# *Case Study: Developing guidance for the responsible use of artificial intelligence in decision-making at Immigration, Refugees and Citizenship Canada*

Authors: Patrick McEvenue, Director, Strategic Policy and Planning, Immigration, Refugees and Citizenship Canada, and Michelle Mann, General Counsel, Justice Canada. [Company address]

**Acknowledgement and Notes**

The authors gratefully acknowledge the assistance of our teams in writing this paper specifically: Gregg Blakely (Assistant Director, Digital Policy), Marjie Brown (Policy Analyst); Robert Hersch (Legal Counsel), Anna Lillicrap (Legal Counsel); Benoit Richer (Policy Analyst); and in editing: Baylie van Gulik (Paralegal) and Christopher Thompson (Legal Counsel).

Note section 3 was authored by Michelle Mann, section 5 was written by both authors and the balance of the paper by Patrick McEvenue. Views expressed in relation to the law are personal and do not constitute the official views of Justice Canada.

This paper was originally prepared for the Law Society of Ontario program Special Lectures 2019: Innovation, Technology, and the Practice of Law, held on November 21-22, 2019.

# *Case Study: Developing guidance for the responsible use of artificial intelligence in decision-making at Immigration, Refugees and Citizenship Canada*

**ABSTRACT**

In recent years, Immigration, Refugees and Citizenship Canada (IRCC) has been exploring the use of artificial intelligence (AI) as a means to enhance its day-to-day business operations. Thus far, efforts have largely been focused on supporting administrative decision-making within targeted, low-complexity business lines. In parallel, the Department together with Justice Canada has undertaken significant work to analyze and address the associated legal, ethical, technological and operational considerations.

To reflect the intertwined nature of these considerations, this paper is a joint effort by legal counsel and policy specialists. It begins with a description of IRCC's pilot project to apply advanced analytics to Temporary Resident Visa applications – the Department's primary use case involving AI and automation. The authors then explore various aspects of public law that are a form of accountability of government institutions in this domain. This is a non-exhaustive and high level exploration of legal issues related to administrative, human rights and privacy law. A series of tips are proposed to assist legal and AI practitioners in aligning AI tools with key legal principles and obligations. Examples of some emerging legal issues are also very briefly explored.

In addition to aligning with the applicable legal frameworks, practitioners need to be aware of several fundamental policy considerations when developing and using data-driven technologies. As such, the paper reviews the core questions that IRCC set out to answer in developing internal policy guidance on the responsible use of AI. Informed by extensive research, external engagement and the Department's own experience with use cases, IRCC has developed a *Policy Playbook on Automated Decision Support*. A set of guiding principles from the Playbook are included as an appendix.

## 1. CONTEXT

Immigration, Refugees and Citizenship Canada (IRCC) facilitates the entry of migrants who wish to come to Canada temporarily, manages the selection and integration of immigrants and refugees who intend to stay in Canada permanently, grants citizenship and issues passports to eligible citizens. These functions are performed with the aim of maximizing migrants' contribution to the country, maintaining Canada's humanitarian tradition, and protecting the health, safety and security of Canadians.

Administrative decision-making[1] is at the core of IRCC's domestic and international operations – from the processing of permanent and temporary resident visas, electronic travel authorizations and work or study permits, to the granting of asylum and the issuance of secure and internationally recognized Canadian citizenship and travel documents. IRCC has long operated a stable and respected decision-making model, but mass scale international mobility and technological change are challenging its sustainability. Canada's conventional approach requires meaningful change, and artificial intelligence (AI) presents one viable avenue for IRCC to improve its day-to-day business and respond to growing pressures in the immigration domain.

[1] Administrative decision-making refers to decisions made by authorized officials of an institution (such as a board, tribunal, commission, government department, agency or Ministry), where the outcome affects legal rights, privileges or interests.

The impetus for change is driven, first, by tremendous growth. Given rising numbers of temporary resident applications, along with increases in permanent resident and citizenship business lines, IRCC's processing networks are continuously required to stretch and expand their capacity. Innovation is needed to maintain reasonable processing times and manage backlogs. Change is also driven by the increasing availability of Big Data and tools to process it, by rising client expectations for quick and easy digital interaction, and by government commitments to program integrity, innovation and service excellence. As a result, the Department is gradually shifting away from traditional decision-making – where one officer, located in the same place as the applicant, reviews all information and renders a decision – to a revamped business model in which decisions can be shared across processing networks, harnessing capacity and expertise globally. With these new capabilities comes an opportunity to modernize operations, including the potential to augment both the consistency and integrity of decision-making.

As IRCC pursues solutions involving AI tools – many of which could become necessary to sustain timely, high-quality operations – it must tread carefully and consider a wide array of ramifications. How to safeguard clients' privacy in a data-rich environment, how to enhance decision-making procedures without affecting procedural fairness and sound decision-making, and how to use data analytics to target risks while guarding against algorithmic bias and discrimination are just a few examples of issues that are front and center for the Department.

In seeking to apply AI to immigration decision-making, IRCC has undertaken extensive analysis to understand the full spectrum of risks and benefits from legal, policy, technological and operational standpoints. A diverse, multi-disciplinary team of experts has been engaged in problem-solving – no longer compartmentalizing issues and tackling each separately, but instead working on the whole thing at once: the technology and data science, the business process, the legal issues, the policy questions. Working in a multi-disciplinary team has added value by increasing the expertise of all, enhancing and sharpening the advice provided, and ultimately ensuring that each discipline could influence the solution. At its core, this team has been working on how to adapt a model of responsible migration management to a new set of tools – ones that are in full evolution both inside and outside of the Department.

## 2. IRCC'S PRIMARY USE CASE

IRCC has been pursuing a few different projects involving AI – including a chatbot to answer client enquiries on social media and the possible creation of a tool powered by machine learning to assist in legal research and litigation management (in partnership with the Justice Canada) – but the most instructive use case is the application of advanced analytics and automation to decision-making on client applications. Administrative decision-making is at the core of what IRCC does as a department, and the majority of the time and effort spent analyzing the risks and opportunities of AI has been devoted to this activity.

In 2018, IRCC introduced a pilot project to test the use of AI to support application processing. The goal of the pilot, still ongoing, is to automate a portion of the Temporary Resident Visa (TRV) business process, focusing on online applications from China and India.

In deciding which type of client applications to test new approaches on, IRCC considered a host of factors, including:

(1) Complexity: Certain applications involve routine examination of information, while other cases involve more nuanced judgment calls;

(2) Stakes: Temporary residence applications are generally lower stakes for applicants than Permanent Residence or Citizenship applications, and certainly lower stakes than Humanitarian and Compassionate cases or Pre-Removal Risk Assessments[2];

(3) Biggest need: Canada has seen (and projects) enormous year-over-year growth in the volume of applications in temporary resident lines of business; and

(4) Best data: The larger the traveller movement the larger the corpus of cases to learn from, and where usage of online applications is widespread, data is more structured. IRCC looked for large caseloads where data integrity was high.

With a business line selected, IRCC data scientists began designing a system to identify very straightforward cases – applications so clear-cut they do not require the expertise of an experienced visa officer. The idea is to clear the easy approvals off the desks of officers, so that they can focus their time and energy on cases that require more scrutiny. Starting with China, IRCC's project team looked at previous decisions to find the factors that led to near unanimous approval by visa officers. They did so by developing and training a model – using a supervised machine learning approach – with data from more than 180,000 TRV applications with final decisions from the Beijing office between January 2015 and December 2017.[3] The model recognizes key factors that are at play in decision-making – more precisely, at the eligibility step of the TRV assessment process – to identify low-complexity cases. These became model rules that were reviewed and approved for use by a senior officer. It also includes a feedback loop, to the extent possible: data about approved visitors who go on to violate the terms of their stay is fed back to the system so that it can adjust its findings to reflect a changing environment.

In operation at the visa office, the system triages incoming applications – first by "disqualifying" any that trigger a key complexity indicator (such as travelling with a minor) – and then by sorting the remainder into three "bins" based on complexity. The system auto-approves the files in the low complexity categories for eligibility requirements, allowing a visa officer to review these files exclusively on the basis of admissibility. More complex cases are triaged to officers for both eligibility and admissibility determinations, according to traditional procedures and without further processing by the predictive model. There are no auto-refusals.

[2] It is worth noting that IRCC is not considering the application of AI assisted automated decision-making across all business lines. There are no plans to automate decisions in high-stakes, discretionary areas like Humanitarian & Compassionate cases or Pre-Removal Risk Assessments. Even where decision-makers are helped by algorithms, humans remain at the centre of IRCC's decision-making models.

[3] No external data was used to create this model.

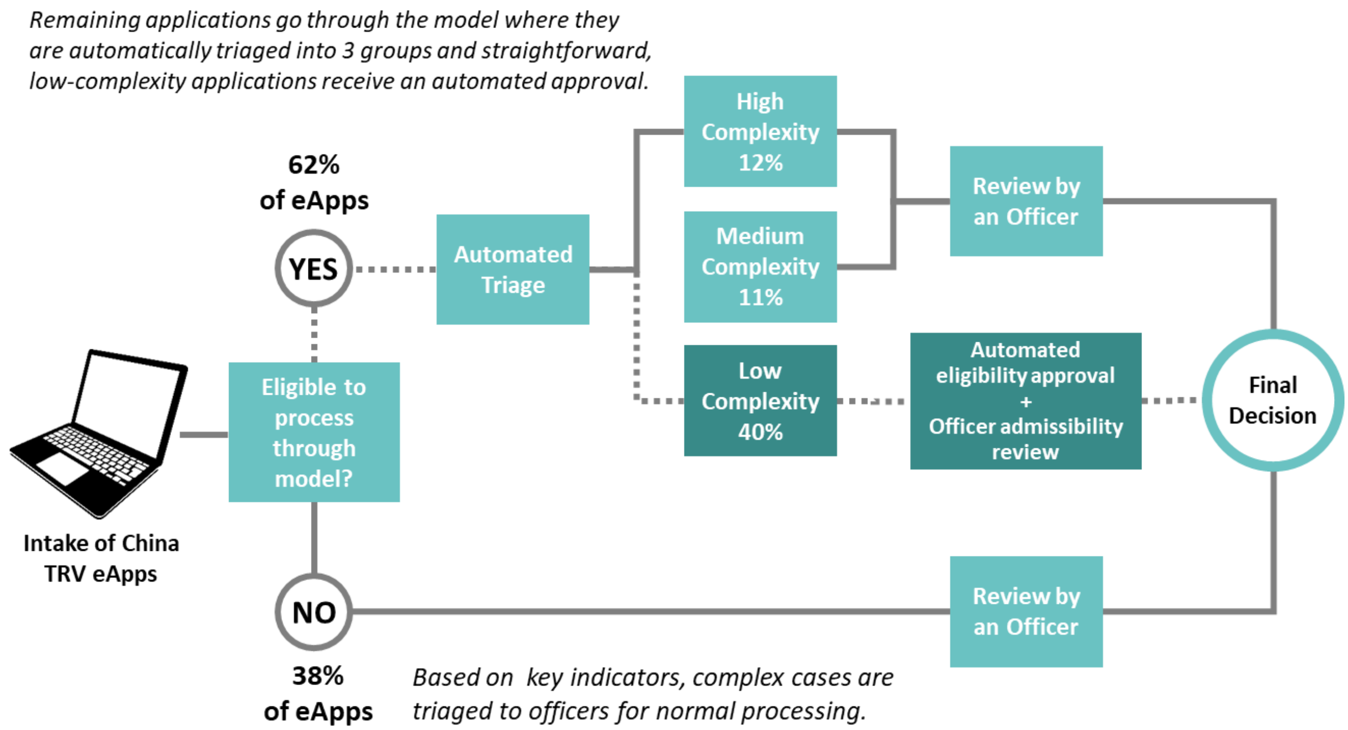


**Figure 1, Sample Process Flow**

As part of the pilot, IRCC established an extremely rigorous quality assurance process, in which a random sample of the cases processed by the automated system (10%) is reviewed by officers each day. The quality assurance regime has demonstrated that officers agree with the model's decisions 99.5% of the time, providing a clear indication that the model's outputs are remarkably consistent with human decision-making. The incidence of adverse outcomes has been very low for cases processed by the model, which further confirms the model's accuracy.

