

A REASONABLE APPREHENSION OF AI BIAS: LESSONS FROM *R. v. R.D.S.*

*Teresa Scassa**

In 1997, the Supreme Court of Canada rendered a split decision in a case in which the central issue was whether an African Nova Scotian judge who had brought her lived experience to bear in a decision involving a confrontation between a black youth and a white police officer had demonstrated a reasonable apprehension of bias. The case, with its multiple opinions across three courts, teaches us that identifying bias in decision-making is a complex and often fraught exercise.

Automated decision systems are poised to dramatically increase in use across a broad range of contexts. They have already been deployed in immigration and refugee determination, benefits allocation, and in assessing recidivism risk. There is also a growing use of AI-assistance in human decision-making that deserves scrutiny. For example, generative AI systems may introduce dangerous unknowns when it comes to the source and quality of briefing materials that are generated to inform decision-makers. Bias and discrimination have been identified as key issues in automated decision-making, and various solutions have been proposed to prevent, monitor, and correct potential issues of bias. This paper uses *R. v. R.D.S.* as a starting point to consider the issues of bias and discrimination in automated decision-making processes, and to evaluate whether the measures proposed to address bias and discrimination are likely to be effective. The fact that *R. v. R.D.S.* does not come from a decisional context in which we currently use AI does not mean that it cannot teach us—not just about bias itself—but perhaps more importantly about how we think about and process issues of bias.

En 1997, la Cour suprême du Canada a rendu une décision partagée dans un dossier où la question principale était de déterminer si une juge afro-néo-écossaise avait fait preuve d'une crainte raisonnable de biais en valorisant son expérience vécue pour rendre une décision. L'affaire dans laquelle la juge tranchait concernait une confrontation entre un jeune noir et un policier blanc. Ce dossier, et ses plusieurs opinions à travers trois tribunaux, nous enseigne que l'identification du biais dans la prise de décision est une tâche complexe et souvent difficile.

Les systèmes de décision automatisés — qu'ils soient entièrement automatisés ou assistés par l'intelligence artificielle (IA) — sont placés à être utilisés d'avantage à travers plusieurs contextes. Des systèmes de décision automatisés ont déjà été déployés dans les domaines de l'immigration et de la détermination du statut de réfugié, de l'attribution des bénéfiques et de l'évaluation du risque de récidivisme. L'utilisation croissante d'assistance de l'IA dans la prise de décision humaine mérite également d'être examinée de près. Par exemple, l'origine et la qualité des rapports produits par des systèmes d'IA générative qu'utilisent les décideurs peuvent introduire des erreurs dangereuses. Le biais et la discrimination ont été identifiées comme des problèmes clés dans la prise de décision automatisée, et plusieurs solutions ont été proposées pour éviter, surveiller et corriger des potentiels problèmes de biais. Cet article utilise l'affaire *R. c. R.D.S.* comme point de départ pour examiner les questions de biais et de discrimination dans les processus de prise de décision automatisée et pour évaluer si les mesures proposées sont susceptibles d'être efficaces. Le fait que l'affaire *R. c. R.D.S.* n'est pas issue d'un contexte décisionnel dans lequel nous utilisons l'IA ne signifie pas qu'elle ne peut pas nous apprendre - non seulement sur le biais — mais peut-être plus importamment, comment nous envisageons et traitons les questions de biais.

* Canada Research Chair in Information Law and Policy, University of Ottawa. I am grateful for helpful feedback on an earlier draft of this paper from Jennifer Raso, as well as for comments on a more recent version from participants at the McGill Law Journal 2024 Symposium, and from the anonymous peer reviewers. Many thanks to the excellent editorial team at the McGill Law Journal. This paper was written without the use of artificial intelligence. I gratefully acknowledge the support of the Canada Research Chairs Program.