AI transparency in a real-world context: what we can learn from past examples of algorithmic and statistical decision-making

McKay, Margaret H.

Proceedings of the 35th Canadian Conference on Artificial Intelligence, 2022-05-27

Canadian AI 2022
30 May–3 June 2022, Toronto

AI Transparency in a Real-World Context: What we can learn from past examples of algorithmic and statistical decision-making

Margaret H. McKay[†,*]
[†] National Research Council of Canada, Osgoode Hall Law School, York University

Abstract

Public discussion about transparency for AI-enabled decisions tends to focus on the challenge of AI explainability. However, there are additional real-world factors which can hamper individuals seeking to understand or challenge decisions impacting them, even when the AI or algorithm is entirely explainable.

Although AI enabled decision tools are relatively new, algorithmic and statistical decision tools are not. This paper examines past efforts by individuals to access algorithms, statistical models, and data used in making decisions which impacted them. The results of those attempts are considered in light of public expectations for transparency of AI-enabled decision tools, as well as current and developing guidance. Legal changes will be needed if governments wish to meet citizen expectations for real-world transparency of AI-enabled decision systems. In the meantime, there are opportunities for AI experts and others to protect the potential for greater transparency through open data, open source licensing, and engagement in policy development.

Keywords: transparency, artificial intelligence, access, law, confidentiality

1. Introduction

The explainability of AI enabled tools is growing, thanks to the efforts of many highly skilled AI experts. Meanwhile, the ability of this work to make a difference to individuals who challenge AI-enabled decisions in the real world remains an open question. This paper considers these questions of real world access and transparency.

Although AI enabled decision tools are relatively new, algorithmic and statistical decision tools are not. The experiences of those who have attempted to challenge decisions based on that earlier generation of tools provide insights to the hurdles facing future challengers of AI-enabled decisions. The results have implications for democracy and human rights.

Transparency has technical, legal, and policy aspects. Countless artificial intelligence ("AI") researchers have taken on the technical challenges of making AI-enabled tools more explainable. Even so, the policies and legal rules applicable in institutions and their jurisdictions will determine the extent to which affected individuals are able to gain insight into the reasons behind these decisions and the processes by which they were made.

Questions about transparency have existed for at least as long as citizens have attempted to hold decision-makers accountable. This paper examines past decisions and current Canadian guidance relating to statistical and automated decision-making ("ADM") to identify the standards set for transparency, non-technical challenges to citizen-level transparency, and questions and opportunities for the future, in a Canadian context.

[*] margaret.mckay@nrc-cnrc.gc.ca

1.1. **Elements of Transparency**

Transparency of decisions can exist at various levels, including transparency in rationale, and transparency in process [1]. AI transparency has been defined to include both interpretability and justifiability [2]. The Supreme Court of Canada has stated "Transparency and accountability are vital to the democratic process" [3].

This paper approaches these issues from the perspective of an average citizen who has been impacted by a decision and wants to challenge it. To this person, transparency is shorthand for their ability to access and challenge all factors, information, and processes which shaped that decision. Thus, transparency is a necessary element in ensuring accountability of decision-makers in their exercise of power.

Seen this way, transparency-type concerns exist at multiple levels. Specifically, the citizen is concerned with transparency in both the mechanics of the decision and in the bases and assumptions which informed it. This paper focuses on these institutional transparency related considerations, including, but not limited to AI transparency. (In contrast, "explainability" is used herein to describe the narrower technical question of why an algorithm produced a particular result from a given input.)

For example, for a citizen attempting to challenge a specific decision, transparency around that decision and its result may include: the availability and fairness of challenge processes; algorithmic explainability and design considerations; and, the ability to assess the appropriateness and accuracy of the assumptions and data underlying the decision at issue. This includes the information used to train the algorithm, information used in the specific decision process, as well as any information produced from that process. Validation data, and particularly its suitability to particular situations, may also be crucial.

The concept of "privacy" is also central. At a functional level it is the laws of privacy and access which provide individuals with the right to know what information institutions hold about them, oblige institutions to ensure information they use is reasonably accurate, and give individuals the right to seek correction of inaccuracies. Privacy rights therefore contribute to transparency and ultimately support accountability.

1.2. **Limits and Expectations of Voluntary Transparency**

Many institutions provide general information on the factors which go into their decisions. However, even where transparency is possible as a technical matter, few institutions voluntarily reveal the details of their decision processes. Reasons for this include: administrative convenience; concerns about trade secrets and obligations to the suppliers of the tools; the risk of being overwhelmed with requests; and, fears that individuals will use the detailed information to game the system to get the decision they want. Where information is not shared voluntarily, individuals must use a formal process of some kind, whether through the courts or freedom of information laws, to seek information on the factors behind the particular decision which affected them.

