

Impact of Artificial Intelligence on Racialized Groups

Focus: AI in Law and Tech

Jake Okechukwu Effoduh
Thursday, 21st September 2023
Visible Minorities Network, Agriculture and Agri-Food Canada

Slide 1

[Introduction and observance of protocol]

Slide 2

This quote, profound in its depth, encapsulates the rise and integration of AI into our daily lives. While scientific in nature, it's a phenomenon that reaches into the very core of our human existence, challenging our beliefs, norms, and notions of equity. AI, the beacon of modern technology, promises innovations beyond our imagination. But, as with all powerful tools, it comes with implications, some of which are starkly evident within our racialized communities. AI's immense potential in sectors like law and tech is undeniable. It powers our search engines, recommends our next binge-worthy show, and even aids in crucial legal decisions. But here lies an inherent challenge. The data that feeds these AI models comes from our world—a world that, as history testifies, has biases. These biases have been palpable. We've seen facial recognition systems failing to recognize darker skin tones like mine, leading to misidentifications and, at times, wrongful arrests. When the algorithms that promise to make our lives more straightforward and safer perpetuate stereotypes and biases, it's a stark reminder that technology, often seen as impartial, absorbs the prejudices in its training data.

Slide 3

Moving to the legal arena, the implications can be even more profound. Predictive policing, an AI tool designed to forecast criminal activity, often targets racialized neighbourhoods due to historical over-policing, creating a self-fulfilling prophecy of crime. In the same vein, algorithms meant to predict recidivism rates sometimes unfairly favour one group over another, impacting parole decisions and sentencing. But, why does this happen? Is it merely a technological glitch? Or does it reflect deeper societal imbalances? The answer is twofold. Yes, the algorithms can and should be refined. But more fundamentally, we need to acknowledge the biases embedded in the datasets we provide to these algorithms. Imagine being handcuffed in front of your neighbours and family for stealing watches. After spending hours behind bars, you learn that the facial recognition software state police used on footage from the store identified you as the thief. But you didn't steal anything; the software pointed cops to the wrong guy. Unfortunately, this is not a hypothetical. This happened three years ago, A facial recognition technology led to the wrongful arrest of a Black man. [Makes reference to the story in Scientific American]: <https://www.scientificamerican.com/article/police-facial-recognition-technology-cant-tell-black-people-apart/#>

Slide 4:

[Shares experience in China as far back as 2015 with an AI system misreading my race but predicting my longevity and other non-visible aspects about me]

Slide 5:

Defining AI in Law

Defining AI is hard:

- A misleading term
- A loaded construct
- “Artificial” is somewhat settled but not “intelligent.”
- The goalpost keeps shifting.
- Must rely on contextualization.

No widely accepted definition.

- Still being queried.
- *The basic locus* of understanding AI in international law is algorithms.
 - A step-by-step procedure for solving a problem.
 - A set of rules to be followed for problem-solving.
 - Analytical methods implemented using a programming language... can make decisions, solve problems, “understand” spoken and written language, and replace humans in the control loops of the system.

“AI is not one thing only, but a constellation of processes and technologies... it is the language of the culture, of companies and governments...” - UN Secretary-General.

“...technology that either assists or replaces the judgement of human decision-makers...from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural networks.” – UN High Commissioner

The *Digital Charter Implementation Act* (“Bill C 27”), containing the *Artificial Intelligence and Data Act* (AIDA) describes AI as a “technological system that, autonomously or partly autonomously, processes data related to human activities through the use of a genetic algorithm, a neural network, machine learning or another technique in order to generate content or make decisions, recommendations or predictions.”

Deconstructions – on racialization and racialized communities: people that society has classified into certain racial categories, often in a manner that emphasizes differences and reinforces hierarchies. These groups are racialized based on certain characteristics, which might include skin colour, facial features, language, culture, or country of origin. Such racialization can result in systemic discrimination, stereotypes, and biases, including with the use of AI, which could even be exacerbated. “Race is the modality in which class is lived, the medium in which class relations are experienced, the form in which it is appropriated and ‘fought through’” - Stuart Hall

Demarcations: How the field of AI intersects with law. It is important to state that “AI *in* Law” may be slightly linked to, but distinct from, other similar frameworks such as “AI *as* Law”, “AI *on* Law” and “AI *and* Law”. These categories are mainly formulated constructs of scholarship and they are not mutually exclusive from each other. However, such cartographic groupings of scholarship in this area may help in the understanding of the varying nature of law’s relationships with AI.

