Christian, C. (2023, April 27–29). *Coded bias: Decoding racism in artificial intelligence technologies* [Paper presentation]. Coding, Computational Modeling, and Equity in Mathematics Education Symposium, St. Catharines, ON, Canada.

Coded Bias: Decoding Racism in Artificial Intelligence Technologies

Gideon Christian
Faculty of Law, University of Calgary

Introduction

Artificial intelligence (AI) can simply be defined as the science and engineering of making computers perform tasks that, if performed by humans, would require human or cognitive intelligence (Scherer, 2016). AI is a broad field of scientific endeavor. This paper will focus on two aspects of AI technology where Black people and people of color should be really concerned about the use of this technology – that is in the area of AI recidivism risk assessment and AI facial recognition technology.

We are witnessing the increasing use of AI risk assessment in the criminal justice system, and AI facial recognition technology in public sectors – from immigration enforcement to criminal law enforcement and beyond. These technologies have been touted as major innovations in the field of artificial intelligence. However, their use comes with increasing concern about their potential to perpetuate race-based bias and discrimination against Blacks and people of color. This phenomenon has been variously referred to by Michele Alexander (2010) as "the new Jim Crow" and by Ruha Benjamin (2019) as "the new Jim Code."

In criminal justice risk assessment, AI tools have been shown to assess Black defendants at a higher rate of recidivism, doing so at "twice the rate as White defendants" (Angwin et al., 2016, para. 15). Similarly, AI facial recognition technology has been shown to have a higher tendency to misidentify Black people and people of color (Najibi, 2020), leading to wrongful arrests and detentions in some cases. Thus, many AI tools appear to exhibit inherent bias against Black people and people of color (Johnson, 2022). This prompts the question: How and why do these tools demonstrate inherent bias against specific racial groups?

Algorithm and Bias

There are many ways biased outcomes could result from AI tools. First, when an AI tool is deliberately designed to generate biased outcomes. This may be rare but is a clear case of explicit bias resulting from bad faith software engineering. The second situation is when biased data is used to train the computer algorithm to make predictions. This would result in a biased outcome, an extension of what is referred to in the computer lexicon as "garbage in, garbage out". The third situation arises when an AI tool trained on data that is responsive to unique factors in a particular environment is deployed for use in a different environment. The second and third cases above would generally result in implicit bias.

AI in criminal justice recidivism risk assessment

Recidivism risk assessment is the process of determining the likelihood that an individual in the criminal justice system – an accused, convicted, or incarcerated person – will reoffend (Christian, 2020). Recidivism risk assessment tools that rely on AI technology are now increasingly being used in the criminal justice systems around the world. Research studies on the use of AI in criminal justice risk assessment have revealed the tendency of AI technologies to automate inequality and perpetuate bias, discrimination, and stereotypes against minority populations characterized by race (Angwin et al., 2016; Chouldechova, 2016).

In Canada, these minority groups are disproportionately represented in the criminal justice system. The unfair representation of Blacks and Indigenous people in the Canadian criminal justice system is well-documented (see *R. v. Jackson*, 2018). Although Blacks represent 4% of the adult population in Canada, they currently constitute 9% of the total federal in-custody population (Department of Justice Canada, Research and Statistics Division, 2022). Similarly, the over-representation of Indigenous persons in correctional facilities in Canada increased by 14% in 2020/2021 (Robinson et al., 2023). Although AI risk assessment tools are not currently being used in Canada, their use in the US criminal justice system is very pronounced. Hence, it is only a matter of time before Canada follows the trend. Thus, there is concern that the use of the tools in Canada may eventually automate the already existing racial bias and inequity evident in the Canadian criminal justice system.

In 2014, then US Attorney General Eric Holder warned of the danger in risk score injecting bias into the court system (Holder, 2014). He noted that while these risk assessment tools might have been designed with the best intention, they inadvertently undermine efforts to ensure justice by exacerbating biases that already exist in the criminal justice system (Holder, 2014).

