Algorithmic Impact Assessment Results

Version: 0.9

Name of Respondent

Adam Bakopolus

Job Title

Data Analyst

Branch

Department of Education

Project Title

Student Risk Prevention Algorithm

Project ID from IT Plan

8675309

Project Phase
Design [Points: 0]

Please provide a project description:

This algorithm is used to predictstudents at risk of being arrested or in crisis by creating amodel that uses a families' "zip codes, incomes, truancynumbers, race, and other indicators." The data used in thesesystems is often shared with law enforcement and othergovernment agencies.

What is motivating your team to introduce automation into this decision-making process? (Check all that apply)

Improve overall quality of decisions
Use innovative approaches
The system is performing tasks that humans could not accomplish in a reasonable period of time

Please check which of the following capabilities apply to your system.

Risk assessment: Analyzing very large data sets to identify patterns and recommend courses of action and in some cases trigger specific actions

Section 1: Impact Level : 2

Current Score: 53

Raw Impact Score: 53

Mitigation Score: 30

Section 2: Requirements Specific to Impact Level 2

Peer Review

At least one of: Qualified expert from a federal, provincial, territorial or municipal government institution. Qualified members of faculty of a post-secondary institution. Qualified researchers from a relevant non-governmental organization. Contracted third-party vendor with a related specialization. Publishing specifications of the Automated Decision System in a peer-reviewed journal. A data and automation advisory board specified by Treasury Board Secretariat.

Notice

Plain language notice posted through all service delivery channels in use (Internet, in person, mail or telephone).

Human-in-the-loop for decisions

Decisions may be rendered without direct human involvement.

Explanation Requirement

In addition to any applicable legal requirement, ensuring that a meaningful explanation is provided with any decision that resulted in the denial of a benefit, a service, or other regulatory action.

Training

Documentation on the design and functionality of the system.

Contingency Planning

None

Approval for the system to operate

None

Other Requirements

The Directive on Automated Decision-Making also includes other requirements that must be met for all impact levels.

Link to the Directive on Automated Decision-Making

Contact your institution's ATIP office to discuss the requirement for a Privacy Impact Assessment as per the Directive on Privacy Impact Assessment.

Section 3: Questions and Answers

Section 3.1: Impact Questions and Answers

Is the project within an area of intense public scrutiny (e.g. because of privacy concerns) and/or

frequent litigation?
Yes [Points: +3]

Are clients in this line of business particularly vulnerable?
Yes [Points: +3]

Are stakes of the decisions very high?
Yes [Points: +4]

Will this project have major impacts on staff, either in terms of their numbers or their roles?
No [Points: +0]

Will you require new policy authority for this project?
Yes [Points: +2]

The algorithm used will be a (trade) secret
Yes [Points: +3]

The algorithmic process will be difficult to interpret or to explain
No [Points: +0]

Does the decision pertain to any of the categories below (check all that apply):
Social assistance (ei, disability claims, etc) [Points: +1]

Will the system only be used to assist a decision-maker?
Yes [Points: +1]

Will the system be replacing a decision that would otherwise be made by a human?
No [Points: +0]

Will the system be replacing human decisions that require judgement or discretion?
No [Points: +0]

Is the system used by a different part of the organization than the ones who developed it?
No [Points: +0]

Are the impacts resulting from the decision reversible?
Reversible [Points: +1]

How long will impacts from the decision last?
Impacts can last years [Points: +3]

Please describe why the impacts resulting from the decision are as per selected option above.

Depending on the required intervention for a student, decisions could have long term impacts on residence, schooling, etc.

The impacts that the decision will have on the rights or freedoms of individuals will likely be:
Moderate impact [Points: +2]

Please describe why the impacts resulting from the decision are (as per selected option above).

This system acts as a risk assessor and in cases where an intervention is required, law enforcement or other government agencies may step in for the best interest of the student for long term success.

The impacts that the decision will have on the health and well-being of individuals will likely be:
High impact [Points: +3]

Please describe why the impacts resulting from the decision are (as per selected option above)

For those flagged by this algorithm, the goal is for well-being to be greatly enhanced as the student will be put in a more favorable environment for success.

The impacts that the decision will have on the economic interests of individuals will likely be:
Little to no impact [Points: +1]

Please describe why the impacts resulting from the decision are (as per selected option above)

The risk assessment will flag students in vulnerable situations and intervene but economic interests are not expected to change as part of this.

The impacts that the decision will have on the ongoing sustainability of an environmental ecosystem, will likely be:
Little to no impact [Points: +1]

Please describe why the impacts resulting from the decision are (as per selected option above)

Not applicable to this tool.

