INSIGHT 5

USING **AI** TO PREDICT **HUMAN BEHAVIOUR** MAY NOT WORK

While AI sometimes makes impressive predictions about human behaviour, many are inaccurate. Basing decisions on these predictions can have dire consequences for people. It might be impossible to improve the technology to a level where its benefits outweigh the costs.

TODAY

More governments and institutions are using AI to predict human behaviour and make decisions about individuals. For example, more than 500 schools in the U.S. use an AI model called Navigate to predict student success.[1] Social workers in the U.S. have used AI to predict which child welfare calls need further investigation.[2] Both are examples of "predictive optimization".[3] Notable AI engineers have argued that predictive optimization algorithms are based on faulty science, with AI predictions being only slightly more accurate than the random flip of a coin.[4] Despite this, they continue to be used because they outsource complex work like developing decision-making rules (e.g. what criteria to investigate for fraudulent behaviour or how to decide if a child is at risk of abuse). Human-generated decision-making rules can appear subjective and inaccurate compared to those of predictive AI models, which claim to reflect objective patterns in the real world.

Predictive optimization

The use of AI to predict future outcomes based on historical data, to make decisions about individuals.

Predictive models are not always right. Predictive AI models are plagued by many issues, including errors due to a mismatch between training data and deployment data. Because predictive AI must be trained on past data, it cannot account for emergent and complex variables in the world and in individual human behaviours. Models may be unable to account for new and unexpected drivers. Moreover, AI cannot filter out the effects of racist real-world practices such as disproportionate policing in Black neighbourhoods or communities, which leads to increased false arrests.[5] This has led to inaccurate predictions for vulnerable people.[6]

Predictive AI models cannot understand why real-world behaviour differs from their predictions. Models may assume that individuals will act rationally and consistently or follow the same rules and patterns of humans in aggregate. Models may not address the structural factors that account for differences between predicted and real-world behaviours. A focus on prediction may hinder the discovery of processes that can lead to new behaviours, such as when simplifying the language used on court summons reduced the rate of people failing to appear in court.[7]

While sometimes justified based on cost savings, some governments have felt significant repercussions after using of predictive optimization models. For example, in 2021, the Dutch government resigned over a scandal involving the tax authority's adoption of a self-learning AI to predict childcare benefits fraud.[8] The AI erroneously identified tens of thousands of families as owing excessive debts to the tax authority. Over 3,000 children were removed from their homes and many families remain separated. The scandal had significant repercussions, with families forced into debt, losing their homes, and some victims dying by suicide.

FUTURES

In the future, predictive optimization may be used in some jurisdictions but not others. It could be forbidden within some jurisdictions, particularly where governments have faced high costs and scrutiny due to failures. That could still allow the private sector to expand its currently opaque uses of predictive optimization.[9] Other jurisdictions may continue to use predictive optimization algorithms despite the risks. This could be because those affected are less able to pursue justice, or because their governments are not bound by democratic norms. Others may view predictive optimization as an inevitably imperfect tool, but one whose use can be justified due to cost savings. Institutions — including governments — that take up AI for predictive optimization and find that the costs outweigh the benefits could keep systems in operation far longer than they should or want to, due to the high amounts already invested or the difficulties involved in undoing a rollout. Some may see predictive AI as ethically unacceptable for decision-making, and instead work on interventions to minimize the predicted negative outcomes.

IMPLICATIONS

- Governments and companies that use predictive optimization without being transparent about the AI's decision-making rules **could be seen as untrustworthy**
- If institutions use AI for predictive optimization while the burden of proof to contest inaccurate predictions is put on affected individuals, **already vulnerable populations may face worsened outcomes**. This could create new bureaucratic bottlenecks and tie up courts with algorithmic harms litigation, including cases related to human rights or Charter violations
- Attempts to sacrifice individual rights for collective gains may **benefit privileged populations at the expense of the vulnerable**, creating greater socio-economic divisions
- The uptake of predictive optimization models could create **initial cost savings that quickly give way to new costs**: to fight litigation from inaccurate predictions; to recontract providers to retrain and retune models; and to create new pathways for complaints and compensation for damages
- If AI decision-making pre-emptively punishes people based on biased assumptions, it could **decrease the individual agency of vulnerable populations** and place new obstacles in their life courses

Endnotes

1 Feathers, Todd. Texas A&M Drops "Race" from Student Risk Algorithm Following Markup Investigation – The Markup, 30 March 2021.

2 PBS News. AP Report: DOJ Examining AI Screening Tool Used by Pa. Child Welfare Agency. PBS NewsHour, 31 January 2023.

3 Wang, Angelina, Sayash Kapoor, Solon Barocas, and Arvind Narayanan. Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms That Optimize Predictive Accuracy. SSRN Scholarly Paper. Rochester, NY, 4 October 2022.

4 Wang, Angelina, Sayash Kapoor, Solon Barocas, and Arvind Narayanan. Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms That Optimize Predictive Accuracy. SSRN Scholarly Paper. Rochester, NY, 4 October 2022.

5 Knox, Dean, Will Lowe, and Jonathan Mummolo. Administrative Records Mask Racially Biased Policing. American Political Science Review 114, no. 3 (August 2020): 619–37.

6 Wang, Angelina, Sayash Kapoor, Solon Barocas, and Arvind Narayanan. Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms That Optimize Predictive Accuracy. SSRN Scholarly Paper. Rochester, NY, 4 October 2022.

7 Fishbane, Alissa, Aurelie Ouss, and Anuj K. Shah. Behavioral Nudges Reduce Failure to Appear for Court. Science 370, no. 6517 (6 November 2020): eabb6591.

8 Brenninkmeijer, Alex, and Björn ten Seldam. The Dutch Benefits Scandal: A Cautionary Tale for Algorithmic Enforcement. EU Law Enforcement (blog), 30 April 2021.

9 Rhea, Alene K., Kelsey Markey, Lauren D'Arinzo, Hilke Schellmann, Mona Sloane, Paul Squires, Falaah Arif Khan, and Julia Stoyanovich. An external stability audit framework to test the validity of personality prediction in AI hiring. Data Mining and Knowledge Discovery 36, no. 6 (2022): 2153-2193.

© His Majesty the King in Right of Canada, 2025

For information regarding reproduction rights: https://horizons.gc.ca/en/contact-us/

PDF: PH4-215/2025E-PDF
ISBN: 978-0-660-76889-2

Aussi disponible en français sous le titre : L'utilisation de l'IA pour prédire le comportement humain pourrait s'avérer inefficace.

DISCLAIMER
Policy Horizons Canada (Policy Horizons) is the Government of Canada's centre of excellence in foresight. Our mandate is to empower the Government of Canada with a future-oriented mindset and outlook to strengthen decision making. The content of this document does not necessarily represent the views of the Government of Canada, or participating departments and agencies.