In 2021, the Province of Ontario consulted residents regarding their expectations for AI-enabled decision tools. Key transparency-related concerns expressed included: a right to plain language; a right to contest decisions and get human review, or request a non-automated decision process; awareness of how data for decision-making is collected; independent review and auditing of AI tools and their use; means to verify that the government is following its own rules; and, addressing bias [4]. To date, decisions relating to transparency in statistical and algorithmic decision systems have failed to meet several of these expectations, especially in relation to independent review and audits, and meaningful rights to contest decisions.

This paper begins in Part I by examining the nature and historical outcomes of formal processes launched by those seeking to understand statistical or algorithmic decisions affecting them. In Part II we consider Canadian and international requirements and proposals and their potential impacts on transparency. Finally, Part III discusses how public and institutional conceptions of reasonable transparency may differ, what this could mean for policy development in this area, and the opportunity and role for AI experts and average citizens in enhancing real-world transparency.

2. Part I: Historical outcomes from attempts to access reasons or processes for statistical or algorithmic decisions

While AI enabled decision tools are comparatively new, algorithmic decision tools of various forms have existed and been challenged for a long time. Examples include statistical decision tools of various types, machine implemented assessments (measurements) with individual impacts, and guidance manuals which effectively require human decision makers to apply criteria mechanically to reach a conclusion. Canadian tribunal decisions on past cases give insights on how they are likely to treat similar requests relating to AI-enabled decision tools, unless there are changes to the law.

2.1. Current processes used in seeking transparency

Individuals seeking information on how a decision affecting them was made will normally need to justify their request. The two primary legal justifications for seeking access to information held by another party are: (a) freedom of information and privacy rights created by legislation; and, (b) legal entitlements to discovery and disclosure in the context of court proceedings [5–7].

Freedom of information (also known as access) laws generally apply only to information held by governmental organizations [8, 9]. This process starts when the individual submits an application for access to the institution holding the information. The institution must respond within set time limits. Applicants who are dissatisfied with the institution's response can ask the appropriate information and privacy commissioner to look into the matter. In some cases, this will result in the matter coming to a specialized tribunal for decision. (Most of the orders cited later in this section are decisions of tribunals of this kind.)

Privacy laws generally provide focused rights in respect of one's own personal information. Such laws generally oblige organizations to ensure that personal information they hold is as accurate, complete and up-to-date as necessary for the purposes for which it is to be used. They also give individuals the right to access personal information held about them, to challenge its accuracy, and to have it amended when appropriate [10, 11]. Most companies and public-sector organizations are subject to at least one privacy law in their day-to-day operations. The requirement to ensure that data is sufficiently accurate for its intended use appears to include derived information based on analytical processes, as well as the suitability of training data used for the system [12].

In contrast, legal entitlements to discovery and disclosure in the context of court proceedings apply to most proceedings regardless of the nature of the organization holding the information. The scope of discovery and disclosure rights depend on the nature of the proceeding, the allegations being made, the relationship of the custodian of the material to the court proceedings, and other factors.

2.2. Freedom of Information

Freedom of Information ("FoI", also known as "access to information") legislation exists to enable transparency in the day-to-day operations of governmental bodies, and forms a

foundational element for democracy [3, 13]. Unfortunately, FoI legislation does not apply to software or code in most Canadian jurisdictions. (See for example [14–16].) In Ontario, some decisions from the Office of the Information and Privacy Commissioner (OIPC) suggest that access to software can be ordered through FoI [17].

The historical exclusion of software from FoI in many jurisdictions is understandable. At the time these statutes were drafted, software was primarily seen as a neutral tool. It is unlikely that legislative drafters envisioned software as something adding substance to government information holdings. (In one case, the Court analogized software to a camera used to produce a film [16].) Certainly, they would not have predicted software becoming capable of generating new information about an individual. Today it is clear that software, and particularly AI tools, can create new information. Thus, access for the purposes of review can be essential to understanding related decisions.

Most automated decision making ("ADM")-related data and general documentation are theoretically in scope for FoI requests, even in jurisdictions where software itself is not. Requests for disclosure of such material are subject to exemptions from disclosure. Such exemptions can prevent the individual from accessing information necessary to challenge a decision. For example, individuals may be blocked from access where the institution can demonstrate that the documents are confidential information, are used in law enforcement or investigations (including bylaws), or are published or soon-to-be-published information [11, 18].