The fear expressed by Eric Holder was confirmed by a research study undertaken by the US non-profit newsroom ProPublica, which revealed the tendency of algorithmic risk assessment tools to perpetuate existing racial bias and stereotypes in society (Angwin et al., 2016). The study focused on recidivism risk predictions by COMPAS – the most widely used AI risk assessment tool in the US criminal justice system. In addition to raising some doubts about the accuracy of predictions made by this risk assessment tool, the study raised even more serious concerns regarding the racial bias evident in COMPAS risk predictions. The study revealed that COMPAS not only falsely assesses Black defendants as future criminals but again does so twice as often as White defendants (Angwin et al., 2016). In fact, White defendants were more often misclassified as "low risk" than Black defendants. In some cases, Black defendants labeled by COMPAS as "high risk" did not reoffend, while White defendants labeled as "low risk" went on to reoffend. So, the study from ProPublica seems to establish that the use of COMPAS in risk assessment results in White defendants getting an unfair pass while Black defendants, on the other hand, get an unfair penalty. It is surprising then that, notwithstanding these findings, the COMPAS AI risk assessment software remains one of the most dominant AI-based risk assessment tools in the US criminal justice system.

Algorithmic Racism

The tendency of results from AI risk assessment tools to be biased against people of colour has been referred to as algorithmic racism, which is defined as "systemic, race-based bias arising from the use of AI-powered tools in the analysis of data in decision making resulting in unfair outcomes to individuals from a particular segment of the society distinguished by race" (Christian, 2020, para. 3). Algorithmic racism may result from the use of historical data collected from the era of biased policing and carceral justice system to develop or train AI technology used in risk assessment.

The fact, though, is that AI risk assessment cannot be better than the data used to train or develop the tools used in the assessment. The problem with using data collected from these discriminatory carceral and law enforcement practices is that individuals from minority communities are disproportionately represented in the data, thus implying that those communities are more prone to crime or more likely to commit crime. An AI tool trained on such data will

inevitably regurgitate the bias and stereotypes implicit in the data. Such bias would even become more blurred by a false belief in technology as being race-neutral and color-blind.

Algorithmic bias – Square Peg in a Round Hole

Algorithmic bias would also arise when an AI tool developed and trained on data from one predominant racial group is sought to be used or deployed for use in another racial group not adequately represented in the training data. This problem has been evident in facial recognition software, which is discussed below. To illustrate this situation in the context of criminal justice risk assessment, consider the Supreme Court of Canada case of *Ewert v. Canada* (2018). In *Ewert*, an Indigenous man was serving consecutive life sentences. While he was serving his sentence, Correctional Services Canada (CSC) reviewed and upgraded his security classification after assessing his risk using an actuarial risk assessment tool.

The tools used by CSC were developed and tested on a predominantly non-Indigenous population. Applying such tools to assess the risk posed by an Indigenous person – a person from a racial group not adequately represented in the training data used to develop the tool – would reasonably raise some serious concerns as to the accuracy or validity of any assessment made by such tools. The Supreme Court in *Ewert* (2018) noted that such cross-cultural application of risk assessment tools would result in "cultural bias" (para. 71). Hence, deploying AI tools for use in an environment different from those for which the tools were originally and purposefully designed will result in biased and inaccurate predictions of risk.

AI risk assessment tools are not one-size-fits-all. These tools must be confined to the environment or groups for which they were originally designed. Efforts should also be made to ensure that even in the environment for which the tools are designed or created, the population in that environment is adequately represented in the training data used to develop the technology. Thus, representation is tied to the accuracy or lack of accuracy in the predictions made by AI tools.

Facial Recognition Technology – the new Jim Crow

Jim Crow laws were a series of state and local laws in the United States that enforced racial segregation and discrimination against Black people and people of color from the late 19th century until the mid-20th century. These laws essentially denied recognition to Black people in certain spaces, as evidenced by their institutionalization of segregation, among other discriminatory practices. The denial of recognition to these groups is also evident today in facial recognition technology, which struggles to accurately recognize Black faces, as evidenced by its high error rate in recognizing Black faces, especially Black women.

Realistically, Jim Crow laws differ from AI facial recognition technology in the sense that the former was a deliberate state policy, and the later is not. However, both phenomena are similar in that they both involve systemic discrimination against Black individuals, they both result in the perpetuation of discrimination which disproportionately affects Black individuals, and the end result is that Black individuals bear the brunt of the negative consequences in both cases.

The adverse impacts resulting from the use of facial recognition technology in law enforcement are evident from various incidents in the United States of wrongful arrest and detention of Black people as a result of false identification by AI facial recognition technology (Johnson, 2022). An example is the case of Nijeer Parks, a Black man from New Jersey who in 2018 was wrongfully arrested and charged with serious crimes after facial recognition technology incorrectly matched his photo to that of the perpetrator of the crime (General & Sarlin, 2021). He spent 10 days in jail, and it took almost a year before the charges against him were dropped after

he found a transaction receipt which placed him 30 miles away from the crime scene, thus providing a perfect alibi. Nijeer's case represents one among many others in which innocent Black people have been wrongly arrested and, in some cases, falsely charged and locked up for crimes they never committed, all as a result of false identification by AI facial recognition technology.