- “AI *in* Law” could simply be regarded as the role AI plays within law scholarship, or ways that the technology is innovating the practice of law itself (how AI can play assistive roles *in* law through fact-finding, evidence-identifying, etc)

- For “*AI on Law*” could demonstrate the technology's contestations towards the field of Law, and how it is reacting to the principles of law, such as with regulation and enforcement.
- For “*AI as Law*”, this metaphor could refer to the ways that AI systems act as instruments, agents or actors of the law itself. For example, AI systems have adjudicated cases as judges, mediated settlements, identified criminals, and even rejected human attempts to disobey traffic rules etc.
- “*AI and Law*” is a much broader concept than all the above, as it encompasses different ways that the two domains relate – how the law arises out of the use and activity of AI, and how, in turn, AI reacts back upon the instrumentality of the law, including the patterns and symmetries occurring within these interactions.

Slide 6:

AI is creating a new world order, but we can see a pattern for racialized communities. Each of the four industrial revolutions has brought about significant technological advancements, but alongside the benefits, these revolutions have also had consequences that disproportionately impacted racialized communities:

- First Industrial Revolution (Late 18th to early 19th Century) - The Age of Mechanical Production: The introduction of water and steam-powered mechanical manufacturing. The cotton gin, invented by Eli Whitney in 1793, revolutionized cotton production but intensified the demand for slave labour in the American South. The device made it possible to clean cotton faster than ever before, making cotton crops more profitable. As a result, slave labor became even more entrenched in the Southern economy, leading to increased suffering and displacement of African-American communities.
- Second Industrial Revolution (Late 19th to early 20th Century) - The Age of Science and Mass Production: The development of electrical energy, the internal combustion engine, and mass production techniques. Urbanization and mass production led to the migration of many people, including racialized groups, to cities in search of work. In the U.S., for example, African Americans moved from the South to the North in the Great Migration. However, upon arrival, they often faced discrimination, were given the lowest-paying jobs, and were subjected to poor living conditions. The burgeoning industries exploited racialized labour while denying them equal rights and opportunities.
- Third Industrial Revolution (Late 20th Century) - The Digital Revolution: The advent of digital technology, computers, and the internet. The digital divide became apparent, where access to computers and the internet was not evenly distributed. Racialized and poor communities often found themselves without the same access to digital resources as their wealthier, predominantly white counterparts. This lack of access hindered opportunities for education, employment, and upward mobility for many in these communities.
- Fourth Industrial Revolution (21st Century) - Age of AI: Inherent biases, e.g., facial recognition software have been shown to misidentify individuals from certain racial and ethnic groups at higher rates than others. Similarly, algorithms used in sectors like finance or criminal justice sometimes show racial biases based on the data they're trained on.

[Makes reference to the story in MIT Tech Review]:

<https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/>

Slide 7:

- Law as Hegemony: Legal frameworks often reinforce the dominant power structures, ensuring the continued supremacy of ruling classes while marginalizing others.
- Data as Extraction: The collection and commodification of data often strip individuals, particularly those from vulnerable communities, of their privacy, autonomy, and cultural richness, mirroring extractive industries.
- Tech as Weapon: Technological advancements, without ethical considerations, can be weaponized to oppress, surveil, and discriminate against marginalized groups.
- Race as Error: Societal constructs of race, underpinned by biases and prejudices, can wrongly categorize and disadvantage racialized groups, perpetuating systemic and algorithmic inequities.

[Makes reference to the story in the Guardian]:

<https://www.theguardian.com/newsletters/2023/aug/15/techscape-facial-recognition-software-detroit-porcha-woodruff-black-people-ai:>

Porcha Woodruff was eight months pregnant when police in Detroit, Michigan came to arrest her on carjacking and robbery charges. She was getting her two children ready for school when six police officers knocked on her door and presented her with an arrest warrant. She thought it was a prank. “Are you kidding, carjacking? Do you see that I am eight months pregnant?” the lawsuit Woodruff filed against Detroit police reads. She sent her children upstairs to tell her fiancé, “Mommy’s going to jail”. She was detained and questioned for 11 hours and released on a \$100,000 bond. She immediately went to the hospital, where she was treated for dehydration. Woodruff later found out that she was the latest victim of false identification by facial recognition. After her image was incorrectly matched to video footage of a woman at the gas station where the carjacking took place, her picture was shown to the victim in a photo lineup. According to the lawsuit, the victim allegedly chose Woodruff’s picture as the woman associated with the robbery’s perpetrator. Nowhere in the investigator’s report did it say the woman in the video footage was pregnant. A month later, the charges were dismissed due to insufficient evidence.