Example: Ontario Property Value Assessment Decisions

In Ontario (where software is considered to be in FoI scope), a relevant body of decisions relates to FoI requests made to the Municipal Property Assessment Corporation ("MPAC"). MPAC is a not-for-profit corporation jointly owned by Ontario municipalities. It is responsible for producing the property valuations upon which municipal property tax assessments are based [19]. MPAC staff carry out sophisticated statistical processes, involving the creation of models for various regions.

Model Access - Confidentiality:

There are several stages to these analyses, beginning with data collection, geographical model specification, and model calibration. The result of these processes is a syntax file capable of generating outputs based on sales data [20]. The syntax file is used to produce a model record comprising statistical command files and output data. The models themselves embody the decision-making process, and there is no separate record which sets out the specific equations used. MPAC has been able to demonstrate that the model records should be protected from disclosure as trade secrets or confidential information, despite having been developed according to generally accepted practices. Thus, individuals who wish to challenge their property value assessments cannot get access to the statistical models and precise factors used.

Data Access – Revenue-driven Public Access as a Barrier:

FoI legislation permits institutions to refuse to allow individuals to use the freedom of information process to access information which is available to the public through other means [11, 21]. The terms of that other access can be expensive and this approach has long been of concern [22]. In an AI and big data context, these concerns mount because meaningful access may require access to thousands of individually priced records, each of which formed part of a training or validation set for the AI tool in question. The fees charged by MPAC for data used to build property models are far in excess of what the allowable FoI fee would be. While in theory a fee could be so high that it amounts to a denial of access, both in British Columbia and Ontario, tribunals have been reluctant to make such findings, with access fees of $30,000 and $5,184,000, respectively, being found acceptable [21, 23, 24]. This exemption has made it cost-prohibitive to access sufficient data to challenge MPAC assessments.

What about Public Interest?

In some situations, tribunals have the opportunity to consider whether the public interest in a disclosure outweighs the reasons given for denying it. In the case of MPAC property value assessments, the Ontario tribunal has recognized that there is some public interest in understanding how properties are assessed. However, this interest has not been interpreted to extend to the ability to actually verify a particular analysis [25]. Adjudicators have considered unverifiable MPAC web page information on the assessment process as sufficient to satisfy the public interest. As a result, those seeking the ability to verify a particular decision are not characterized as advancing a public interest, only their private interests. When such a private interest is balanced against the government interest in MPAC revenue, MPAC's economic interests have prevailed, and the individual has been denied access [25].

In some instances, concerns have been raised that software developers or vendors have claimed confidentiality in more than they should, not acknowledging their use of open source or non-proprietary portions of code and design approaches. In the absence of evidence regarding the code in question, such claims can be difficult to refute [26].

Law Enforcement and Investigations

Information regarding investigations and law enforcement may need to be controlled to ensure the success of the investigation. On the other hand, if investigations and law enforcement activities are unfair in their nature or their application to specific populations, public transparency may be the best approach to end injustice.

FoI laws generally exclude law enforcement and investigation information from disclosure. The scope provided to "law enforcement" is very broad. Many decisions which impact individuals, civilly or criminally, can be seen as relating to the investigation or enforcement of some kind of law. For example, audit practices and internal review decision guidance have been found to be law enforcement, both in the context of federal tax audits and property tax reviews [18, 27]. This potentially enables the exemption of broad areas of governmental activity from disclosure. In light of increasing concerns about algorithmic policing tools and the the data on which they are trained, a mechanism to enable independent expert fairness and bias reviews is needed.

2.3. Court-related discovery and disclosure processes

Individuals involved in court proceedings usually have a right to obtain access to documents and other material relating to the case which is held by the other party [5–7]. In order to get access to information considered confidential, or information held by outsiders to the court case (even if, like the police in a criminal case they are somewhat aligned with the other party), one must first demonstrate that the material will probably provide evidence relevant to the case. This can be difficult to prove prior to getting access to the material.

Blood Alcohol Testing

It is a worth considering the approach which has evolved for driver blood alcohol testing. Although these "Breathalyzer" readings are merely sensor outputs, with no AI or algorithmic component, they do represent a machine assessment which is very difficult to challenge. Breath alcohol testing is therefore important to consider as an example of an approach which could someday be put in place for challenges to AI-enabled decisions as well.