AI facial recognition technology in Canada

For an unknown period of time, AI facial recognition technology was deployed and used by various law enforcement agencies in Canada, including the RCMP and Toronto Police (see Boutilier, 2021). The clandestine use of AI facial recognition technology by these law enforcement agencies sparked outrage when it became public that they were using the controversial Clearview AI software. This resulted in a joint investigation by the Office of the Privacy Commissioner of Canada, Commission d'accès à l'information du Québec, Information and Privacy Commissioner for British Columbia, and Information Privacy Commissioner of Alberta (2021) and even a Parliamentary committee hearing (Standing Committee on Access to Information, Privacy and Ethics, 2022). It is important to note, though, that the outcry against the use of AI facial recognition technology by these public agencies in Canada was not based on the racist nature of the technology – that is, its high propensity to misidentify people of color. Rather, the outcry related to the unethical and indeed unlawful methods adopted by the developer in scraping the internet for people's photos without their consent and using such photos to build their database. Hence, a case of privacy breach/concern.

Aside from the privacy issue evident from the Clearview software, which was concerning to many White people, Black people and people of color indeed have even more concerns – the high error rate in the identification of individuals from these racial groups by the technology. This is what critical race theory refers to as interest convergence – a situation where the interest of the minority group is advanced because it aligns or converges with that of the dominant group.

Unlike the US, in Canada, we do not have any highly publicized cases of Black people being wrongfully arrested because of misidentification by AI facial recognition technology. While we cannot categorically state that such incidents never existed in Canada, it is important to note that police use of this tool in Canada was clandestine. Hence, even where there might have been actual cases of wrongful arrests arising from misidentification by the police use of the tool in Canada, the use of such information in effecting the arrest might have been concealed by the same police department secretly using the technology.

However, the use of facial recognition technology in the Canadian immigration system began to gather attention with the publication of the Federal Court of Canada decision in the case of *Barre v. Canada (Citizenship and Immigration)* (2022). This case involved two Black Somali women who made a successful refugee claim in Canada. The Minister of Public Safety later sought to revoke their refugee status on the ground that their faces matched those of other women who entered Canada as Kenyan citizens. The women challenged the revocation of their refugee status at the Federal Court. While the women asserted that the government used the discredited Clearview software in matching their faces to those of other women, the Minister refused to disclose whether it used the technology, asserting that the information was protected by law. The Federal Court not only accepted the women's assertion that the facial recognition technology was used by the government, but the court also held that the disclosure of information relating to the government's use of the technology here was not protected by law (the Privacy Act). Referring to facial recognition software as an "unreliable pseudoscience," the court also expressed a great deal of concern about the use of the technology by the government on the women, considering "that

darker-skinned females are the most misclassified group with error rates of up to 34.7%, as compared to the error rate for lighter-skinned males at 0.8%" (*Barre v. Canada*, 2022, para. 25; see also Lohr, 2018).

Unfortunately, while *Barre* was the first, it was not the last reported case of Canadian government agencies seeking to use evidence from facial recognition technologies to revoke the refugee status of successful refugee claimants of Black race. In fact, to date, all reported Federal Court cases in which the Canadian government has sought to use evidence from facial recognition technology to revoke successful refugee claims have involved individuals of Black race and predominantly Black women (in addition to *Barre vs. Canada*, 2022, see also *AB v. Canada*, 2023; *Abdulle v. Canada*, 2023; *Ali v. Canada*, 2023). The same racial/gender group where the technology has it highest error rate.

What is also concerning about the use of this technology by the Canadian government agencies is the lack of transparency in its use by the government. In all the Federal Court cases to date, the government agencies have refused to even admit their use of the technology. Rather, they attempt to rely on legal provisions (though unsuccessfully) to avoid the disclosure of such information. AI facial recognition technology is a Blackbox, responsible government use of the technology in context affecting the rights of citizens (irrespective of race) should be characterised by openness and transparency.

While ethical guidelines for the use of AI facial recognition technology are still evolving, the following common-sense principle should guide every public sector use of the technology in Canada: When the use of the technology affects an individual's rights, the government agency must inform the impacted individual and provide them with an opportunity to challenge the decision made by, or using the technology. Hopefully, the government agencies will incorporate this common-sense principle into their use of the technology to achieve the measure of transparency and openness that is characteristic of the system of government in Canada.