At the time of writing, the model had triaged over 160,000 cases, of which more than 55,000 were identified as very straightforward and given an automatic positive eligibility determination without further officer intervention. These cases were then screened for admissibility by officers in the traditional manner.

## 3. A FOCUS ON SOME LEGAL ISSUES THAT ARISE IN THE CONTEXT OF GOVERNMENT USE OF AI IN ADMINISTRATIVE DECISION-MAKING

The law provides a framework for ensuring that legal authority is exercised accountably, in the manner that Parliament intended it to be. Each area of public law highlighted below (administrative, human rights and privacy law) speaks to government responsibility.

> "By virtue of the rule of law principle, all exercises of public authority must find their source in law."[4]

[4] *Dunsmuir v New Brunswick*, 2008 SCC 9 at para 28.

The rule of law helps to ensure that the law is administered in a transparent and predictable manner and in doing so also provides a form of accountability to those who are affected by administrative decisions. Administrative decisions, such as whether to approve an immigration application, have often been highlighted as key sites where individuals interact with the state and can provide vital opportunities for affected individuals to have their voice heard and have their rights and interests recognized and duly taken into consideration by the state.[5] The same fundamental legal principles and obligations of public law continue to apply when AI is used as part of the administrative decision-making process. Ideally, the use of advanced technologies should complement and even augment respect for the rule of law and the promotion of dignity through a respectful interaction between individuals and the state.

The law has always had to adapt to new situations and technologies with creativity and flexibility through incremental growth of the common law and dialogue between courts and legislatures. The same is true with AI. In time we will see how the law responds to AI and whether issues will emerge where there are gaps, uncertainties or undesirable results. It remains to be seen whether courts can adequately deal with the issues, whether government policy thinkers and legislatures must respond or solutions will be found, in some instances, in technology and related standards. Ultimately there will be no panacea approach or solution. Our role is to anticipate and understand these gaps, uncertainties and to inform and advance solutions.

## Legal Framework for AI Tools in Government

What follows is a brief synopsis of some issues raised in public law – specifically administrative, human rights, and privacy law. Highlights in each of these areas are noted along with some novel questions and tips for practitioners to apply and align with the law. Finally, select emerging issues are very briefly explored when current legal frameworks are applied to this evolving technology.

### 3.1 Administrative Law

In this context, administrative law accountability focuses on three essential concepts: legality, fairness and reasonableness of decision-making.[6] The following sub-sections outline some of the ways that administrative law may play out when decisions are made with the assistance of AI. It will highlight how the core administrative law principles and obligations continue to apply, and may develop incrementally where needed, to respond to the use of this technology.

The interpretation and application of administrative law principles in the context of AI tools in automation will often depend upon the juxtaposition of the technological solution and human decision-makers in the decision-making process. While there are many variations possible[7], this can be envisioned as a spectrum of AI and human involvement (see figure 2). In general, the greater the implication and impact of the AI solution in the decision-making process, the more consideration should be given to ensuring principles of administrative law continue to be respected.

---

[5] See Geneviève Cartier, "Administrative Discretion and the Spirit of Legality: From Theory to Practice" (2009) 24:3 CJLS 313. See also Geneviève Cartier, "Administrative Discretion as Dialogue: A Response to John Willis (Or: From Theology to Secularization)" (2005) 55:3 UTLJ 629.

[6] Donald J.M Brown & the Honourable John M Evans, *Judicial Review of Administrative Action in Canada*, (Toronto: Carswell, 2009) generally describes vires, process, and merits.

[7] See CIO Strategy Council, *CAN/CIOSC 101:2019, Ethical design and use of automated decision systems*, (National Standard of Canada), (Ottawa: CIOSC, 2019) at 11 [*CAN/CIOSC 101:2019],* online (pdf): <https://ciostrategycouncil.com/members-login/can-ciosc-101-2019_2019-09-12_en/> accessed 10 October 2019 for references to "human in the loop", "human on the loop", and "human out of the loop".

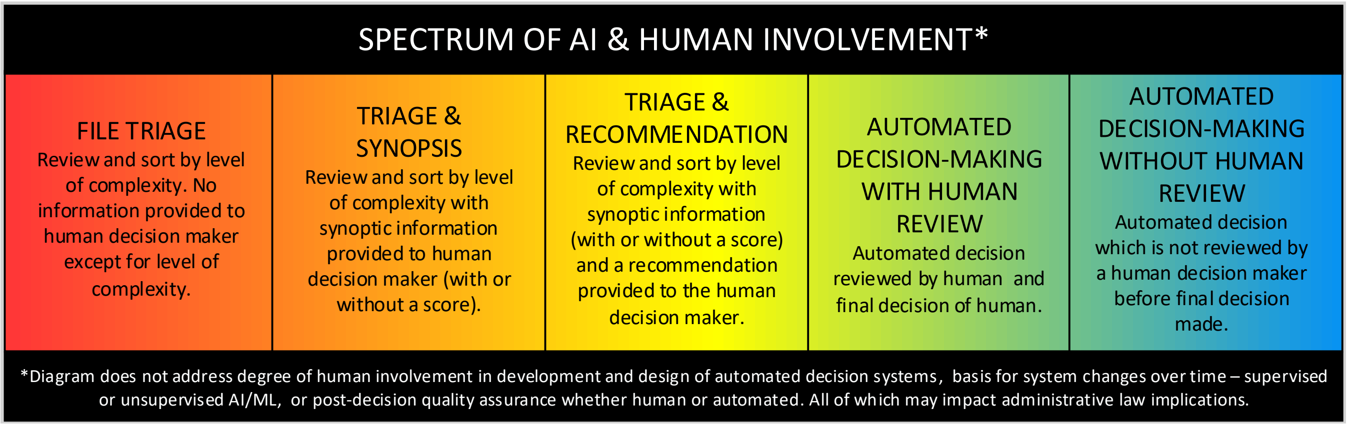


**Figure 2, Spectrum of AI & Human Involvement**

### *3.1.1 Ensuring a Fair Process*

Administrative decisions affecting the "rights, privileges or interests" of individuals are subject to a duty of procedural fairness.[8] Although the ability of non-Canadians to enter Canada is considered a privilege, these decisions must still be made fairly.[9] An essential element of procedural fairness is meaningful participation. This means that the individual knows the case to be met and has an opportunity to respond to it. The content of procedural fairness varies and is context specific having regard to a number of factors including the nature of decision, the statutory scheme, importance of the decision to the individual, legitimate expectation, and procedure of the decision-maker.[10]

The extent to which the use of technology, specifically AI assisted decision-making, may change what is legally required to meet the duty of procedural fairness will depend – among other things – on the role that AI plays in the decision-making process. As mentioned above, the more AI tools affect the decision-making process (see Figure 2 above), the more care must be taken to assessing in this case, existing procedures, to ensure that they continue to provide individuals with a meaningful opportunity to participate.

In the context of immigration applications, the legislative criteria on which applications will be evaluated are generally known to the applicant in advance. Applicants are usually expected to satisfy the decision-maker that they meet the criteria through the information submitted in their application. These are two essential building blocks – the legislative criteria and the factual information used. Further follow-up with applicants may only be needed where there are additional concerns or information (i.e. extrinsic evidence/not known to the applicant) and the applicant should be provided with a fair opportunity to respond.[11]

---

[8] *Canada (AG) v Mavi*, 2011 SCC 30, [2011] 2 SCR 504.
[9] *Chiau v Canada (Minister of Citizenship and Immigration)*, [2001] 2 FC 297, 195 DLR (4th) 422; *Phuti v Canada (Minister of Citizenship and Immigration)*, [2000] FCJ No 1233 at para 24.
[10] *Baker v Canada (Minister of Citizenship and Immigration),* [1999] 2 SCR 817, [1999] SCJ No 39 at paras 23-27.
[11] *Sharma v Canada (Public Safety and Emergency Preparedness)*, 2016 FCA 319, [2017] 3 FCR 492 (the Federal Court of Appeal found that an inadmissibility report did not have to be disclosed to the individual prior to its finalisation reiterating the legal test frequently used by courts to determine whether evidence needs to be disclosed before a decision is made. The Court distinguished documents "whose essential purpose is to list relevant information from the file" from those that are "more akin to advocacy tools" at para 33).

The same appears generally to be true when AI is used to assist in the decision-making process and the same essential building blocks are used in the AI solution. However, what precisely a meaningful opportunity to participate entails depends on the context, including the nature of the decision being made, how AI was used in the processing of their application and what information was relied on by the AI system or the human decision-maker.

### *3.1.2 What is Defensible Decision-Making?*

Decisions do not have to be perfect, but have to be reasonable with a sufficiently transparent and intelligible justification. Depending on the context, decisions within a range of acceptable outcomes are defensible.[12] This includes, for example, avoiding undue reliance or fettering discretion[13], avoiding bias, applying the legislative criteria, appropriate consideration of relevant information and cogent reasons. Some of these concepts are briefly explored below.[14]

### *3.1.3 Avoiding Undue or Inappropriate Reliance on AI*

While AI can help to improve the efficiency and quality of decision-making, it is important to ensure that decision-makers understand the AI solution and use it for its intended scope and purpose. For example, where an AI tool is intended to be used to flag certain aspects of an application for a decision-maker to review, the decision-maker should still conduct their own review of the entire file and not treat the flag as determinative of the outcome of the application. Providing a clear, transparent explanation to decision-makers about how the AI solution works, what it reviews, and why it is flagging certain aspects of the file to the decision-maker can help the officer to put these flags in context and use them appropriately. [15]

### *3.1.4 Considering the Appropriate Criteria, Information and Avoiding Bias*

The use of AI does not change the legislative criteria that must be applied to make a particular decision unless it were to be specifically addressed. As such, the information and considerations taken into account by AI tools should include those that are relevant to and tied to those legislative criteria. The use of such systems should also aim to augment decision-making by enabling better decisions. AI tools can be used to improve the overall

---

[12] *Newfoundland and Labrador Nurses' Union v Newfoundland and Labrador (Treasury Board)*, 2011 SCC 62, [2011] 3 SCR 708 at para 16 [*Newfoundland*].

[13] See e.g. *Kanthasamy v Canada (Minister of Citizenship and Immigration)*, 2015 SCC 61, [2015] 3 SCR 909 (long standing body of jurisprudence in immigration law in respect of reasonableness of decision-making and holding that the department cannot dictate rules for a decision-maker to a make his or her decision).

[14] *Newfoundland*, *supra* note 12.