Historically, it was common for accused to obtain access to information on breath alcohol test device maintenance and calibration/validation data, even though this is held by the police (who are not a party to the court proceeding). This access enabled defendants to raise questions about whether the device was functioning properly. This frequently led to long court challenges which were expensive and inconvenient for the government. In the face of this, Parliament amended the Criminal Code [28], the result of which is a presumption of device accuracy, subject to only a single-concentration test validation by a qualified police

technician [6, 29, 30]. The Supreme Court of Canada has affirmed that there is no longer an assumption that calibration/validation information will be relevant (and therefore available) to a defendant (despite the fact that the single concentration validated will often be different from the concentration "blown" by the accused individual). Thus, a defendant who wishes to access this information must first prove that that information which they have never seen is likely to support their claim that the device was inaccurate.

There is a subtle line between, one the one hand, enforcing acceptance of the results of well-conducted science, and on the other hand, bald claims of superior knowledge which will not tolerate outside examination. Distinguishing these two situations may not always be easy. Rigorous and independent review processes, and transparent sharing of these results can provide good evidence that the conclusions are based on strong science.

In the case of blood alcohol testing, the Court's position was based on work by the Alcohol Test Committee which stated that a single point calibration was adequate to verify device function and accuracy at any level tested. This committee stands as the Canadian authority on blood alcohol testing. Of its eight members, five are from the Royal Canadian Mounted Police, and the other three are government employees in law enforcement-related scientific roles [29, 31].

2.4. **Data Accuracy - Glimmers of Hope**

In both a governmental and a private-sector company context, institutions using personal information are generally obliged to ensure that it is accurate enough for its intended use [10, 11, 32]. A 2018 decision in the context of an application for parole suggests that this can include a requirement for reasonable accuracy or appropriateness of the training data used in statistical tools informing decisions about an individual [12]. In this case, the assessment process for parole applications included the use of statistically-based tests to estimate the risk posed by the offender if released. The tests in use were well known and had been the subject of peer-reviewed papers. This enabled the individual affected by the decision, Mr. Ewert, to demonstrate that the accuracy of the assessment system was vulnerable to errors when applied to other cultural groups and that it had not been validated for his group (Métis). This was found to violate the requirement for accurate information [12].

2.5. **Private Sector Use of AI Tools on Critical Communications Infrastructure**

The Canadian Radio-Television and Telecommunications Commission ("CRTC") recently addressed questions of transparency in relation to the private-sector deployment of an AI-enabled decision system [33]. The situation involved an application by Bell Canada to make a pilot AI-enabled call-blocking mechanism permanent. The system was intended to block fraudulent and scam voice calls using the Bell network, regardless of whether or not they had a beginning or end-point on that network or were just transiting through. The system uses AI to identify anomalies in telecommunications voice traffic, flagging calls for review and potential blocking. Despite efforts by challengers to access system information, Bell Canada sought and was granted protection from disclosure of proprietary information related to the system, as well as non-proprietary information which, if released, would assist bad actors in circumventing the system.

Some telephone service providers who use Bell infrastructure filed requests for limitations on the period of approval, as well as for a study of the use of machine learning as a tool for call-blocking, with the goal of developing a regulatory framework. A request was also submitted requesting asking the CRTC to require Bell Canada to carry out an Algorithmic Impact assessment. All these requests were denied. Despite the status of telecommunications networks as critical national infrastructure, they are privately owned and not subject to the federal Directive on Automated Decision-making (discussed below). The call-blocking tool

was being generically deployed across the network and was not an AI service being provided specifically to the federal government. The CRTC decided that neither an algorithmic impact assessment, nor a regulatory framework were needed, and in December 2021, they permitted Bell Canada to make the system permanent, subject to the continuation of routine reporting on known false positives.

3. Part II: Current and proposed Canadian guidelines - Implications for Individuals

3.1. The Canadian Situation

Federal Formal requirements governing the adoption and use of AI tools are under discussion in Canada. At present, only the Government of Canada's "Directive on Automated Decision-making" (the "Directive") is in force [34]. Despite having only limited application, providing no means for an individual citizen to seek its enforcement, and being only an internal government policy, the Directive is the most mature Canadian guidance currently in place, and one of few examples worldwide.

The Directive applies only to ADM systems involved in Government of Canada administrative decisions (specific decisions affecting legal rights, privileges, or interests). Moreover, its application is limited to services where the intended client is external to the Government of Canada and is being served or using a Government of Canada service. This would appear to include applicants for federal benefits and licences, but to exclude internal human resources-type decisions. It is also not clear that it would apply to decisions with direct impact on citizens as a class or group, where there is no individual service or application, for example planning decisions regarding infrastructure needs such as water treatment or bridge repair in areas of federal jurisdiction. In terms of enforceability, the ability of the Treasury Board to enforce such policies is linked to its budgetary controls.