Will the Automated Decision System use personal information as input data?
Yes [Points: +4]

Have you verified that the use of personal information is limited to only what is directly related to delivering a program or service?
No [Points: +4]

Is the personal information of individuals being used in a decision-making process that directly affects those individuals?
Yes [Points: +2]

Have you verified if the system is using personal information in a way that is consistent with: (a) the current Personal Information Banks (PIBs) and Privacy Impact Assessments (PIAs) of your programs or (b) planned or implemented modifications to the PIBs or PIAs that take new uses and processes into account?
No [Points: +1]

What is the highest security classification of the input data used by the system? (Select one)
Classified / Confidential [Points: +2]

Who controls the data?
Other Canadian Government (prov/municipal) [Points: +2]

Will the system use data from multiple different sources?
Yes [Points: +4]

Will the system require input data from an Internet- or telephony-connected device? (e.g. Internet of Things, sensor)
No [Points: +0]

Will the system interface with other IT systems?
No [Points: +0]

Who collected the data used for training the system?
Another level of government [Points: +3]

Who collected the input data used by the system?
Another level of government [Points: +3]

Will the system require the analysis of unstructured data to render a recommendation or a decision?
No [Points: 0]

Section 3.2: Mitigation Questions and Answers

Internal Stakeholders (Strategic Policy and Planning, Data Governance, Program Policy, etc.)
Yes [Points: +1]

Which Internal Stakeholders will you be engaging?

Data Governance
Strategic Policy and Planning
Legal Services

External Stakeholders (Civil Society, Academia, Industry, etc.)
Yes [Points: +1]

Which External Stakeholders will you be engaging?

Other (describe)

Please describe

Law enforcement

Will you have documented processes in place to test datasets against biases and other unexpected outcomes? This could include experience in applying frameworks, methods, guidelines or other assessment tools.
Yes [Points: +2]

Will you be making this information publicly available?
No [Points: +0]

Will you be developing a process to document how data quality issues were resolved during the design process?
Yes [Points: +1]

Will you be making this information publicly available?
No [Points: +0]

Will you undertake a Gender Based Analysis Plus of the data?

No [Points: +0]

Will you be making this information publicly available?
No [Points: +0]

Have you assigned accountability in your institution for the design, development, maintenance, and improvement of the system?
Yes [Points: +2]

Will you have a documented process to manage the risk that outdated or unreliable data is used to make an automated decision?
Yes [Points: +2]

Will you be making this information publicly available?
No [Points: +0]

Will the data used for this system be posted on the Open Government Portal?
No [Points: +0]

Will the audit trail identify the authority or delegated authority identified in legislation?
No [Points: +0]

Will the system provide an audit trail that records all the recommendations or decisions made by the system?
Yes [Points: +2]

Will all key decision points be identifiable in the audit trail?
Yes [Points: +2]

Will all key decision points within the automated system's logic be linked to the relevant legislation, policy or procedure?
No [Points: +0]

Will you maintain a log detailing all of the changes made to the model and the system?
Yes [Points: +2]

Will the audit trail clearly set out all decision points made by the system?
Yes [Points: +1]

Could the audit trail generated by the system be used to help generate a notification of the decision (including a statement of reasons or other notification) where required?
No [Points: +0]

Will the audit trail identify precisely which version of the system was used for each decision it supports?
Yes [Points: +2]

Will the audit trail show who the authorized decision-maker is?
No [Points: +0]

Will the system be able to produce reasons for its decisions or recommendations when required?
Yes [Points: +2]

Will there be a process in place to grant, monitor, and revoke access permission to the system?
No [Points: +0]

Will there be a mechanism to capture feedback by users of the system?
No [Points: +0]

Will there be a recourse process planned or established for clients that wish to challenge the decision?
Yes [Points: +2]

Will the system enable human override of system decisions?
Yes [Points: +2]

Will there be a process in place to log the instances when overrides were performed?
Yes [Points: +1]

Will the audit trail include change control processes to record modifications to the system's operation or performance?
Yes [Points: +2]

Will you be preparing a concept case to the Government of Canada Enterprise Architecture Review Board?
No [Points: +0]

If your system involves the use of personal information, will you undertake or have you undertaken a Privacy Impact Assessment, or updated an existing one?
Yes [Points: +1]

Will you design and build security and privacy into your systems from the concept stage of the project?
Yes [Points: +1]

Will information be used within a closed system (i.e. no connections to the Internet, Intranet or any other system)?
Yes [Points: +1]

If the sharing of personal information is involved, has an agreement or arrangement with appropriate safeguards been established?
No [Points: +0]