[15] *Dhillon v Canada (AG),* 2016 FC 456 (we have little guidance from Canadian courts in relation to the use of automated systems and fettering of discretion. In a different use of an automated system on the spectrum referred to in Figure 2, the Federal Court found that the Previous Offender Process, an automated system used by the Canadian Border Service Agency at the border (ports of entry), did not fetter officer's discretion as it was not automatically referring individuals to secondary examination at every attempt to enter Canada rather it served to "enhance the efficiency of the examination process at points of entry by automating some of the processes an experienced BSO would follow if they had an opportunity to fully review the history of individuals seeking to enter Canada" and functioned as an "institutional memory of future consistent compliance" at paras 39-40).

quality of decisions, such as by mitigating (individual) cognitive bias[16] by identifying and addressing inconsistencies, gaps and unconscious bias.

Bias in decision-making can occur in a number of ways, most obviously where a particular factor leads a decision-maker to prejudge the outcome. This is a very important consideration in the spectrum of AI and human involvement in decision-making (Figure 2) as information is highlighted by AI systems or additional AI-generated information is provided to the human decision-maker. The human decision-maker must be vigilant to avoid bias and rather use this information to enhance and inform their decision.

In addition, bias in automated decision-making can arise in a variety of ways. Bias may exist in the data used to train the model and seep into the model, perpetuating that bias. Bias can be incorporated through the influence of human bias at various stages of the design and development of the AI system. However, properly applied AI can also assist in mitigating individual human bias, for example by facilitating more consistent and robust decision-making. Leveraging AI itself to combat bias is a welcome and active area of development[17] where we see AI being used to detect and reduce human bias or by turning the technology on itself to proactively detect and compensate for bias in data, AI models and outcomes.[18]

In *Ewert v. Canada,*[19] the Supreme Court of Canada touched on bias (accuracy and completeness of data) when it considered statistical tools used by the Correctional Service of Canada to assess the risk of recidivism. A Métis inmate argued that the statistical tools were based on data gathered from predominantly non-Indigenous populations, and that there was insufficient research to confirm that they were unbiased when applied to Indigenous persons. The Court agreed. Although *Ewert* dealt with traditional statistical tools, the case provides insight into how a court may approach data bias in the development of AI tools.

AI can also be used in a way that helps decision-makers to provide clearer, more intelligible reasons for their decisions, which in turn improves accountability to affected individuals and may facilitate access to judicial review.

It is recommended that practitioners review AI tools in the design phase to flag legal considerations and risks early to influence design of the AI tool and its use. Ongoing review during implementation is also recommended, particularly where changes are being made to design. Participation of legal advisors in a multi-disciplinary team of subject matter experts in technology (including data science) and policy and program administration is ideal and in fact essential. Below are a selection of practical tips (Figure 3) that will help translate administrative law considerations into practice.

---

[16] See Haiyan Zhang et al, "The Role of AI in mitigating bias to enhance diversity and inclusion" (New York: IBM Corporation, 2019), online (pdf): <https://www.ibm.com/downloads/cas/2DZELQ4O> accessed 4 October 2019. See also Osonde Osoba, William Welser IV, "An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence" (California: RAND Corporation, 2017), online (pdf): <https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf> accessed 27 October 2019 [RAND].

[17] Jake Silberg & James Manyika, "Notes from the AI frontier: Tackling bias in AI (and in humans)" (McKinsey Global Institute, 2019) at 7, online (pdf):<https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/MGI-Tackling-bias-in-AI-June-2019.ashx> accessed on 8 October 2019. See also RAND, *supra* note 16 at 22.

[18] Silberg & Manyika, *ibid* at 1.

[19] 2018 SCC 30 (the Court found that Correctional Services Canada (CSC) "continued to rely on the results they produced in making decisions about offenders without inquiring into their validity with respect to Indigenous offenders" concluding that "[t]his was a breach of the CSC's obligation under s.24 (1) of the CCRA to take all reasonable steps to ensure that any information about an offender that it uses is as accurate as possible" at para 80).

| Practical Tips - A Starting Point for Administrative Law in Government AI Tools |
|---|
| • Ensure adequate information for and training of decision-makers to ensure an understanding of the AI tool and the respective role of the AI tool and the human decision-maker.<br>• Assess training data for bias.<br>• Ensure data used is accurate, up to date, reliable and as complete as possible.<br>• Evaluate whether the model/rules can either detect and correct, or, potentially embed or exacerbate bias.<br>• Evaluate the model/rules for inclusion and consideration of legislative criteria and relevant facts.<br>• Assess the technology and its use by decision-makers to meet transparency and explainability.<br>• Assess intelligibility and cogency of decisions and reasons available.<br>• Monitor the operation of the AI system on an ongoing basis to detect and avoid administrative law risks. |

**Figure 3, Practical Tips – A Starting Point for Administrative Law in Government AI Tools**

## 3.2 Human Rights Law

Human rights law provides another important lens for thinking through the legal considerations that arise in the context of using AI to assist in administrative decision-making.

Human rights law is relevant both because the nature of certain administrative decisions which may have an effect on human rights and because the use of AI tools may give rise to additional human rights considerations.

### *3.2.1 Using AI in the Context of Decisions that May Have an Effect on Human Rights*

The Supreme Court of Canada has confirmed that a robust administrative law approach is an appropriate lens through which to think about the effect of administrative decisions on *Canadian Charter of Rights and Freedoms'*[20] rights and values.[21] The Court indicated that this involves assessing whether the decision-maker exercised their discretion in a way that took sufficient account of *Charter* values and did so reasonably by proportionately balancing the *Charter* protections at issue, in light of the nature of the decision, the statutory context and the particular facts.[22]

### *3.2.2 Avoiding Discrimination and Preserving Equality*

In addition, the use of AI tools may in certain circumstances give rise to human rights law considerations, even in situations where the administrative decision being made would not ordinarily affect an individual's human rights. Of particular relevance are human rights protections for equality.

[20] *Canadian Charter of Rights and Freedoms*, Part 1 of the *Constitution Act*, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11 [*Charter*].
[21] *Doré v Barreau du Québec*, 2012 SCC 12, [2012] 1 SCR 395.
[22] *Ibid* at paras 32, 57-58.

Government action must not discriminate. In this context, the questions are: does the action create distinctions based on enumerated or analogous grounds and, if so, are the distinctions discriminatory?[23] This may arise due to the adverse effect of a government action on an individual because of his/her particular personal characteristics that people either cannot change (are immutable) or that are changeable only at unacceptable cost to personal identity (constructively immutable).[24] Much has been written about the potential for AI models to identify patterns that will directly or indirectly differentiate between applicants based on what would be considered immutable personal characteristics**.** [25] Enumerated or analogous grounds could be embedded directly or indirectly or by proxy. Indirect may arise where grounds are inadvertently hidden behind other variables used. For example, the address or geographic location information could be a proxy for any number of enumerated or analogous grounds such as national or ethnic origin or membership in a particular religious group. Detailed understanding of the data and context is required to identify otherwise innocuous variables that may be a proxy for, or indirectly correlate to, an identified ground.

Not all differential treatment is discriminatory. It will be in cases where the distinction amounts to perpetuating disadvantage and stereotyping[26] or has "the effect of perpetuating arbitrary disadvantage on the claimant".[27] These guiding principles inform the contextual inquiry into whether there has been discrimination in the substantive sense and must be considered in designing and implementing AI tools.[28]

A key consideration is the correspondence between the grounds on which the claim is based and the actual circumstances of the claimant. Government actions that take into account the claimant's actual circumstances in a way that respects his or her value as a human being are less likely to discriminate.[29]

In many cases, immutable personal characteristics are relevant to the legislative criteria for the decision being made and are appropriate to take into account because they correspond to the applicant's actual circumstances[30]. For example, the legislative criteria for becoming an economic immigrant take into account the applicant's age, but when considered alongside other factors age may be a relevant indicator of such things as the length of time the applicant is likely to remain in the workforce, their ability to adapt to life in Canada, and so forth. AI models may therefore be helpful in identifying and applying factors which are relevant for making good decisions.

---

[23] *Law v Canada (Minister of Employment and Immigration)*, [1999] 1 SCR 497 [*Law*] (the Court described a three-step framework, which it rearticulated in *R v Kapp*, 2008 SCC 41, [2008] 2 SCR 483 [*Kapp*] in two steps: "(1) [d]oes the law create a distinction based on an enumerated or analogous ground? (2) [d]oes the distinction create a disadvantage by perpetuating prejudice or stereotyping?" at para 17). See also *Withler v Canada (AG)*, 2011 SCC 12, [2011] 1 SCR 396 at para 30 [*Withler*]; *Quebec (AG) v A*, 2013 SCC 5, [2013] 1 SCR 61 at paras 324, 418; *Kahkewistahaw First Nation v Taypotat*, 2015 SCC 30, [2015] 2 SCR 548 [*Taypotat*].

[24] See e.g. *Corbiere v Canada*, [1999] 2 SCR 203 at paras 13-15, 58-62; *Withler, supra* note 23.

[25] J. Tashea, "Risk-assessment algorithms challenged in bail, sentencing and parole decisions" (ABA Journal 2017), online: <http://www.abajournal.com/magazine/article/algorithm_bail_sentencing_parole> accessed 15 October 2019. See also RAND, *supra* note 16 at 11.

[26] *Andrews v Law Society of British Columbia*, [1989] 1 SCR 143, 56 DLR (4th) 1; *Kapp*, *supra* note 23 at para 23; *Withler*, *supra* note 23; *Quebec v A*, *supra* note 23.

[27] *Taypotat*, *supra* note 23 at para 16; *Quebec v A*, *supra* note 23 at para 331.

[28] *Quebec v A*, *supra* note 23 at paras 325-331, 418.

[29] See *Lavoie v Canada*, 2002 SCC 23, [2002] 1 SCR 769 at para 44; Wi*nko v British Columbia (Forensic Psychiatric Institute)*, [1999] 2 SCR 625 at para 88; *Nova Scotia (Workers' Compensation Board) v Martin; Nova Scotia (Workers' Compensation Board) v Laseur*, 2003 SCC 54, [2003] 2 SCR 504 at para 93; *Canadian Foundation for Children, Youth and the Law v Canada*, 2004 SCC 4, [2004] 1 SCR 76 at paras 58-62.

[30] *Law, supra* note 23.

As noted above, it has also been recognized that AI models may increase quality of decisions by mitigating individual human decision-maker cognitive bias or discrimination (along with being more comprehensive and consistent). AI systems can be designed to enhance comprehensive analysis and review by, for example, leveraging their capacity to consider a larger number of individual circumstances (variables). AI tools may enhance consistency in decision-making and mitigate cognitive bias and discrimination at the level of the individual human decision-maker[31]. Finally, AI tools should be leveraged to detect and remedy existing bias - a challenge that our technology sector is embracing.[32]

### *3.2.3 Human Rights and AI Tools - Practical Tips for the AI Practitioner*

When using AI to assist with administrative decision-making, potential human rights implications should be assessed in both the design and implementation phases. Reviewing for potential discriminatory impacts and ensuring due regard to *Charter* values in a way that is consistent with the administrative law approach is essential as discussed above. A number of steps can be taken to achieve this.

As noted above, legal practitioners working in multi-disciplinary teams should review the AI tools early to identify legal implications and risks in order to influence design decisions and engage in ongoing review, where needed. Below are a selection of practical tips (see Figure 4) or steps that will help embed respect for human rights law into the AI solution.