Slide 8:

When scientific or technological decisions are based on a narrow set of systemic, structural, or social concepts and norms, the resulting technology can privilege certain groups and harm others. Bias occurs when an AI algorithm produces systemically prejudiced results due to erroneous assumptions in the machine learning process. An exploration into the fallout of MIT Media Lab researcher Joy Buolamwini's discovery of racial bias in facial recognition algorithms. the Ghanian-American starts "Coded Bias" by showing a massive discovery of how this facial recognition problem, in which the AI did not recognize her face. But when she put on a white mask, it did.

Slide 9:

[Makes reference to my review of the case here]: <https://obiter-dicta.ca/2020/02/25/should-a-court-rely-on-the-proprietary-algorithm-of-an-artificial-intelligence-system-to-make-a-sentencing-decision/>

In 2013, Eric Loomis was found guilty of his role in a drive-by shooting in La Crosse, Wisconsin. During the sentencing phase, the court considered a report generated by COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), a risk assessment tool. The COMPAS report indicated that Loomis was at a high risk of committing another crime in the future. Loomis challenged his sentence, arguing that the use of the COMPAS report violated his due process rights. He raised concerns about the tool's transparency since its algorithm is proprietary and not available for scrutiny. He also mentioned potential biases in the tool. In 2016, the Wisconsin Supreme Court upheld Loomis's sentence, ruling that the judge's use of the COMPAS report was appropriate. However, the court issued cautionary notes on the use of such tools, stating that these assessments should not be the sole determinants in sentencing decisions and that their limitations and potential biases should be clearly acknowledged. The case highlights the growing tension between the use of AI and algorithmic tools in the justice system and the concerns about fairness, transparency, and due process.

Slide 10:

We often think of bias resulting from preferences or exclusions in training data, but bias can also be introduced by how data is obtained, how algorithms are designed, and how AI outputs are interpreted, and more.

"A lot of times, the failings are not in AI. They're human failings, and we're not willing to address the fact that there isn't a lot of diversity in the teams building the systems in the first place. And somewhat innocently, they aren't as thoughtful about balancing training sets to get the thing to work correctly. But then teams let that occur again, and again. And you realize, if you're not thinking about the human problem, then AI isn't going to solve it for you." – Vivienne Ming

- Measurement bias is a systematic distortion of the results of treatment comparisons that can occur when people measuring outcomes - whether researchers, healthcare providers, or study participants - are aware of the treatment allocation when they are measuring outcomes.
- Temporal bias occurs when we assume a wrong sequence of events to someone's disability which misleads our reasoning about causality.
- Omitted-variable bias occurs when a health statistical model leaves out one or more relevant variables. The bias results in the model attributing the effect of the missing variables to those that were included. E.g., the age of the car is negatively correlated with the price of the car and positively correlated with the car's mileage. Hence, omitting the variable age in your regression results in an omitted variable bias. A person's age is negatively correlated with their ability to be a donor, e.g., blood, sperm, bone marrow, and even negatively correlated with the strength of their immune response but positively correlated with their resilience levels.
- Algorithmic bias - systematic and repeatable errors that create unfair outcomes, such as privileging one arbitrary group of users over others.
- Evaluation bias - an unfair evaluation of a person's work or competence
- Aggregation bias is the expected difference between the effects of the group reaction to a punishment and the effects of the individual reaction to the same punishment.

- Popularity bias means that popular items are overly exposed in recommendations at the expense of less popular items that people may find interesting.
- Linkage bias describes the origin of the bias (i.e., in the linkage procedure) but it is not a bias mechanism as such.
- Behavioural bias: irrational beliefs or behaviours that can unconsciously influence our decision-making process. (They are generally considered to be split into two subtypes – emotional biases and cognitive biases)
- Presentation bias: Efforts to present oneself more favourably than others are commonly referred to as self-presentation bias. Self-serving bias.
- Social bias: Any bias that results in unfair or incorrect social assumptions and judgements can be considered a social bias. – Bernie Sanders (ageism), Beauty Bias, Gender Bias (Binary).

Slide 11:

Evaluating the pros and cons of AI for racialized groups and communities within various sectors.

Slide 12:

Based on the projection of countries and sectors that will lead in AI, their past engagements with racialized communities can tell us something about our future.

Slide 13:

Just as AI revolutionizes various sectors, its impact on the legal industry (as well as other populations) is equally transformative.