Conclusion

Algorithmic bias impeded in AI technologies plays a significant role in generating unfair outcomes along racial lines. Whether arising from deliberately designed biased algorithms, biased training data, or the misapplication of tools to different racial groups not adequately represented by the training data, the result is the same: perpetuation of discrimination and inequality. To achieve accurate and just results, AI tools must be developed and trained with adequate representation from the communities they will affect. Ignoring this crucial aspect leads to systemic discrimination, as evidenced in the cases of AI risk assessment tools and facial recognition technology.

To ensure responsible and ethical use of AI facial recognition technology, transparency and openness must be embraced by Canadian government agencies. Public sector use of the technology should prioritize informing individuals impacted by its use and allowing them an opportunity to challenge any decisions made based on the technology's outcomes. This opened and transparency will enable the decoding of racial bias imbedded in AI technologies.

References

AB v. Canada (Citizenship and Immigration), 2023 FC 29 (CanLII). https://canlii.ca/t/jts52

Abdulle v. Canada (Citizenship and Immigration), 2023 FC 162 (CanLII). https://canlii.ca/t/jvdcp

Alexander, M. (2010). *The new Jim Crow: Mass incarceration in the age of colorblindness*. The New Press.

Ali v. Canada (Citizenship and Immigration), 2023 FC 671 (CanLII). https://canlii.ca/t/jx64r

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Barre v. Canada (Citizenship and Immigration), 2022 FC 1078 (CanLII). https://canlii.ca/t/jr6r8

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity.

Boutilier, A. (2021, June 10). RCMP broke privacy laws in using controversial Clearview AI facial recognition tools, watchdog says. *Toronto Star*. www.thestar.com/politics/federal/2021/06/10/rcmp-broke-privacy-laws-in-using-controversial-clearview-ai-facial-recognition-tools-watchdog-says.html

Chouldechova, A. (2016, October 24). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *arXiv*. https://doi.org/10.48550/arXiv.1610.07524

Christian, G. (2020, October 26). Artificial intelligence, algorithmic racism and the Canadian criminal justice system. *Slaw*. https://www.slaw.ca/2020/10/26/artificial-intelligence-algorithmic-racism-and-the-canadian-criminal-justice-system/

Department of Justice Canada, Research and Statistics Division. (2022, December). Overrepresentation of Black people in the Canadian criminal justice system. *JustFacts*. https://www.justice.gc.ca/eng/rp-pr/jr/obpccjs-spnsjpc/index.html#_ftnref26

Ewert v. Canada, 2018 SCC 30, [2018] 2 S.C.R. 165. https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/17133/index.do

General, J., & Sarlin, J. (2021, April 29). *A false facial recognition match sent this innocent Black man to jail*. CNN. https://www.cnn.com/2021/04/29/tech/nijeer-parks-facial-recognition-police-arrest/index.html

Holder, E. H., Jr. (2014, August 1). *Attorney General Eric Holder speaks at the National Association of Criminal Defense Lawyers 57th annual meeting and 9th State Criminal Justice Network conference* [Speech]. https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th

Johnson, K. (2022, March 7). *How wrongful arrests based on AI derailed 3 men's lives*. *Wired*. https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/

Lohr, S. (2018, February 9). Facial recognition is accurate, if you're a White guy. *The New York Times*. www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html

Najibi, A. (2020, October 24). Racial discrimination in face recognition technology. *SITN*. https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

Office of the Privacy Commissioner of Canada, Commission d'accès à l'information du Québec, Information and Privacy Commissioner for British Columbia, and Information Privacy Commissioner of Alberta. (2021, February 2). *PIPEDA Findings #2021-001, Joint investigation of Clearview AI, Inc*. https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/

R. v. Jackson, 2018 ONSC 2527 (CanLII). https://canlii.ca/t/hrm8w

Robinson, P., Small, T., Chen, A., & Irving, M. (2023, July 12). Over-representation of Indigenous persons in adult provincial custody, 2019/2020 and 2020/2021. *Juristat*. https://www150.statcan.gc.ca/n1/en/catalogue/85-002-X202300100004

Scherer, M. U. (2016). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law and Technology*, *29*(2), 354–400. https://doi.org/10.2139/ssrn.2609777

Standing Committee on Access to Information, Privacy and Ethics (ETHI). (2022, October). *Use and impact of facial recognition technology: Report of the Standing Committee on Access to Information, Privacy and Ethics*. https://www.ourcommons.ca/DocumentViewer/en/44-1/ETHI/report-6