**Practical Tips - A Starting Point for Human Rights in Government AI Tools**

- Evaluate *training data* suitability for the intended purpose including based on (developing) best practice and industry standards including for reliability as a representative sample.
- Evaluate the *algorithm* and its suitability for the intended purpose including based on best practice and industry standards (if any).
- Identify whether enumerated or analogous grounds are relied on in the AI tool in variables embedded in the model/rules, directly or indirectly by proxy. Detailed understanding of the data and context is required.
- Before deployment, review of the impacts of reliance in the model/rules on identified grounds and if negative, a substantive assessment of whether discriminatory. If yes, further analysis of the root cause (algorithm, training data, human influence on model / rule design) and take steps to address.
- Ongoing macro level analysis of model outputs to determine whether any residual discriminatory impacts can be detected.

**Figure 4, Practical Tips – A Starting Point for Human Rights in Government AI Tools**

[31] *Haiyan Zhang et al*, *supra* note 16; *Silberg & Manyika, supra* note 17.

[32] *Silberg & Manyika*, *supra* note 17; On embedding values into autonomous and intelligent systems see also "The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems" 1st ed, (IEEE, 2019), online: < https://standards.ieee.org/industry-connections/ec/autonomous-systems.html> accessed 7 October 2019.

## 3.3 Privacy Law

Privacy law is another framework for ensuring accountability of government for its actions. The *Privacy Act* is designed to be technologically neutral, as are the relevant program-specific legislative frameworks. As with the other domains of law the fundamental privacy principles and obligations on government institutions are unchanged and can be applied to AI tools; however, there are potential impacts and nuances that merit careful consideration.

Awareness is key to ensuring that design and implementation approaches enhance privacy protection and mitigate the potential for privacy intrusion. On a practical level, there are differences as to how personal information is managed in AI tools – as compared to traditional manual approaches. Uncovering these differences is central to the analysis.

Under the *Privacy Act* accountability for government institutions focuses on a number of principles and obligations related to the collection, notice, use, accuracy, disclosure, retention and disposal of personal information.[33] A policy principle that imbues some of these obligations is avoiding undue privacy intrusion by minimising the amount of personal information collected, used or disclosed. [34] Work must be done to address privacy accountability through careful design and responsible use of AI.

### *3.3.1 Ensuring Minimal Collection of Personal Information*

The definition of 'personal information' is broad,[35] and the collection and use of the large amounts of data needed to train algorithms and develop AI models in government institutions implicates *Privacy Act* obligations. In terms of collection authority, where new collections of data that include personal information are anticipated, it will be incumbent on the institutions to ensure that a rationale for the collection is developed that describes an adequately direct[36] and necessary[37] link between the data to be collected and an operating program or activity of the institution.[38] Similarly, where publicly available personal information is collected, a rationale for the collection is also required; however, under the *Privacy Act*, the use and disclosure obligations do not apply to such publicly available information.[39] In the absence of legal obligation, relevant privacy policy principles should nonetheless be considered.

---

[33] *Privacy Act*, RSC 1985, c P-21.

[34] Information and Privacy Commissioner of Ontario, *Big Data Guidelines*, (Canada: Ontario Privacy Commissioner of Ontario, 2017) at 5, online (pdf): <https://www.ipc.on.ca/wp-content/uploads/2017/05/bigdata-guidelines.pdf> accessed 10 October 2019.

[35] *Privacy Act*, *supra* note 33, s 3.

[36] *Ibid,* s 4.

[37] Treasury Board Secretariat of Canada, *Directive on Privacy Practices*, online: <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18309> accessed 1 October 2019.

[38] While government institutions are obligated by the *Directive on Privacy Practices* to limit collections to personal information that is demonstrably necessary for a program or activity, in *Union of Canadian Correctional Officers/Syndicat des Agents Correctionnels du Canada Confédération des Syndicats Nationaux CSN (UCCO-SACC-CSN)* v *Canada (AG)* 2016 FC 1289 at para 141 the *Directive's* standard of necessity was rejected in favour of adopting a new test for collection authority that requires establishing "a direct, immediate relationship with no intermediary between the information collected and the operating programs or activities of the government." The decision of the Federal Court was affirmed in *Canada (Union of Correctional Officers) c Canada (AG)*, 2019 FCA 212.

[39] *Privacy Act, supra note 33,* s 69(2).

### *3.3.2 Evaluating Secondary Use of Personal Information*

Where government institutions have already collected personal information needed to develop AI tools, the institutions must determine whether the information can be used for the purposes of developing and deploying these solutions. Where personal information was not collected for the purposes of developing an AI solution, the institutions will need to determine whether the intended use is permitted by the *Privacy Act*.[40] Institutions may re-use personal information where that use is either with consent of the individual,[41] or it is a consistent use in that there is a sufficiently direct connection between the purpose of the initial collection (the primary purpose) and the purpose of the proposed use (the secondary purpose) such that the individual would reasonably expect their information to be used in that manner.[42]

In many cases, the application of these principles and obligations will be straightforward. However, there are interesting and, some would argue, novel functionalities that merit consideration, for example:

- Is there a material change to the character of use of personal information when it is used as part of large data sets in the AI solution development[43]?
- Does the AI solution learn and generate net "new" and previously unknown information about an individual[44]?
- AI tools generate analysis, new information and outputs that may be expected or unexpected. Where they are unexpected, can the value of the data to be collected and used in the AI tool (i.e. the purpose) be adequately determined prior to the analysis of the data by the tool?[45]

There is some debate as to whether these functionalities as used in AI tools today are novel and represent, from a legal perspective, a material departure from the functionalities of other technologies currently in use.[46] Or, perhaps these considerations are something to watch for on the horizon as more complex solutions are developed. In either case, it is important for AI developers to step back and assess the implications of the functionalities of AI tools with a view to both compliance with the applicable legal framework and importantly both near and long term privacy policy implications.

---

[40] *Privacy Act, supra note 33,* ss 7(a), 7(b).
[41] Consent is not further explored in detail in this paper.
[42] *Bernard v Canada (AG)*, 2014 SCC 13, [2014] 1 SCR 227 at para 31.
[43] See *Big Data Guidelines*, *supra* note 34 at 5 (reference to unintended uses).
[44] The concept of generated information is not entirely new and similar concepts appear in some federal legislation. See e.g. *Privacy Act, supra* note 33, 7(a), 8(2)(a), 9(4), 11(1)(a)(4); *Department of Employment and Social Development Act*, SC 2005, c 34, s 31; *Canada Revenue Agency Act*, SC 1999, c 17, ss 56, 88. See also Jason Millar, "Core Privacy: A Problem for Predictive Data Mining" in Ian Kerr, Val Steeves & Carole Lucock, eds, *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society* (New York: Oxford University Press, 2009) at 110, online (pdf):<https://www.academia.edu/7372357/Core_Privacy_A_Problem_for_Predictive_Data_Mining> accessed 8 October 2019.
[45] Or to put it another way, can the "outputs" of the AI system be adequately anticipated to inform the rationale for the collection and use of the selected data? See also *Big Data Guidelines*, *supra* note 34 at 5.
[46] Treasury Board Secretariat of Canada, Responsible Artificial Intelligence in the Government of Canada (2018), online (pdf): <https://buyandsell.gc.ca/cds/public/2018/06/28/05baa93c08b6f2d3000855170f831066/ABES.PROD.PW__EE.B017.E33657.EBSU000.PDF> accessed 28 October 2019.

### *3.3.3 Ensuring Adequate Notice of Collection and Use of Personal Information*

Transparency is another very important principle and is seen in obligations under the *Privacy Act* [47]and in government policy such as the *Directive on Privacy Practices*[48]. The *Privacy Act* requires that government institutions inform individuals of the purpose for which that information is being collected at the time of its collection.[49] The *Directive on Privacy Practices* is prescriptive and requires a description of the purpose for the collection and also "any uses or disclosures that are consistent with the original purpose"[50]. The Treasury Board of Canada *Directive on Automated Decision Systems* requires various notices and information to be provided depending on the impact level of the AI solution which impact is to be assessed using an assessment tool that is a companion to the Directive.[51] There is therefore an opportunity for government institutions to further mainstream transparency around the use of AI technology by notifying individuals at the time of collection that their personal information may (or will, as appropriate) be used in development of AI models.

Info Source[52], which publishes the description of the personal information banks of each government institution, may also serve the purposes of increasing transparency around use of AI technology. The Personal Information Indexes must include statements such as: the purpose for obtaining (collecting) or compiling personal information in the relevant bank, the consistent uses or disclosures of that information and the retention and disposal standards. Including information in Info Source to reflect the development or use of AI models is another opportunity for government institutions to further mainstream transparency.

### *3.3.4 Assessing Disclosure of Personal Information*

The obligations of government institutions in regards to disclosure of personal information do not change in the context of the development and deployment of AI tools. In many systems that automate processing and decision-making in existing programs, there may be few or no new disclosures of personal information rendering this a neutral aspect of using AI tools.

### *3.3.5 Ensuring Accuracy of Personal Information Used for Administrative Purposes*

With respect to accuracy of information used for an administrative purpose, both traditional and AI model use of personal information require ensuring the accuracy of this personal information[53]. In practice, manual processing may allow for needed verification for any errors in personal information before its use, while AI models may process the data without this double-checking function built-in. Consequently, while the privacy-related obligations remain the same, there may be an additional level of scrutiny required when using AI models since use of inaccurate personal information to develop an AI model may invalidate any findings or decisions

---

[47] *Privacy Act, supra* note 33.
[48] *Directive on Privacy Practices*, *supra* note 37.
[49] *Privacy Act, supra* note 33, s 5(2).
[50] *Directive on Privacy Practices*, *supra* note 37 at 6.2.9.
[51] Treasury Board Secretariat of Canada, *Directive on Automated Decision-Making*, online: <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> accessed 10 October 2019. There are various other transparency and disclosure requirements of this *Directive* that are relevant. However, this section of the paper dealing with the impacts of privacy law, does not address these in detail.
[52] Government of Canada's website describes Info Source as a mechanism which "provides information about the functions, programs, activities and related information holdings of government institutions subject to the *Access to Information Act* and the *Privacy Act*"(2019), online: <https://www.canada.ca/en/impact-assessment-agency/corporate/transparency/info-source.html#toc001> accessed 10 October 2019. See also *Privacy Act*, *supra* note 33, ss 10, 11.
[53] *Privacy Act*, *supra* note 33, s 6(2).

made by that model. Mitigation of this risk can be achieved by careful analysis of the accuracy of the personal information in the data set over the lifecycle of its intended use. Where weaknesses are found in its accuracy, strategies must be developed to address it.

### *3.3.6 Using De-identification to Enhance Privacy*

De-identification – the process of modifying data such that it can no longer be attributed to an identifiable individual and cannot be reengineered to re-identify an individual – is recommended in developing and deploying AI technology because such de-identified data is no longer subject to *Privacy Act* obligations. De-identification can also be very effective in minimising privacy intrusion. However, effective de-identification can be difficult unless the identifying data is completely purged from a data controller's system[54]. Further, de-identification is an evergreen process because the potential for linking information to an identifiable individual can occur as new information is collected or technical means to defeat de-identification become more sophisticated. De-identification may also not be desirable in many instances, for example, where AI is used to process data as part of an administrative process. However, it may be that this is not an all or nothing proposition. Whether de-identification can be used depends on the trajectory of the design/development cycle of an AI solution and careful review of what personal information is in fact necessary.