- Automation and Efficiency: AI technologies have the capability to streamline traditionally labour-intensive processes. However, while greater efficiency can make legal research and legal services more affordable and enhance access to justice, it could lead to job reductions, including for racialized groups, potentially limiting employment opportunities.
- Predictive Analytics: Legal practitioners have the remarkable ability to employ AI algorithms for predicting case outcomes through analysis of historical data, ultimately equipping lawyers to make decisions rooted in data-driven insights. Yet, it is crucial to recognize that the reliability of AI in this context hinges on the fairness, comprehensiveness, and adequacy of the data that is fed into the AI system. This, in turn, raises questions about data quality, bias, and the need for continuous scrutiny and improvement in our quest for fairness in AI-assisted legal decision-making.
- Access to Justice: It is essential to recognize the potential benefits that AI tools can bring to racialized groups, as they can bridge longstanding gaps in access to legal resources. However, realizing these benefits hinges on the careful design of these AI systems to be culturally sensitive and inclusive, ensuring that they consider the diverse needs of their users. AI may not fully grasp certain social cues and non-verbal communication that human professionals can discern, potentially leading to the oversight of critical information that ought to be conveyed.

[Makes reference to this article]: <https://www.fasken.com/en/knowledge/2023/04/federal-court-affirms-the-assistance-of-artificial-intelligence>

Slide 14:

- OECD Principles on AI
 - EU Ethics Guidelines for Trustworthy AI
 - G20 AI Principles
 - UNESCO's Principles for AI.
 - Montreal Declaration for Responsible AI
 - Global Partnership on AI
 - AI Bill of Rights, US
-
- Machine autonomy must not undermine human autonomy.
 - Promoting human well-being and safety and the public interest. AI technologies should not harm people. They should satisfy regulatory requirements for safety, accuracy, and efficacy.
 - Ensuring transparency, explainability and intelligibility. AI should be intelligible or understandable to developers, users, and regulators.
 - Fostering responsibility and accountability. Responsibility can be assured by the application of “human warranty”, which implies evaluation by patients and clinicians in the development and deployment of AI technologies.
 - Ensuring inclusiveness and equity, irrespective of age, gender, income, ability, or other characteristics.
 - Promoting AI that is responsive and sustainable. Responsiveness requires that designers, developers, and users continuously examine an AI technology to determine whether it is responding adequately, appropriately, and according to the context in which it is used.

Slide 15:

- Damages: the sum of money the law imposes for a breach of some duty or violation of some right. Generally, there are two types of damages: compensatory and punitive.
- Product liability vs. strict liability: When a person sues another for such cause of action, the Plaintiff (the person suing) needs to prove four elements: duty of care, breach of that duty, causation, and damages. If one of those elements is missing, you'll lose your case. However, certain companies and manufacturers promote products or activities that are so inherently dangerous if the end user is injured, they'll be able to sue for strict liability. No need to prove negligence. Examples of such products include certain chemicals, flammables, explosives, or owning wild animals. If the activity involves a serious risk of harm, and that harm cannot be eliminated by using reasonable care, once someone gets injured, they'll be able to sue for strict liability. (e.g., EU and US courts have regarded medical software as a medical device because of its intended use). Strict liability can be imposed where harm from bias occurs, even in the absence of negligence, recklessness, or intent to harm. (Strict liability – e.g. dog bite). Product liability refers to when you're trying to hold a manufacturer liable for an injury you sustained while using their product. Unlike in strict liability cases, you must be able to show some sort of negligence on the part of the Defendant (the entity getting sued). There are three ways you can do so in these kinds of cases: Design Defects, Manufacturing Defects, Failure to Warn

- No fault, no liability compensation funds. Providing payments to individuals who suffer injuries due to the use of AI technologies. In New Zealand, for example, patients seek compensation for medical injuries through a no-fault, no-liability scheme. Injured patients receive government-funded compensation, thereby giving up the right to seek damages, except in rare cases of reckless conduct.
- Pre-emption laws: If a commercial algorithm is approved by a regulatory agency, the doctrine of pre-emption may apply, i.e., that a decision taken by a central government agency to validate an AI technology will supersede any cause of action guided by civil laws.
- Data protection laws and guidelines on the use of sensitive/indigenous data for AI
- Government co-regulation with the private sector to understand how to deal with harms from bias in AI systems without limiting independent regulatory oversight, also PPPs.