The Directive establishes a requirement for an "Algorithmic Impact Assessment" and peer review of ADM systems proportionate to a risk score determined by the department proposing to adopt them. Unlike freedom of information enquiries, such reviews can include confidential access to proprietary elements. Ongoing monitoring, testing, and quality checks are also required. To the extent that there might be custom source code owned by the government involved in the ADM, this is to be released, subject to various exceptions. Thus, the Directive offers some safeguards in relation to the assessment of some AI tools proposed for some uses by the Canadian government. It does not apply to private organizations, or to other levels of government, and individuals have no way to enforce it, or to use it to challenge a decision impacting them.

Both the federal Competition Act [35] and the Ontario Class Proceedings Act [36] prohibit the use of statistical evidence against an individual without an opportunity to cross-examine all those who supervised the preparation of the information. These requirements support fairness for those whose interests are impacted by complex data-driven statistical assessments. ADM systems are at least equally data-driven and complex; yet, up to this point no equivalent protections have been given to individuals subject to them.

Ontario Ontario is developing an AI framework [4]. Consultations on guidelines for transparency and ethical use are ongoing [37]. Broad public feedback has been received, reflecting concerns to ensure that residents can challenge decisions, and obtain transparency in relation to the algorithmic tools themselves. To its credit, Government of Ontario's draft Transparency Guidelines propose strong transparency requirements, including the possibility for meaningful access to algorithmic technologies for external researchers and auditors, even in respect of proprietary technologies. While it is important to note that these are only guidelines, if they were adopted as law within Ontario they would rank among the

most advanced AI transparency requirements in the world at this time. If they are not incorporated into law, existing requirements including exceptions to freedom of information and "relevance" enquiries for access to materials held by third parties will allow institutions to prevent access or disclosure in many cases, regardless of what the guidelines recommend. Moreover, as a practical matter public institutions are under constant pressure to seek the lowest possible price for the goods and services they procure. AI transparency requirements increase risk for vendors and will therefore drive up some costs. Thus, unless there is law requiring institutions to follow these guidelines, it may be very difficult for them to do so even if that would be their preference.

The most pressing concern identified in the Ontario Consultation was accountability, including: individual rights to address potential bias in the AI tool; the right to contest decisions; the right to opt out of algorithmic processes; and, the right to plain language. These citizen concerns will not be adequately addressed unless the current guidelines or something like them are passed into law.

The CIO Strategy Council of Canada [38] has also created a proprietary voluntary national standard on "Ethical design and use of automated decision systems" [39]. This standard is useful to the extent that it clearly intends to apply to a broad range of organizations, both public and private. It states that AI ethics should be considered a compliance matter. Likely reflecting the commercial interests of some contributing stakeholders, the guidelines suggest only internal reviews, and provide only a cursory note regarding the importance of having an appeals and escalation process for negatively impacted persons. Commercial users wishing to adopt the standard are obliged to pay.

3.2. **The International Situation**

The World Economic Forum [40], in collaboration with the UK Government Office for AI [41] and other stakeholders have developed guidelines for government AI procurement [42]. This document draws on a range of sources, and is broadly consistent with both the Canadian Directive [34], and the Alan Turing Institute Report [2]. These guidelines encourage an open-by-default approach to government AI-enabled processes, with narrow exceptions when justified. This approach would provide greater transparency than is currently prescribed in most parts of Canada. It will be important to monitor the extent to which the scope of permitted exceptions undercuts transparency in practice.

In the United States, a senate bill has been reintroduced which, if passed, would create the "Algorithmic Assessment Act" [43]. The proposed law would require initial and annual impact assessments of automated decision systems used for critical decisions by large businesses, under the supervision of the Federal Trade Commission. The proposed U.S. law would apply to the organization deploying the technology, and not necessarily to the one selling it. It establishes requirements for these assessments and includes reporting requirements. Summary versions of the reports would be made publicly available by the Federal Trade Commission. It also provides for the creation of a Bureau of Technology within the Federal Trade Commission. This would serve as an expert body to advise the Commission and provide technical assistance in relation to enforcement activities.

The European Union is consulting on a general regulation for AI products offered for sale in the E.U. [44]. The objectives of such rules would include avoiding AI market fragmentation within Europe by establishing minimum EU standards, and ensuring that AI systems placed on the market in the EU are safe and trustworthy. The proposed EU regulation would make it mandatory to disclose the use of AI systems when they are interacting with humans, and would require the creation and use of AI regulatory sandboxes as part of governmental regulatory oversight for AI systems intended for sale in the EU. Also required would be risk-proportionate post-market monitoring, information sharing, and market surveillance for

high-risk AI systems. Where necessary, the authorities would also have access to the source code of the AI system. Market surveillance authorities would be entitled to full access to all data and documentation related to their activities, including training, validation, and testing datasets. Administrative fines of up to €30M or 6 percent of total worldwide annual turnover would be possible in cases of serious non-compliance.