### *3.3.7 Retaining Records*

Record keeping is another key aspect of accountability of government institutions for actions taken and a foundation of transparency. Other drivers for record keeping related to AI tools are addressed above in the context of administrative law and human rights law; however, under the *Privacy Act* and *Regulations*, where information is used for an administrative purpose (i.e. in a decision-making process), it must be retained for a minimum of two years.[55] Where AI models are being developed or used, it will be important to ensure record-keeping practices meet this requirement. From a more general privacy due diligence perspective, consideration should be given to keeping records in relation to "privacy by design" choices that were made in designing and deploying AI tools.[56]

### *3.3.8 Human Rights and Privacy Considerations Related to Using AI*

The use of AI may give rise to additional privacy considerations particularly in situations where new information is collected to support the development and deployment of AI models or in situations where information generated by AI models reveals more information about an applicant than they would have realized when they provided it to the government.[57] Where there is a proposal to collect or use personal information for

---

[54] See generally Ann Cavoukian & Khaled El Emam, *Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy*, (Canada: Information and Privacy Commissioner of Ontario, 2011) at 4-5, online (pdf):<https://fpf.org/wp-content/uploads/2011/07/Dispelling%20the%20Myths%20Surrounding%20De-identification%20Anonymization%20Remains%20a%20Strong%20Tool%20for%20Protecting%20Privacy.pdf> accessed 8 October 2019. See also Australia National Data Service, ANDS Guide: De-identification (Australia National Data Service, 2018), online:< http://www.ands.org.au/working-with-data/sensitive-data/de-identifying-data> accessed 10 October 2019.

[55] *Privacy Act*, *supra* note 33, s 6(1) and *Privacy Regulations*, SOR/83-508, s 4(1)(a).

[56] See *Directive on Automated Decision-Making*, *supra* note 51 (various requirements depending upon the impact assessment performed).

[57] Office of the Privacy Commissioner of Canada, *2017-2018 Annual Report to Parliament on the Personal Information Protection and Electronic Documents Act and the Privacy Act*, (Gatineau: Office of the Privacy Commissioner of Canada,

development or deployment of an AI model, analysis in relation to s. 8 of the *Charter* will focus on the nature of the personal information sought, and whether there is a reasonable expectation of privacy in relation to this information. There is a reasonable expectation of privacy in information that is of a personal and confidential nature, or that reveals a personal core of biographical information[58] - this includes information which tends to reveal intimate details of the lifestyle and personal choices of the individual.[59] The context in which the personal information is to be used will also be important, however, to ensure that the privacy notion "of control over, access to and use of information"[60] is respected. The aim of this analysis should be to protect informational privacy and prevent intrusions before they happen[61] and, where possible, de-identification of personal information is likely to aid in attaining this goal.

Engaging on s. 8 *Charter* rights should be done early so that, as the AI system is developed, measures to mitigate risks of privacy intrusion can be built in. For example, business records will be less likely to attract a reasonable expectation of privacy, whereas lifestyle, intimate relations, political or religious opinions will generally engage higher privacy interests.[62] Consequently, where developers seek to use personal information as a predictive tool, it will be prudent to consider what kind of personal information is needed to accomplish this task and whether this can be done without using information that is likely to attract a reasonable expectation of privacy. For instance, an AI model that analyses address data from applications in order to predict application volumes so that additional immigration officers can be proactively dispatched to deal with anticipated influxes of applications, is likely to be at low risk of engaging s. 8 concerns.

The Practical Tips below (Figure 5) are the third set of tips to ensure that practitioners can put privacy law considerations into practice through early review of the AI solution with a view to influencing design and implementation decisions.

---

2018) at 53-54, online: <https://www.priv.gc.ca/en/opc-actions-and-decisions/ar_index/201718/ar_201718/> accessed 1 October 2019.

[58] *R v Mills*, [1999] 3 SCR 668 at para 81; *R v Plant*, [1993] 3 SCR 281 at 292-94 [*Plant*].

[59] *Plant*, *ibid* at 293; *R v Tessling*, 2004 SCC 67, [2004] 3 SCR 432 at para 62.

[60] *R v Spencer*, 2014 SCC 43, [2014] 2 SCR 212 at para 40.

[61] *Hunter v Southam Inc*, [1984] 2 SCR 145 at 160 (prevent unjustified searches and seizures in regard to personal information before they happen).

[62] *Thompson Newspapers Ltd v Canada (Director of Investigation and Research, Restrictive Trade Practices Commission)*, [1990] 1 SCR 425 at 517.

| Practical Tips - A Starting Point for Privacy Law in Government AI Tools |
|---|
| • Evaluate data flows to ensure collection, use, and disclosures of personal information meets legal and policy requirements, including conduct of a privacy impact assessment.<br>• Consider functionality differences with AI solutions and whether they impact privacy obligations.<br>• Ensure notification obligations are met in relation to new collections of personal information or new uses of previously collected personal information.<br>• Assess accuracy of personal information in the data set over the lifecycle of its intended use.<br>• Assess the extent to which de-identification can be used.<br>• Evaluate record keeping requirements.<br>• Consider what personal information may be used to accomplish an intended purpose, and use the information with the lowest privacy interest.<br>• Monitor the operation of the system on an ongoing basis to detect and avoid privacy risks. |

**Figure 5, Practical Tips – A Starting Point for Privacy Law in Government AI Tools**

Consideration of privacy implications should be paramount in any government program or activity that seeks to use AI technologies. This is particularly true within the immigration domain due to the inherent vulnerability of many immigration applicants. The notion of implicating privacy considerations at the design phase is mentioned above and is also the hallmark of “Privacy by Design”[63]. It is essential to consider these principles and obligations in the design and ongoing deployment of AI tools to ensure that personal information is being managed appropriately. Using AI technologies does not change the obligations of government institutions, but these technologies do offer an opportunity to interpret and apply privacy law in new and interesting ways.

### *3.3.9 Engaging in the AI Project Lifecycle*

The Practical Tips found above (Figures 3, 4 and 5) suggest some steps to effectively “put the law into practice” on AI projects in relation to those specific areas of law. They are designed to assist practitioners in moving from interpreting and applying the legal frameworks, to identifying the risks and on to aligning with the principles and obligations of the legal frameworks to mitigate risk.

Practitioners and multi-disciplinary expert teams can use the *Engaging in the AI Project Lifecycle* below *(Figure 6)* as a high level snapshot of steps to be taken over the project lifecycle to assess the legal implications at the outset and ensure that legal risks are identified and mitigated. Please note that this is a guide and not exhaustive of areas of the law that might be engaged, nor does it describe all steps to be taken for effective risk management. Note that it does mention other areas of the law not addressed in this paper, specifically commercial issues related to intellectual property.

[63] Ann Cavoukian, *Privacy by Design: The 7 Foundational Principles*,(Information and Privacy Commissioner of Ontario, 2009).

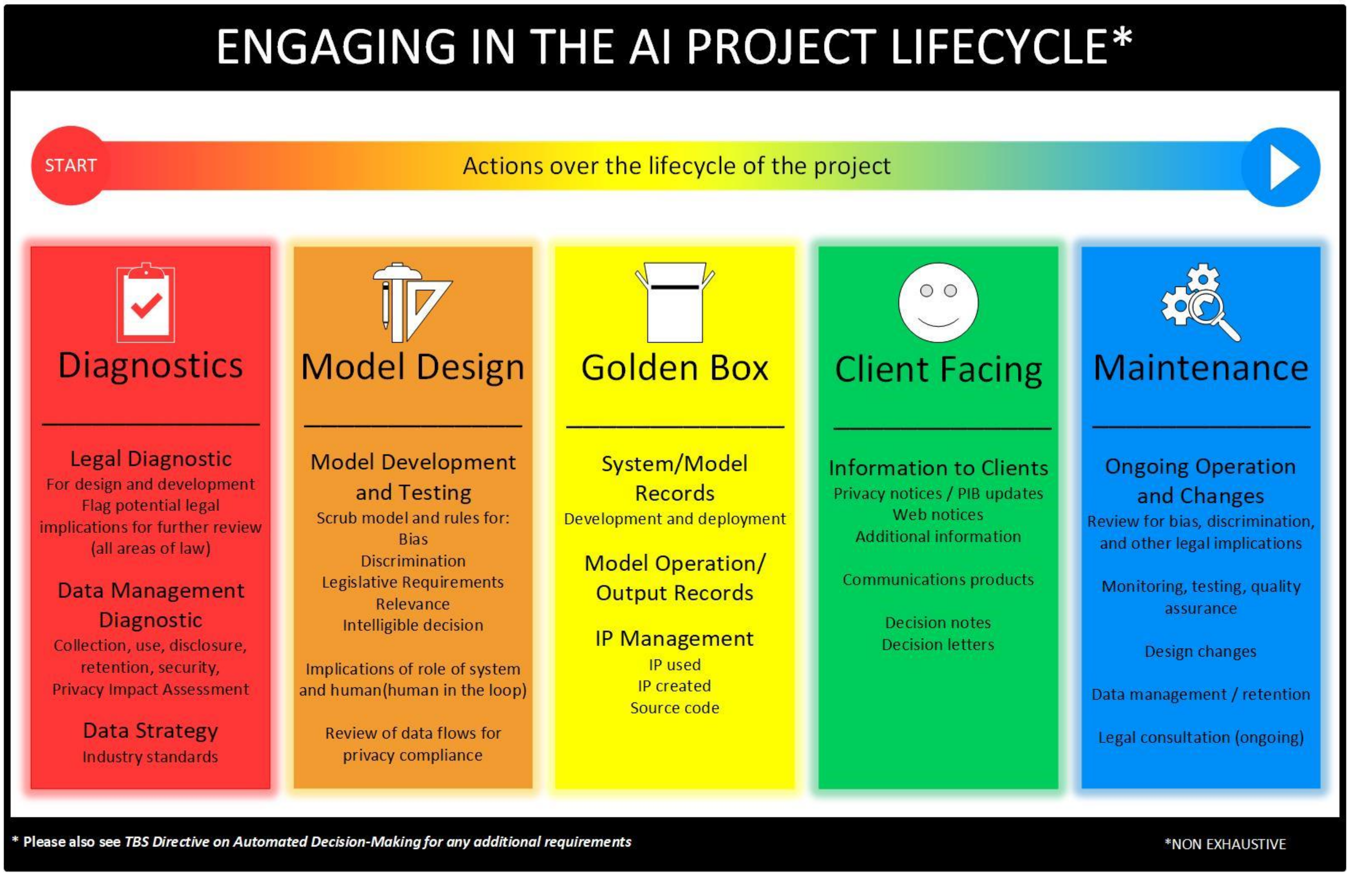


**Figure 6, Practical Tips – Engaging in the AI Project Lifecycle**

### *3.3.10 Following Standards and Best Practices*

Technology solutions and standards in design of AI systems can be an important part of the solution in addressing legal and policy implications and risks. Failure to follow standards and industry best practices can also have legal implications.

Work is ongoing on AI standards with two recent notable developments being the IEEE *Ethically Aligned Design, First Edition*[64] and the Canadian CIO Strategy Council that earned its accreditation to develop National Standards of Canada from the Standards Council of Canada earlier this year. It recently published the *National Standard of Canada on Automated Decision Systems*[65] "creating the world's first consensus-based standard on ethical design and use of automated decision systems. This Standard goes beyond aspirational principles to provide a framework and process that can be both measured and tested for conformity, providing confidence in these systems."[66]

---

[64] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*, 1st ed, (IEEE, 2019), online: < https://standards.ieee.org/industry-connections/ec/autonomous-systems.html> accessed 7 October 2019.

[65] *CAN/CIOSC 101:2019, supra* note 7.