Slide 16:

[Makes reference to the Call for Code for Racial Justice is an initiative external to IBM]:
<https://www.ibm.com/blog/call-for-code-for-racial-justice-projects/>

- Counteraction: AI can be used to counteract bias. E.g., decision support tools such as legal decision support systems, multi-disciplinary data science teams
- Control mechanisms: “human-in-the-loop” systems – human decision-maker as caveat; train algo developers – Professional Responsibility
- Affirmative data and Reverse engineering: Use of preferential randomized data instead of observational data to support representative development of the AI system; Technological reverse-engineering drivers of bias and use of counterfactuals.
- Bias prediction models and sensitivity checks: to ensure outputs are not reinforcing existing social biases. Also, oversampling underrepresented populations to mitigate statistical bias.

Slide 17:

1. Moratorium on AI systems that are high-human risk: In September 2021, the United Nations High Commissioner for Human Rights proposed that any AI applications that cannot be operated in compliance with international human rights law should be expressly banned, and moratoriums imposed on the sale and use of AI systems that carry a high risk for the enjoyment of the right to health and other rights.

2. Enforce the right to contest bias (and right to meaningful explanation where bias from an AI system may have occurred: Office of the Privacy Commissioner of Canada’s recommendations for the reform of the Personal Information Protection and Electronic Documents Act, 2000 (PIPEDA) which proposes the the right to meaningful explanation where bias from an AI system may have occurred (like Art 15(1)(h) of the GDPR) and right to contest bias (similar to Art 22, GDPR).

3. Limit the use of automated decision-making by AI for health (maybe as last resort?). Or introduce “human in the loop” to adhere to obligations of accountability and explainability. Human-in-the-loop artificial intelligence refers to AI decision-making processes where humans may provide feedback or

confirmation. rethink this phrase and consider a future where humans remain at the center of all AI technologies. AI should effectively communicate and collaborate with people to augment their capabilities and make their lives better and more enjoyable. Humans are not simply “in the loop.” Humans are in charge; AI is “in the loop.” ... “Lost in the loop”- Human rights in the loop?

4. Prescription of debiasing orders: Debiasing techniques can help provide some remedy to biased AI for health systems that cause harm. Google has rolled out AI debiasing initiatives, including responsible AI practices featuring advice on making AI algorithms fairer. Also, mandating periodic “algorithmic hygiene” can ensure that as much as possible training data are representative.

5. A Human Rights Algorithmic Impact Assessment on AI for Health is valuable: In 2019, two senators in the USA co-sponsored the “Algorithmic Accountability Act”, which would require companies to study and adjust flawed algorithms that result in inaccurate, unfair, biased, or discriminatory decisions that would affect people in the USA. (Looked at the Act, no definition of Algorithm). ... EU AI Act? Canada’s AIDA?

6. Right to be Forgotten from AI for health systems: Patients should have the right to ensure that any private information about them can be permanently removed from AI systems to prevent being stigmatized by their past. The right to be forgotten derives from the case Google Spain SL, Google Inc v Agencia Española de Protección de Datos, Mario Costeja González (2014). For the first time, the right to be forgotten became codified and to be found in the GDPR, in addition to the right to erasure.

7. Third World approach to harmonization of standards: A decolonized global harmonization of regulatory standards that is pro-poor can ensure that all countries, including countries in the third world, can benefit from rigorous testing, transparent communication of outcomes and monitoring of AI. Third World Approaches to International Law (TWAIL) is a critical school of international legal scholarship and an intellectual and political movement. TWAIL is a dynamic, intentionally open-ended and decentralized... mutually constitutive relationship between international law and the third world/global South; the TWAIL network enables solidarity and mutual support through a shared political commitment to advocating for the interests of the global South. It endeavours to give voice to viewpoints systemically underrepresented or silenced.

8. Upskilling: Good support and training will ensure that workers can avoid common pitfalls such as automation bias when using AI technologies. Research: Into how best providers and programmers can address any biases that will manifest in AI applications against racialized people. Non techno-solutionist interventions. E.g., ensuring diversity in schools, inclusive AI design teams, co-production and deploying non-AI solutions where practicable.

9. Incentives: To encourage developers to identify and avoid bias (e.g., the precertification program of the US Food and Drug Administration, the Digital Health Innovation Action Plan). The Software Precertification (Pre-Cert) Pilot Program will help inform the development of a future regulatory model that will provide more streamlined and efficient regulatory oversight of software-based medical devices. This pilot, limited to manufacturers of software as a medical device (SaMD), is an important first step to help us explore and evaluate the program model to inform how we establish the Pre-Cert Pilot Program

Slide 18:

Conclusion: It brings us back to our starting quote. AI has happened, and it's "very big indeed." But as it weaves into the fabric of our lives, we must ensure it embraces the diverse tapestry of our societies. It isn't just a tech challenge. It's an ethical, humanistic, and aesthetic one.

Slide 19:

Thank you/Questions.