Student Risk Prevention Algorithm
Class Concepts: Transparency, Power, Auditing

MEMORANDUM

TO: Joe Principal
Imaginary High School

FROM: Adam Bakopolus
Data Scientist, Imaginary High School

DATE: November 14th, 2021

RE: Canadian Algorithm Impact Assessment

As our school district further discusses rolling out the Student Risk Prevention Algorithm as a means of preemptively identifying students at-risk of arrest or in vulnerable situations, I'm writing to ensure that we've considered all ethical issues. As school and community leaders, we are accountable to our student body and must take all necessary actions to ensure that the algorithm is developed appropriately while mitigating the effect felt by our students. To this end, I have recently completed the Canadian Algorithm Impact Assessment (AIA), and the results are attached. The AIA effectively evaluates a particular algorithm and assesses both the impact a system will have on its "target" as well as how this impact itself is mitigated through documentation, auditing, transparency requirements, etc. I found this tool to be extremely valuable, and I recommend that its findings be reviewed in upcoming discussions.

Many of the AIA's questions centered around transparency and the ability to understand the algorithm's findings:

"Will the system be able to produce reasons for its decisions or recommendations when required?"
"The algorithm used will be a (trade) secret [Yes or No]"
"The algorithmic process will be difficult to interpret or to explain [Yes or No]"

The students flagged by the Risk Prevention Algorithm and the code used to make this determination will not be made public. However, these questions effectively highlight "the limits of the transparency ideal".[1] Even if auditing in the form of algorithm transparency was allowed, oftentimes technical expertise is a barrier to full understanding of an algorithm's process.[2] Additionally, while we do expect the Risk Prevention Algorithm to be on the more straightforward side to interpret (a risk score is assigned based on school attendance, household income, etc.), the AIA makes clear that the lack of transparency into the algorithm's "black box" must be accounted for. Fortunately,

---

[1] Ananny, Mike, and Kate Crawford. "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability." *New Media & Society*, vol. 20, no. 3, 2016, pp. 973–89. *Sage Journals*, https://journals-sagepub-com.proxy.lib.umich.edu/doi/full/10.1177/1461444816676645.

[2] Sandvig, Christian, et all. *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*. International Communication Association, 22 May 2014, http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf.

we are developing a thorough process that allows individuals to challenge an output risk score, but this itself, as will be seen next, results in its own challenges.

The AIA included a question that touched upon the importance of power, or the disproportionate way in which individuals are affected by a system or process due perhaps to their situation or social standing[3]:

"Will there be a recourse process planned or established for clients that wish to challenge the decision?"

This Risk Prevention Algorithm has the potential to significantly impact a flagged student, as the information can and will be made available to law enforcement and social services. While we do plan to have a recourse process established for challenges, this question does highlight its criticality. It was also an important reminder that this challenge process is regressive in regards to power. Students from low-income households, for example, may not have the same level of time and resources higher-income students are afforded due to their societal standing. This AIA question stresses that, alongside this algorithm's rollout, we institute a review process that accounts for these discrepancies.

As I hope can be seen through these example questions, the AIA is a powerful tool in assessing how the Risk Prevention Algorithm will impact students and how effectively we as a school have established guardrails to ensure that the impact can be mitigated as much as possible. With the AIA's focus on transparency, auditing, and power differentials, I strongly recommend that this assessment be considered in future meetings discussing the rollout and development of the Risk Prevention Algorithm. With this algorithm having the potential for long-standing significant impacts to a vulnerable student population, we must hold ourselves accountable and ensure that the process is fair to all affected parties.

[3] Sandvig, Christian. "Johnny Cash's keyword: "Power" (a.k.a. Is this regressive or progressive?)." *Coursera: SIADS 503: Data Science Ethics*, Oct. 2021, https://www.coursera.org/learn/siads503/lecture/IcvoQ/johnny-cashs-keyword-power-a-k-a-is-this-regressive-or-progressive.

References:

Ananny, Mike, and Kate Crawford. “Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability.” *New Media & Society*, vol. 20, no. 3, 2016, pp. 973–89. *Sage Journals*, https://journals-sagepub-com.proxy.lib.umich.edu/doi/full/10.1177/1461444816676645.

Sandvig, Christian, et all. *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*. International Communication Association, 22 May 2014, http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf.

Sandvig, Christian. “Johnny Cash's keyword: "Power" (a.k.a. Is this regressive or progressive?).” *Coursera: SIADS 503: Data Science Ethics*, Oct. 2021, https://www.coursera.org/learn/siads503/lecture/IcvoQ/johnny-cashs-keyword-power-a-k-a-is-this-regressive-or-progressive.