[66] CIO Strategy Council, CIO Strategy Council Publishes National Standard of Canada for Automated Decision Systems (2 October 2019), online: <https://ciostrategycouncil.com/2019/10/cio-strategy-council-publishes-national-standard-of-canada-for-automated-decision-systems/> accessed 9 October 2019.

*3.3.11 Emerging Legal Issues*

Writing a list of emerging legal issues is a daunting task as the list keeps morphing with the evolving technology and the seemingly endless array of potential solutions. There are many emerging legal issues related to AI that merit attention but only a limited selection are referred to below.

i. Collisions of Domains of Law

High profile examples of litigation in the US have highlighted key challenges in relation to the use of AI and other automated tools in various context including assisting with decisions in criminal sentencing and decision-making in administering social benefit programs.[67] In *State v Loomis*, the Wisconsin Supreme Court grappled with balancing competing legal interests of protecting the confidentiality of proprietary AI systems with transparency and disclosure required to permit a challenge of such system to ensure that due process rights of the accused are protected.[68] We must work to find a balance where adequate disclosure can be provided to permit fulsome challenge to such systems while at the same time respecting confidentiality rights in proprietary systems and confidentiality interests in open source systems to the extent that system integrity (e.g. gaming) could be a risk.[69]

Within the legal system itself, tensions are beginning to arise between legal norms, such as the open court principle and respect for the independence of the judiciary, and emerging social concerns, such as the open data movement and the potential for technology to assist in fostering accountability and lack of bias. For example, in Canada, a website was ordered to refrain from republishing Canadian court and tribunal decisions because they contained personally identifiable information.[70] As another example, the French government's recently enacted *Justice Reform Act* contains a provision banning the use of analytical techniques to understand decision-making of individual judges, partly in response to the development of a web tool conceived to assist in understanding judicial decision-making.[71] Legal academics have also expressed concern that differential access to new legal technologies may create access to justice concerns. Finding a balance in responding to the tensions created by novel analytical technologies and their impact on the legal system will likely be an ongoing challenge.

ii. Choice of technology, red lines and trade offs

Challenges are increasingly evident from the choice of technology where the opaque nature of the more complex AI technologies, such as neural networks, may not allow adequate explanation as could be required.

---

[67] See *State v Loomis,* 371 Wis (2d) 235 (2016) [*Loomis*]; *KW v Armstrong,* 180 F Supp (3d) 703 (D Idaho 2016).
[68] *Loomis, ibid*.
[69] K. Hosangar & V. Jair, "We Need Transparency in Algorithms, But Too Much Can Backfire" (Harvard Business Review, 2018), online: <https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire> accessed 15 October 2019.

[70] *A.T. v Globe24h.com*, 2017 FC 114.
[71] J. Tashea, "France bans publishing of judicial analytics and prompts criminal penalty" (ABA Journal, 2019); M. Benesty, "The Judge Statistical Data Ban – My Story, Artificial Lawyer" (Artificial Lawyer, 2019) <https://www.artificiallawyer.com/2019/06/07/the-judge-statistical-data-ban-my-story-michael-benesty> accessed October 15 2019.

This may be relevant whether issues raised are grounded in human rights, administrative law or privacy.[72] Significant tradeoffs have been identified in terms of lost opportunities for use of more advanced AI technologies supporting better AI tools that would more significantly enhance decision-making. [73] We must work with technologists to inform appropriate design of these evolving technologies to meet legal requirements whether for transparency, explainability or other requirements. Ongoing development of new technical methods to make AI models more explainable and interpretable is welcome.[74]

iii. Gaps, uncertainty, or unacceptable results in the law

Experts, academics and stakeholders are frequently debating whether the law needs to be adjusted for AI. While there are some international examples we are essentially at ground zero in Canada. A simple framework to aid answering this question follows. When you apply the current law, is the result:

- A gap since the answer cannot be found?
- Significant uncertainty in the answer?
- A clear but unacceptable answer? For example the public good is not served or the policy intent of the law is compromised.

If the answer to any of the above questions is yes, an adjustment to the legal framework may be needed.

Many questions follow such as who is best placed to adjust the legal framework – the courts, the legislatures or leave it to government policy rather than the law where possible? Is alignment with new laws in other countries desirable or necessary? Is a technology solution viable? There are questions of timing and whether regulating now would be premature. Is it too soon or perhaps more appropriately is an incremental and iterative approach needed by many players to keep up with the evolution of this technology? Each area of law is unique and there will be no panacea solution.

If we are to achieve the ideal where use of advanced technologies complement and even augment respect for the rule of law and the promotion of dignity, addressing these and many other emerging issues is a key challenge. Careful solutions likely need to be explored and developed by the private and public sector in collaboration rather than leaving these extremely complex questions exclusively to the legislatures and the courts, which may lag in responding to new developments with respect to this rapidly evolving technology.

## 4. IRCC'S POLICY RESPONSE

As IRCC has embarked on the project of using AI tools, it has become a place to engage on and explore these issues. Certainly, reaping the benefits of new technologies requires more than technical expertise and sound

---

[72] Austl, Commonwealth, Data61 Commonwealth Scientific and Industrial Research Organisation 2019, *Artificial Intelligence: Australia's Ethics Framework (A Discussion Paper)* by D Dawson et al (CSIRO, 2019) at 34, online: <https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/> accessed 10 October 2019.

[73] Accenture, "Understanding Machines: Explainable AI" (Accenture, 2018), online (pdf):<https://www.accenture.com/_acnmedia/pdf-85/accenture-understanding-machines-explainable-ai.pdf> accessed 10 October 2019.

[74] See e.g., R. Schmelzer, "Understanding Explainable AI", *Forbes Cognitive World* (2019), online: <https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/> accessed 15 October 2019; *Hosangar & Jair*, *supra* note 69.

legal advice. As IRCC made its initial foray into automated decision support, work began on a number of fronts to prepare for a possible future where the innovative becomes ordinary.

In particular, we set out to answer five key questions.

### *4.1 How do we better structure our organization?*

To support the design and responsible use of AI, as well as broader organizational transformation, IRCC has made a variety of changes to its organizational structure. This includes: an in-house lab of technical experts to lead on the development and implementation of algorithmic systems; a dedicated digital policy team concerned with ethics, legal and policy authorities, and connecting the dots for senior management; and a new sector, led by IRCC's Chief Information Officer, focused on transformation and digital solutions. The Department has also established an internal steering committee to serve as a focal point for policy, legal, and privacy issues related to IRCC's use of technology.

While the Department already has officials responsible for providing broad directional guidance on the use and governance of data, IRCC is also working to establish a Chief Data Officer, who will take a broader role in addressing data related issues throughout the Department. IRCC is taking these steps in recognition of the fact that strong data quality and governance are foundational requirements for the responsible and effective use of technologies like AI.

### *4.2 How do we bring in outside perspectives and expertise?*

This is an emerging field and we recognize that the vast majority of the expertise lies outside of IRCC. Engaging external experts will help IRCC undertake this work in a technically sound and responsible manner. We also want to adhere to inclusive design principles, incorporating where possible the perspectives of clients, both typical and outliers.

IRCC has actively engaged federal partners, stakeholders and academics regarding its use of electronic tools in order to glean best practices, counter misconceptions, and to more effectively respond to stakeholder concerns. In 2019, engagement included: consultations with experts in AI ethics, law, and privacy; an IRCC-led symposium on algorithmic government; a presentation and panel discussion at the International Metropolis Conference; and presentations to stakeholder groups such as immigration consultants, the Law Commission of Ontario and the Law Society of Ontario. The Department of Justice has also established a Task Force to research and examine legal and ethical issues in relation to AI. The Task Force is also considering the use of AI as a legal- tech tool and is pursuing and participating in pilots of AI tools.

Maintaining public trust is critical to ensuring a successful immigration program. Potential fear and concerns surrounding AI compound the challenge of reassuring Canadians that our programs are in good hands. Open and transparent external engagement helps to ensure that the public is aware that policymakers are seized with questions of technology ethics, and provides avenues for both critics and proponents to contribute to the process.

### *4.3 How do we interact with government-wide policy in this area?*

In a government context, we need to ensure that we are aligned with other directives and guidelines to which we are held to account.

The Treasury Board Secretariat (TBS) *Directive on Automated Decision-Making*[75], which will come into force in April 2020, provides high-level requirements for federal departments using systems, tools or statistical models to recommend or make administrative decisions about clients. IRCC actively engaged with the TBS during consultations on the development of the Directive, and provided detailed feedback on the design of the Algorithm Impact Assessment tool that accompanies the Directive. IRCC's TRV automation pilot afforded the department valuable insight during the development of the TBS Directive as it represented one of the most concrete uses cases of automated decision-making in the Government of Canada at time.

IRCC's experience with the pilot and its work to introduce a supportive legal framework for the use of technology have given IRCC a strong foundation for ongoing collaboration with Government of Canada partners on AI issues.

### *4.4 How do we address some tricky policy debates?*

Over the course of three years of research, legal advice and technical development of the advanced analytics pilot project for visa applications, IRCC uncovered a number of challenging policy issues. For starters, many of the legal issues discussed in this paper had to be equally considered from a policy perspective. It is a maxim of policy professionals that there can be considerable daylight between what is legally allowable and what constitutes a best practice. In this sense, we need to consider both what we must do and what we should do. For example, with respect to transparency, what is the gap between what must legally be disclosed to the public and what is in keeping with the notion of "open government" in 2019?

As discussed above with respect to the legal framework, many of the risks associated with using AI to support decision-making also have a flip side that is positive. This leaves policymakers much to consider. Bias, for example, is one of the chief concerns identified with respect to data-driven technologies. As noted above, critics are rightly concerned that flaws in data sets or poorly constructed business rules could lead to adverse outcomes multiplied on a large scale. But AI-powered tools can also be used to study existing programs in new ways, to make sure they are achieving their intended results, and to shine a light on any bias that may have inadvertently crept into long-standing practices of human decision-makers. Concern for good data continues to fuel policy work and is behind a renewed focus on strong data management and governance.

Another is that complex algorithmic systems will conceal from outsiders the logic underpinning how decisions are made. The flip side of this is that by codifying heuristics long used by human decision-makers into hard-wired business rules, and auto-generating detailed audit trails for all files that pass through a system, decision-making bodies can lay bare the reasoning behind decision-making in a way that was previously opaque. The question for an organization like IRCC is how much can be disclosed before individuals with malicious intent find themselves with instructions for "gaming the system." These considerations underlie our choice of technology, where we apply it, and how we employ it to augment the decision-making process.

A third example of a two-way policy street is balancing risks to clients (applicants, often foreign nationals) and to the Canadian population. Most critiques of algorithmic systems, and indeed most standards and principles in the field of ethical AI, take an individual rights perspective – that is, they are concerned first and foremost with preventing harm to individuals who are processed by a system. From this perspective, IRCC should focus on ensuring that any automated systems supporting administrative decision-making do not unfairly refuse an applicant. In other words, the benefit of the doubt should go to applicants seeking access to Canada. But part of IRCC's core mandate is protecting the health, safety and security of Canadians. From this perspective, designing a system that automates approvals and always resolves borderline cases in the applicant's favour can carry significant consequences. Canadians take on the health, safety and security risks of allowing visitors to Canada,

[75] *Directive on Automated Decision-Making*, *supra* note 51.

and it is understandable for Canadians to be more concerned about a system mistakenly approving risky individuals than about it mistakenly refusing *bona fide* candidates. IRCC needs to take a balanced view, considering fairness from both the individual and collective perspectives.