It is clear that many jurisdictions are interested in regulating aspects of AI-enabled decision system sales and use. Whereas Canada was a leader in this area when the federal Directive was introduced in 2019, we are now falling behind as other jurisdictions push ahead with proposals having much broader potential impact.

4. Part III: Understandings of "transparency": implications for policy and a role for AI experts

A review of existing and developing guidance documents reveals what appears to be a significant gap between the average citizen's understanding of AI transparency and what is being offered by governments. While this gap is not new, it does raise the question of where the bar for transparency should be set in relation to AI-enabled decision tools.

Institutions are not consistently open to sharing the details of their decision-making, regardless of whether the tools are AI enabled or simply a departmental manual. At one level this is understandable: expert individuals and teams need to be able to do their work efficiently, without the obligation to explain approaches which frequently involve complex expert knowledge and the delicate balancing of diverse demands. Complete openness could require a radical change to many operational models, with huge implications for resourcing. At the same time, individuals seriously impacted by a decision have a legitimate interest in knowing how that specific decision was made and what factors went into it.

In considering options to address the gaps between citizen expectations and historical practice on ADM transparency, one must not lose sight of the legitimacy of certain claims to limit disclosure. Full assessment of such claims requires expertise in both the technology and the law. At the same time, processes which require individuals to pay expensive technical and legal experts will leave the most vulnerable populations unprotected.

A tribunal with the authority and expertise to confidentially review materials that institutions do not want to reveal to individuals could produce better, more informed decisions. For example, the tribunal might order disclosure of parts of the code, data, or design information in question, or perhaps could even carry out its own assessment of the risk of a particular kind of bias in an algorithm. Public funding for technical and legal research capacity within these tribunals would reduce individual financial barriers to access, increasing accessibility for vulnerable populations and overall effectiveness.

Thus, in a world of excellent transparency, adoption and use of AI-enabled tools would involve at least three factors, applied in an objective and risk-based manner:

(1) Technical explainability of systems (where possible)
(2) Requirements with respect to independent reviews of systems and related data and open publication of the resulting reports
(3) Independent expert tribunals with the authority and resourcing to investigate and address individual concerns, including remedies and penalties

Approaches like the one proposed in Europe come with costs, including slower product releases, increased regulatory burden, and the possibility that some tools may never be introduced into the jurisdiction. Ultimately, each society will need to balance the risks and costs of various approaches, as well as their obligations to protect the rights of individuals and groups.

Governments in Canada are not likely to achieve an ideal transparency system in the near term. Thus, individuals seeking to challenge AI-enabled decisions can expect to encounter the same obstacles previously encountered in relation to algorithmic and statistical tools.

There are ways in which both experts and non-experts can lay the groundwork for challenges to AI-enabled decisions during what is likely to be a long transitional period. As we have seen, both privacy and freedom of information laws provide some limited means for individuals to challenge incorrect data used in or produced by AI-enabled systems. These laws also provide access points for the gathering of relevant information. For example, the viewpoints and opinions of government officials conducting assessments of AI tools will normally be disclosable (except where they reveal the actual information supplied by the third party). Additionally, information on how the government carries out its approval processes is generally disclosable![45]. Taken together, such information could provide insights into the systems in use and could support further challenges, raising awareness and public pressure to improve shortcomings in some tools.

Contracting practices are another key area for attention, particularly in respect of contract terms on confidentiality. In one case, system error reports prepared in part by city staff were found to contain confidential information of the software vendor![46]. While confidentiality claims will continue to occur, careful and informed contacting can help prevent over-reach by vendors. To the extent that claims of confidentiality can be limited to only key software components, it will be easier for individuals to access information and challenge decisions.

The AI community is well positioned to contribute to increases in policy and legal transparency, in addition to their ongoing good work on technical aspects of transparency. Areas for further engagement include: (a) increasing the availability, findability, and identifiability of non-proprietary code elements and data to aid in restricting the scope of vendor claims that their products are entirely confidential; and, (b) publicly reporting shortcomings of AI-enabled tools and their impacts in AI-enabled systems to aid those seeking to challenge decisions. Additionally, as open source licences evolve, augmentation of licence terms to strengthen requirements for disclosure of specific open source components would help to chip away at over-reaching claims of confidentiality.

Conclusions

AI-enabled tools may be new, but questions of transparency in algorithmic decision-making are not. Historically, individuals affected by algorithmic decision-making have not experienced the level of transparency that many appear to expect in relation to new AI-enabled systems. Legal requirements supporting transparency are needed, as are expert tribunals with the authority and resources to hold institutions accountable for their use of AI-enabled tools, and to fully investigate and make findings on complaints by individuals.

Ongoing contributions by AI researchers and others in the area of explainability, open source software, and open data repositories will support progress in this area. Going forward, a continuation of these efforts, and a sharpening of procurement and open source software licence terms to require more detailed disclosure regarding components used will continue to advance transparency. Furthermore, governments will be consulting and proposing approaches to this issue. All those interested in AI should ensure that their voices are heard as these discussions continue.

Acknowledgements

The author wishes to thank Svetlana Kiritchenko, Joel Martin, and Andrea Slane for comments and advice on this paper and the topic in general. This work is aligned with

the National Research Council of Canada AI for Logistics Program focus on ethics and responsible data. Nothing contained herein is legal advice. Opinions expressed are personal to the author and may not be shared by her employer.

References

[1] K. de Fine Licht and J. de Fine Licht. "Artificial intelligence, transparency, and public decision-making". In: *AI Society* 35 (2020), pp. 917 –926. URL: https://link.springer.com/article/10.1007/s00146-020-00960-w.

[2] D. Leslie. *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*. The Alan Turing Institute, 2019. URL: https://doi.org/10.5281/zenodo.3240529.

[3] *Dagg v Canada (Minister of Finance)*. 1997. URL: https://www.canlii.org/en/ca/scc/doc/1997/1997canlii358/1997canlii358.html.

[4] *Consultation: Ontario's Trustworthy Artificial Intelligence (AI) Framework*. 2021. URL: https://www.ontario.ca/page/ontarios-trustworthy-artificial-intelligence-ai-framework-consultations.

[5] *Rules of Civil Procedure, R.R.O. 1990 Reg. 194*. URL: https://www.ontario.ca/laws/regulation/900194.

[6] *R. v. Gubbins [2018] 3 SCR 35*. 2018. URL: https://www.canlii.org/en/ca/scc/doc/2018/2018scc44/2018scc44.html.

[7] *R. v. Khan [2004] Ontario Judgements 3811*. 2004.

[8] *Freedom of Information and Protection of Privacy Act, R.S.O. 1990, c. F.31*. URL: https://www.ontario.ca/laws/statute/90f31?search=freedom+of+information.

[9] *Access to Information Act, R.S.C.1985, c. A-1*. URL: https://www.laws-lois.justice.gc.ca/eng/acts/A-1/index.html.

[10] *Personal Information Protection and Electronic Documents Act, SC 2000, c 5*. 2000. URL: https://www.laws-lois.justice.gc.ca/eng/acts/P-8.6/index.html.

[11] *Municipal Freedom Of Information and Privacy Protection Act R.S.O. c.M-56*. 1990. URL: https://www.ontario.ca/laws/statute/90m56?search=municipal+freedom+of+information.

[12] *R.v.Ewert*. 2018. URL: https://www.canlii.org/en/ca/scc/doc/2018/2018scc30/2018scc30.html?autocompleteStr=Ewert%20v%20Canada&autocompletePos=1.

[13] *Ontario (Public Safety and Security) v Criminal Lawyers' Association, [2010]*. 2010. URL: https://www.canlii.org/en/ca/scc/doc/2010/2010scc23/2010scc23.html.

[14] *Freedom Of Information and Privacy Protection Act, R.S.B.C. c.165*. 1996. URL: https://www.bclaws.gov.bc.ca/civix/content/complete/statreg/1198514681/96165/?xsl=/templates/browse.xsl.

[15] *Freedom Of Information and Privacy Protection Act, R.S.N.S. c.5*. 1993. URL: https://nslegislature.ca/sites/default/files/legc/statutes/freedom%20of%20information%20and%20protection%20of%20privacy.pdf.

[16] *Yeager v Canada (Correctional Service) (CA), [2003] 3 FC 107*. 2003. URL: Ontario(PublicSafetyandSecurity)vCriminalLawyers'Association,[2010].

[17] *Order MO-3701, Municipal Property Assessment Corporation*. 2018. URL: https://decisions.ipc.on.ca/ipc-cipvp/orders/en/item/356237/index.do?q=MO-3701.

[18] *Order F21-22 City of Vancouver*. 2021. URL: https://www.oipc.bc.ca/orders/3545.