The appended guiding principles from the *Policy Playbook on Automated Decision Support* are a reflection of where we have landed, as a department, with respect to these and other questions.

### *4.5 How do we communicate what we learned to the rest of our department?*

Following an extensive period of research, culminating in an internal white paper on digital transformation in 2018, and with a catalogue of promising practices established through the advanced analytics model for visa applications in China, the Department was ready to issue some guidance for the policy and program experts and technologists working toward the next breakthrough in augmented decision-making. Consideration was given to various formats, and in the end a "Playbook" model was selected.[76] In June of 2019, IRCC completed the first iteration of its *Policy Playbook on Automated Decision Support* ("the Playbook") and began consulting with partners.

Designed as a complement to the TBS *Directive on Automated Decision-Making,*[77] the Playbook defines key terms and concepts in the immigration context, and gives staff concrete guidelines related to responsible design, privacy, transparency, procedural fairness and respecting the role of decision-makers. The document starts with a set of 12 core principles, which provide a coherent basis for strategic choices about whether and how to make use of new tools in the future, and then offers a more concrete how-to guide for practitioners. Together, the two halves of the Playbook help to answer questions like: *When, if ever, are "black box" algorithms appropriate? Who is responsible for individual decisions made by an automated system? How do we meaningfully explain automated decisions to clients? How can we give IRCC's officers the benefit of insights gleaned from analytics without fettering their decision-making?*

In developing the *Policy Playbook on Automated Decision Support*, IRCC chose to orient the guidelines in several deliberate and fundamental ways. First, the scope is broad – guidance presented in the Playbook is intended to capture all automated systems that make or support administrative decisions, either in whole or in part. This means the guidance applies even when a decision is not entirely automated (such as when a system classifies cases according to the level of scrutiny they require, flags cases for human review or investigation, or provides recommendations about whether applications should be approved). This broad scope also aligns with the overall approach of providing guidelines rather than a strict compliance regime, and it ensures that the Playbook covers a wide variety of scenarios in which AI could play a role in decision-making. While the scope is broad, the Playbook targets the application of automation and AI in the area of *decision-making only* and is not designed to directly address other uses of these technologies, such as the automation of repetitive clerical tasks or the creation of a chatbot to answer clients' questions.

A second fundamental orientation of the Playbook is that the goal of efficiency (though a driver for innovation) is not in itself sufficient to warrant automated decision support. The Playbook is concerned first and foremost with a larger public good. Advanced analytics and AI present opportunities to *strengthen* administrative decision-making by providing a richness of relevant information to decision-makers, and by focusing their

---

[76] Why "Playbook"? A term not commonly used in government policymaking circles, "Playbook" has been defined as "a series of strategies and approaches an organization adopts in its digital delivery," "a collection of best practice and organisational strategies," and "a digital service manual." Playbooks aim to educate people about how digital works and why it is different. They also standardize and formalize the way an organization works.
[77] *Directive on Automated Decision-Making*, *supra* note 51.

attention on the determinative factors. In this way, new tools should help officers make not only faster, but *better* decisions.

Thirdly, IRCC's Playbook does not impose any definitive limitations on the use of automation and AI. Rather than ban certain applications of AI, such as 'black box' algorithms[78], the Playbook establishes a series of checks and balances that promotes a cautious approach and a commitment to providing meaningful explanations. The Playbook helps program designers and system developers to consider their particular use case and how defensible their proposal would be. This approach was thought to be more supportive of innovation, particularly in light of the difficulty of imagining possible use cases in the future, given the speed at which data science is advancing. Although in practical terms certain cases will likely remain off limits for automation and AI (for example, decisions on asylum decisions), adopting a permissive yet careful approach allows for innovation while also promoting transparency and accountability. Orienting policy guidelines in this way will help build public trust and confidence in new approaches and applications of AI in the sometimes contentious immigration domain.

The practical handbook portion of the Playbook guides innovators through a linear process when considering the development of a new automated decision system, equipping them to consider the right questions at the right times.

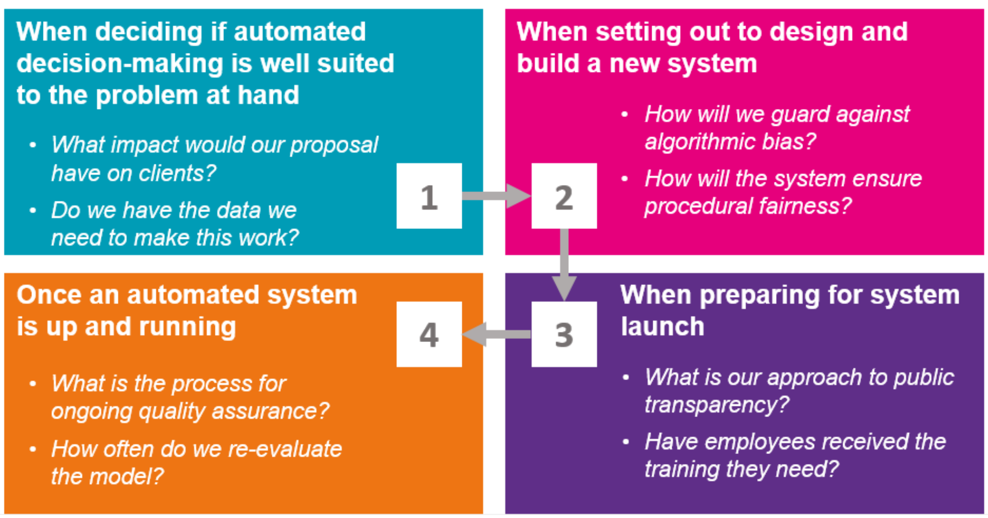


**Figure 7, Four-Stage Approach for Program Designers and System Developers, *Policy Playbook on Automated Decision-Making***

## 5. CONCLUSION

[78] The principle put forward in IRCC's Playbook is that "black box" algorithms cannot be the sole determinant of final decisions on client applications. True "black box" AI tools are of limited utility for administrative decision-making because the Department should not make decisions it cannot meaningfully explain to clients. However, not everything IRCC does concerns the substantive rights of clients and needs to be explained to them. For example, "black box" algorithms could be used for the purposes of business optimization, such as figuring out where to locate offices or how to best distribute resources among them.

Over the past two to three years, IRCC has significantly advanced its thinking on responsible design and ethical use of advanced analytics, AI and automation. Informed by extensive research, engagement of multi-disciplinary experts both within and external to the Department, and experimentation with automated decision-making in targeted low-risk business lines, the development of internal guidance will assist IRCC in navigating complex legal, policy, technological and operational questions.

However, we are still in the early days of adopting AI tools and more work lies ahead of us than behind. Using AI does not change the obligations of government institutions in public law but these technologies do offer an opportunity to interpret and breathe new life into these frameworks to ensure rule of law is respected and that the government is held to account in meaningful ways. Linkages to departmental business planning and decision-making regimes will need to be strengthened in order to ensure that these structures serve as effective checkpoints for the application and enforcement of IRCC and TBS guidelines. As IRCC implements and refines its current policy principles, which are focused on administrative decision-making use cases, it will also need to consider the extension of AI into other domains. This is likely to raise new questions and necessitate further guidance. An inventive research agenda and continued engagement with stakeholders and technical experts will be key to success. So too will be continued collaboration between public and private sectors to address challenges and emerging issues and find solutions that ensure that the potential of AI is fully realised and ethics are intact.

# APPENDIX: GUIDING PRINCIPLES

## FROM: *IRCC POLICY PLAYBOOK ON AUTOMATED DECISION SUPPORT*

## OVERARCHING GOALS

1. ***The use of new tools should deliver a clear public benefit. IRCC should use automated decision support wherever it can do so responsibly, effectively and efficiently – in that order.***

- **IRCC's use of automated decision support should build trust in Canada's immigration system**. Maintaining the confidence of Canadians is critical to the Department's success, and IRCC should actively address potential anxiety over fairness and privacy risks associated with Big Data and AI. The Department should take measured steps, particularly at the outset, to build legitimacy around its use of automation, analytics and AI. New tools should be designed and implemented in a manner that commands trust and understanding.
- There is an opportunity cost to <u>not</u> pursuing the use of automation, advanced analytics and AI. IRCC should not be so cautious as to miss out on the tremendous potential of data-driven technologies. As in any era, **making use of the best tools available is part of effective and efficient use of public money.** In some cases, algorithmic systems are the only efficient way to process the massive amount of information needed for the operation of modern government services.
- However, **IRCC can only employ these technologies to the extent that their use remains consistent with Canadian norms and standards** regarding equality and non-discrimination, procedural fairness, transparency, privacy and accountability. The requirements of administrative law, and of overarching frameworks like the *Charter of Rights and Freedoms* and the *Privacy Act*, are not changed by technology.
- **IRCC should not pursue automating decisions (or aspects of decisions) at the expense of program integrity.** Risk management is at the heart of IRCC's work. Automation may change risks or introduce new ones, but innovation should also seek to boost program integrity.
- Big data and automation should **strengthen administrative decision-making by providing a richness of relevant information to decision-makers, and by focusing their attention on the determinative factors**. In this way, new tools should help officers make not only faster, but *better* decisions.
- **IRCC should make every effort to obtain reliable outcomes data** in order to make a direct link between clients' ultimate outcomes and the information they submit at the application stage. At the same time, **the Department needs to carefully consider the strength of the inferences it draws from new data sets**. Differentiating correlation and causation is a perennial challenge, and human intentions are hard to predict reliably.
- **Big data should be used to help IRCC look backward as well as forward.** IRCC should look to use new tools to detect, assess and remove any bias or inconsistencies that may have inadvertently crept into long-standing business practices.
- **IRCC should strive to constantly improve the quality of its data holdings, given that good data is foundational to advanced analytics and AI.** Prioritizing the development of data-driven tools means prioritizing the breadth, depth and reliability of our data sets.

2. ***Administrative decisions are about people, and they are made by people, even when we use AI. Humans, not computer systems, are accountable for decisions.***

- **The introduction of automated decision support can change the *time and place* of human intervention in the decision-making process** (for example, humans can take on new roles, like setting business rules for an automated triage system to later carry out) **but does not displace the central role of human judgment in decision-making**. There is an essential role for humans in deciding which types of systems to use, which cases to apply them to, and which values to encode.
- **Keeping a human in the loop is important for public confidence. Models that incorporate automated decision support should maintain a role for people in the decision-making process.** Human-in-the-loop systems represent a form of transparency and accountability that is more familiar to the public than automated processes. In cases where algorithms play a role in administrative decisions that significantly affect people's lives, it is reasonable for applicants and Canadians to expect that a real person has exercised human judgment in developing the process, and that a person is ultimately accountable for its final results.
- Automation, advanced analytics and AI are force multipliers; they **assist and augment the capabilities of people**. Using these technologies to allow IRCC staff to focus on more creative, problem-solving, or client-facing work should improve both service delivery and job satisfaction.
- Accountabilities are not passed to machines when they take over tasks previously done by humans. **IRCC must take ownership for the successes and failures of its systems.** Simply referring to systems or tools as "biased" wrongly puts the responsibility for negative outcomes solely on algorithms.