[19] *Order MO-1881, Municipal Property Assessment Corporation*. 2004. URL: https://decisions.ipc.on.ca/ipc-cipvp/orders/en/item/132306/index.do?q=MO-1881.

[20] *Order MO-1564, Municipal Property Assessment Corporation*. 2002. URL: https://decisions.ipc.on.ca/ipc-cipvp/orders/en/item/131600/index.do?q=MO-1564.

[21] *Order MO-1573, Niagara Regional Police Services Board*. 2002. URL: https://decisions.ipc.on.ca/ipc-cipvp/orders/en/item/131614/index.do?q=MO-1573.

[22] *Order P-496, Ontario Securities Commission*. 1993. URL: https://decisions.ipc.on.ca/ipc-cipvp/orders/en/item/128488/index.do?q=P-496.

[23] *Order 91-1996 Inquiry Re: A decision by the Ministry of Environment, Lands and Parks to withhold Digital Map Data from the Western Canada Wilderness Committee (WCWC)*. 1996. URL: https://www.oipc.bc.ca/orders/240.

[24] *Order 1948 Municipal Property Assessment Corporation)*. 2005. URL: https://decisions.ipc.on.ca/ipc-cipvp/orders/en/item/132365/index.do?q=MO-1948.

[25] *Order 2412 Municipal Property Assessment Corporation)*. 2009. URL: https://decisions.ipc.on.ca/ipc-cipvp/orders/en/item/133239/index.do?q=MO-2412.

[26] S. Katyal. "The Paradox of Source Code Secrecy". In: *Cornell Law Review* 104(5) (2019), p. 1183. URL: https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?article=4803&context=clr.

[27] *3412229 Canada Inc. v. Canada (Revenue Agency)*. URL: https://decisions.fct-cf.gc.ca/fc-cf/decisions/en/item/490285/index.do.

[28] *Criminal Code, RSC 1985, c C-46*. URL: https://laws.justice.gc.ca/eng/acts/C-46/index.html.

[29] *Evidential breath testing for alcohol, Parliament, the science and the courts (Part 2)*. 2020.

[30] *Canadian Artificial Intelligence Conference*. 2020. URL: https://commons.allard.ubc.ca/fac_pubs/536/.

[31] URL: https://www.csfs.ca/what-we-do/csfs-committees/atc-alcohol-test-committee.

[32] *Corrections and Conditional Release Act, SC 1992, c 20, s.24*. URL: https://laws-lois.justice.gc.ca/eng/acts/C-44.6/.

[33] *Compliance and Enforcement and Telecom Decision [2021] CRTC 2021-403*. 2021. URL: https://crtc.gc.ca/eng/archive/2021/2021-403.htm.

[34] *"Directive on Automated Decision-making", Directive under the Policy on Service and Digital*. 2019. URL: https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.

[35] *Competition Act, R.S.C. 1985*. URL: https://www.laws-lois.justice.gc.ca/eng/acts/C-34/index.html.

[36] *Class Proceedings Act, S.O. 1992, c. 6 s.23(6)*. URL: https://www.ontario.ca/laws/statute/92c06?search=class+proceedings+act.

[37] *Ontario, "Transparency Guidelines"*. 2021. URL: https://github.com/ongov/Transparency-Guidelines.

[38] URL: https://ciostrategycouncil.com/.

[39] *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, CAN/CIOSC 101:2019*. CIO Strategy Council, Standards Council of Canada, 2020. URL: https://ciostrategycouncil.com/standards/101_2019/#gf_12.

[40] URL: https://www.weforum.org/.

[41] URL: https://www.gov.uk/government/organisations/office-for-artificial-intelligence.

[42] *Guidelines for AI Procurement*. World Economic Forum, 2019. URL: https://www3.weforum.org/docs/WEF_Guidelines_for_AI_Procurement.pdf.

[43] *Algorithmic Accountability Act of 2022*. 2022. URL: https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202022%20Bill%20Text.pdf.

[44] *Proposal for a Regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts (SEC(2021) 167 final - SWD(2021) 84 final - SWD(2021) 85 final)*. European Commission, 2021. URL: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52021PC0206.

[45] *AstraZeneca Canada Inc v Canada (Health)*, URL: https://decisions.fct-cf.gc.ca/fc-cf/decisions/en/item/45310/index.do?q=AstraZeneca+Canada+Inc+c+Canada+%28Sant%C3%A9%29%2C+%5B2005%5D+.

[46] *Order MO-3151, City of Toronto*. 2015. URL: https://decisions.ipc.on.ca/ipc-cipvp/orders/en/item/134573/index.do?q=MO-3151.