## THE RIGHT TOOLS IN THE RIGHT CIRCUMSTANCES

### 3. *Because IRCC's decisions have significant impacts on the lives of clients and Canadians, the Department should prioritize approaches that carry the least risk.*

- Automating final decisions is the highest risk approach to achieving efficiencies. IRCC should first **explore whether efficiencies can be achieved by automating other aspects of the business flow** (e.g. workload distribution, risk-tiering, note-taking, research, communications, etc.).
- **The degree of human involvement in decision-making should be calibrated to:**
    1. the gravity of the decision, including its impact on clients;
    2. the degree of complexity and amount of subjective judgment involved; and
    3. the quality of the data that can be brought to bear.

  These considerations are more important than whether the action in question is a final decision or an interim step, or whether it is a positive (approval) or negative (refusal) decision.
- When using a predictive model for decision-making, automating approvals is usually considered less impactful than automating refusals (as doing the latter can be viewed as unjust profiling), and this is generally true from the perspective of an individual applicant. However, it is important to view the situation from the perspective of other applicants – particularly one who is applying to a business line where space is limited. **In situations where demand exceeds supply, the stakes for positive decisions are raised**: it may be perceived that an approval (or invitation or acceptance) of one candidate's application comes at the expense of another's.
- **A third perspective is that of Canadians**, who accrue the economic, social and cultural benefits of temporary and permanent migration, but who also take on the safety and security risks. **From a collective perspective, automated approvals based on prediction can carry significant consequences**. It is understandable for Canadians to be more concerned about mistakenly approving risky individuals than about mistakenly

refusing *bona fide* candidates. IRCC needs to take a balanced view, considering fairness from both the individual and collective perspectives.

### *4. "Black box" algorithms can be useful, but cannot be the sole determinant of final decisions on client applications.*

- **True "black box" AI tools are of limited utility for administrative decision-making** because the Department should not make decisions it cannot meaningfully explain. "Black box" tools, such as a facial recognition application, can be used in a supporting role – but even then, **confidence/reliability levels should be sufficiently high, and where they are not, humans must have an opportunity to review the system's results before final decisions are made.**
- Not all decisions taken by IRCC concern the substantive rights of clients and need to be explained to them. **Black box algorithms should be permitted for the purposes of business optimization**, such as figuring out where to locate offices or how to best distribute resources among them.

## RESPONSIBLE DESIGN

### *5. IRCC must recognize the limitations of data-driven technologies and take all reasonable steps to minimize unintended bias.*

- Every machine-learning algorithm operates entirely within the world defined by the data used to calibrate it, so **limitations or flaws in data sets will bias outcomes**, sometimes severely. In creating tools designed to make decisions on immigration, citizenship and passport applications, **IRCC must be careful not to replicate and "hard wire" historical biases** under a guise of technological neutrality. Understanding our data and planning for its responsible use is not a purely technical question – interdisciplinary collaboration is required to address fairness meaningfully.
- All analytical tools are limited in their ability to predict and describe outcomes. While overall patterns and trends are useful, they cannot substitute for evidence of personal actions. For this reason, **IRCC should be very careful when applying information about a group to an individual**.
- Analytics may suggest that an applicant is low-risk, based on how their personal profile aligns with that of successful applicants in the past, and in this case **IRCC may wish to take on the risk of an automated approval. However, the calculus changes in the reverse scenario**: if analytics find that an applicant's profile is consistent with unsuccessful past applicants, suggesting that a refusal may be in order, IRCC should not allow an automated refusal (i.e. the client takes on the risk). As in the case of a humans-only decision-making model, **refusals must be based on *evidence* about the particular person in question, and not solely on their similarity to a general profile of applicants** who have been refused in the past or otherwise indicate riskiness. For example, an automated refusal based on a TRV applicant's passport being in Interpol's Stolen and Lost Travel Documents database would be justifiable because it is based on an objective, verifiable fact about the applicant in question (i.e. they do not meet the requirement for a valid travel document). Conversely, it would not be justifiable to base an automated refusal on the fact that the applicant's country of origin has a high incidence of passport fraud – this could serve as a risk indicator amongst others, but it is not itself a sufficient basis for an outright refusal.

### *6. Officers should be informed, not led to conclusions.*

- Decision support systems can be used to assist officers in exercising their discretion. In these cases, systems must be designed so that they **do not fetter decision-makers** in the exercise of their power.
- **Officers need appropriate training and context to understand what, precisely, a decision support system is telling them**. For example, labels matter: when a system automatically triages clients' applications into

groups, it makes a difference whether these groups are labelled "green/yellow/red" or "straightforward/average/complex." The former may appear to represent instructions to officers (green=approve, red=refuse), the latter a simple indication of how much time and effort an officer will need invest on the way to a decision.

- ❖ A system that provides **no real opportunity for officers to reflect** is a *de facto* automated decision-making system, even when officers click the last button.
- ❖ It is important to establish a **high level of consistency between how officers input data and how the results of data analytics are presented back to them**.

### *7. Humans and algorithmic systems play complementary roles. IRCC should continually strive to optimize these roles and find the right balance, in order to get the best out of each.*

- ❖ **Let computers focus on their strength** – reliably analyzing large volumes of data and considering millions of possibilities without fatigue – **and let humans focus on theirs**: intuition, creativity, empathy, social skills, shaping a larger strategy. Humans are best at framing questions, machines are best at finding the answers. **Algorithmic systems should free people to focus on things that require their expertise and judgment.**
- ❖ **Systems that make administrative decisions on their own** – as opposed to assisting a human decision-maker in making a decision – **are suitable for decisions involving straightforward, factual determinations** (such as whether a client has a criminal record or has visited Canada before). Automation should be focused first and foremost on routine elements – where desired outcomes are easily described and where human judgment is not required.
- ❖ When determining whether a task or step in a decision-making process could be shifted from humans to machines, program designers should ask themselves, "**Is this a situation in which reasonable minds may differ?**" If yes, automation is ill-advised. Conversely, if it is felt that virtually any officer would reach the same conclusion given the facts at hand, automation should be pursued.
- ❖ **Having a human in the loop not only safeguards against procedural fairness risks, it gets better results.** For straight-forward types of assessments (e.g. photo matching), the performance of data-driven algorithms is often superior to that of humans. But when machines and humans team up, the results are better still. One way to boost performance is to have experts apply judgment to the output of mathematical models; another is to flip the sequence and have program experts and decision-makers contribute to the design of the model in the first place.
- ❖ **IRCC should regularly review and adjust its automated systems** to ensure they reflect the reality on the ground and are operating in alignment with the policy priorities of the day. Feedback loops are critical to responsible design and quality control. For example, feedback from frontline officers must be able to reach system developers, so that IRCC can constantly reassess whether a system's conclusions can be trusted.
- ❖ It **is essential that human officers continue to review a portion of applications manually.** This will help to ensure that officer skills do not atrophy and that knowledge about country conditions and fraud trends remains up-to-date.
- ❖ **Technical experts (and their managers) are in a position of trust, and have a special responsibility to thoroughly evaluate computer systems and their possible impacts.** As always, the Department expects objective assessments and frank advice. With the advent of machine learning systems, extraordinary care should be taken to identify, explain and mitigate potential risks, especially when those risks may only be apparent to those with a technical understanding of those systems. Managers need to create conditions that allow experts to consult peers, reassess risks as systems evolve and report hazards without fear of reprisal.

- IRCC should make every effort to **give its employees the skills, tools and support they need** to pursue innovation. Support includes time to learn and space to experiment responsibly.

8. ***On top of respecting Canada's current privacy protection framework, IRCC should continually adopt emerging privacy-related best practices in a rapidly evolving field.***

- Data scientists, program designers and IT specialists should be conversant in privacy issues, including **understanding the rights and responsibilities associated with the collection, use, disclosure and retention of personal information,** as well as **emerging best practices and standards regarding the use of other information related to individuals**.
- The Department should **be proactive in preparing a possible data breach** related to its algorithmic systems. This means having the right team and standard operating procedures in place.

## TRANSPARENCY AND EXPLAINABILITY

9. ***IRCC should subject all systems to ongoing oversight, to ensure they are technically sound, consistent with legal and policy authorities, fair and functioning as intended.***

- Inviting **external scrutiny** of the design of our systems is an important aspect of ensuring legitimacy. Where possible, new algorithmic systems should be made available to external experts, such as an advisory body, for auditing, testing and review. If testing is kept internal, the testing methods, assumptions and results should be documented and made available to the advisory body.

10. ***IRCC must always be able to provide a meaningful explanation of decisions made on client applications.***

- Regardless of the degree to which IRCC relies on algorithmic systems to support decision-making, applicants will have the same rights they have today to understand the basis of decisions made on their applications. Explanations are key to applicants' ability to pursue recourse mechanisms, and to government accountability more generally. Explaining how algorithmic systems work can be technically challenging, but **explanations of automated decisions should not hinge on the general public's understanding of how algorithmic systems function**. If an explanation provides little meaningful information, it offers little value.
- One way to make explanations understandable is to limit a system's performance to those functions that can be easily explained. This may, however, mean squandering the potential of AI systems that can make sense of complex data in ways that humans cannot. **Rather than imposing sharp limits on what new tools can *do*, IRCC should endeavour to find new and better ways to *explain* them.**
- Given a prohibition on black box systems for administrative decision-making in full, the use of automated decision support systems should make decisions **more explainable, not less**. For example, audit trails capturing every factor and step in the decision-making process can be used, where needed, to systematically generate and record reasons for acceptance or refusal.
- An **audit trail** is critical to respecting an applicant's right to understand the basis of the decision on their application, and may be needed to satisfy the courts in the context of a legal challenge. Any new tool that lacks the capacity to **automatically generate a comprehensive audit trail of the administrative decision-making path** is not yet ready for implementation.
- As with explanations of individual administrative decisions, **IRCC needs to find ways to be meaningfully transparent about the workings of whole systems.** While strategies for explaining the Department's work may differ depending on the audience, at a minimum:
  - Clients need to understand how their applications will be processed and their personal information used

— Canadians need to understand how IRCC spends public money to facilitate the entry of family, friends and legitimate visitors, while at the same time protecting against threats to health, safety and security

— Stakeholders, media and technical experts need a basis for an informed critique of Canada's approach

### *11. IRCC must balance transparency with the need to protect the safety and security of Canadians.*

- The Department has an interest in not disclosing too much information about its decision systems for program integrity reasons. Divulging too much information can allow individuals to manipulate decision-making, or "game the system." Explanations of how systems work, and of the individual decisions they help the Department to make, should be **detailed enough to inform a client generally about how decision will be made, but not so comprehensive as to jeopardize program integrity and, by extension, the safety and security of Canadians.**
- Given that all automated decision-making rules could ultimately find their way into the public domain, **IRCC should endeavour to design systems, to the greatest extent possible, around factors/data points that are difficult to falsify or misrepresent**. For example, bank records are more difficult to fabricate than a reason for travel, which is simply stated by the applicant. As much as we can, we should rely on objective facts, in order to mitigate potential fraud risks. The weight assigned to different data elements in IRCC's algorithms should account for the degree to which they may have been targeted for fraud in the past, or are likely to be targeted going forward.

### *12. Clients will continue to have access to the same recourse mechanisms, and IRCC's use of automated systems should not diminish a person's ability to pursue these avenues.*

- When a client challenges a decision that was made in full or in part by an automated system, and IRCC has the authority and agrees to review the decision, **a human officer will conduct the review**.
- However, **IRCC should not proactively offer clients the choice to have a human officer review and decide on their case in lieu of the automated system at the beginning of the application process**. IRCC's objective is to implement responsible, effective and efficient systems. Allowing clients to opt out of them would compromise the Department's ability to ensure consistent and fast processing, while also erroneously implying that its automated systems provide inferior service.