REVIEWING ALGORITHMIC DECISION MAKING IN ADMINISTRATIVE LAW

Michael Lang

Résumé

L'intelligence artificielle est peut-être le changement technologique le plus important depuis la popularisation de l'Internet dans les années de déclin du XXe siècle. L'intelligence artificielle promet d'affecter la plupart des secteurs de l'économie moderne, du camionnage et du transport aux soins médicaux et à la recherche. Notre système juridique a déjà commencé à envisager comment des systèmes de prise de décision artificiellement intelligents sont susceptibles d'affecter l'équité des procédures et l'accès à la justice. Ces effets ont été sous-évalués dans le domaine du droit administratif, dans lequel des systèmes artificiellement intelligents pourraient être utilisés pour accélérer la prise de décision, assurer un traitement relativement égal de cas similaires et lutter contre la discrimination. Mais l'adoption de systèmes artificiellement intelligents par les décideurs administratifs soulève également de graves questions. Cet essai se concentre sur une de ces questions : les décisions administratives prises par des systèmes artificiellement intelligents sont-elles capables de répondre au devoir d'équité procédurale dû aux sujets de ces décisions ? L'essai est divisé en trois parties. Dans la première, j'expose brièvement l'utilisation croissante des systèmes artificiellement intelligents dans le contexte administratif. Nous étudierons principalement les algorithmes d'apprentissage machine en décrivant le défi technique d'inexplicabilité qu'ils posent. Dans la deuxième section, j'expose le devoir des décideurs administratifs d'expliquer leur raisonnement dans certains contextes. Dans la troisième section, je soutiens que les processus administratifs qui utilisent des systèmes artificiellement intelligents compliqueront l'accomplissement efficace de cette tâche. Les personnes soumises à certains types de décisions administratives peuvent être privées des motifs auxquels elles ont droit. Je soutiens que l'intelligence artificielle pourrait nous amener à repenser la justification des décisions en droit administratif.

Michael Lang
Reviewing algorithmic decision making in administrative law

Abstract

Artificial intelligence is perhaps the most significant technological shift since the popularization of the Internet in the waning years of the 20th century. Artificial intelligence promises to affect most parts of the modern economy, from trucking and transportation to medical care and research. Our legal system has already begun to contemplate how artificially intelligent decision making systems are likely to affect procedural fairness and access to justice. These effects have been underexamined in the area of administrative law, in which artificially intelligent systems might be used to expedite decision making, ensure the relatively equal treatment of like cases, and ward against discrimination. But the adoption of artificially intelligent systems by administrative decision makers also raises serious questions. This essay focuses on one such question: whether the administrative decisions taken by artificially intelligent systems are capable of meeting the duty of procedural fairness owed to the subjects of such decisions. This essay is arranged in three sections. In the first, I briefly outline the increasing use of artificially intelligent systems in the administrative context. I focus primarily on machine learning algorithms and will describe the technical challenge of inexplicability that they raise. In the second section, I set out the duty of administrative decision makers to explain their reasoning in certain contexts. In the third section, I argue that administrative processes that use artificially intelligent systems will likely complicate the effective discharge of this duty. Individuals subject to certain kinds of administrative decisions may be deprived of the reasons to which they are entitled. I argue that artificial intelligence might prompt us to rethink reason giving practices in administrative law.

INTRODUCTION

[1] Artificially intelligent machines are being used in a wide and expanding set of domains, from finance and banking (Lin, 2019 at 531), to healthcare and research (Topol, 2019 at 44). Artificial intelligence (AI) promises to make decision making more efficient and precise, in some cases surpassing the speed and accuracy of human decision makers (Groete & Berens, 2020 at 46). AI promises to reshape our economies, make our workplaces more efficient, and make our healthcare systems more effective. Delegating increasing decisional power to AI systems could eliminate redundancy and inaccuracy in important fields, but it could also replicate and entrench existing oppression and inequality (Kim, 2017 at 857). Many of the well documented risks associated with AI decision making might be especially concerning when these systems are used by agents of the state. Much attention has been given to the criminal law context, in which the use of artificially intelligent systems used to make determinations about bail, sentencing, and incarceration conditions could unfairly discriminate against already vulnerable populations (*Ewert v. Canada*, para 66). But state actors in Canada and elsewhere have also signalled an interest in using AI to facilitate administrative decision making (Government of Canada, 2019).

[2] These systems could be used to make numerous kinds of administrative decisions, from municipal governments deciding where to place rodent control bait (Coglianese & Lehr, 2017 at 1161) to national immigration agencies using machine learning algorithms to monitor and disrupt irregular migration (Azizi & Yektansani, 2020 at 183). Challenges raised by AI facilitated decision making in the administrative context will likely be distinct from those relevant in other domains. This essay will focus narrowly on one such challenge: whether the use of AI systems in administrative decision making will frustrate the state's obligations of procedural fairness. It will argue that AI decision systems powered by machine learning, and by deep learning in particular, encounter an explicability challenge, one in which the reasons that support one decisional output over others are unknowable. Insofar as government agencies adopt decisions not supported by reasons, it may be that they are unable to discharge the obligation to provide oral or written reasons

Michael Lang
Reviewing algorithmic decision making in administrative law

Lex-Electronica.org **n°26-2** 2021 Dossier Spécial

for administrative decisions governed by procedural fairness. The essay will be organized in three parts.

Deep learning and inexplicability: In the first part, I briefly outline the technical incapacity of certain machine learning systems to provide reasons justifying their decisional outputs.

Procedural fairness and the duty to give reasons: In the second part, I broadly describe the duty of procedural fairness owed by administrative decision makers in some contexts. This duty sometimes includes an obligation to give oral or written reasons justifying a decision. I will summarize the general conditions under which this obligation may be applicable.

Reasons and machine learning in administrative law: In the third part, I argue that the use of machine learning to make decisions in administrative law is likely to frustrate the capacity of the state to effectively discharge its obligation to give reasons. I suggest that the introduction of AI systems into the administrative context will likely motivate reconsideration of the duty of procedural fairness.

1. DEEP LEARNING AND INEXPLICABILITY

[3] AI decision making takes numerous forms; machine learning is perhaps the most notorious and challenging. Over the past two decades, dramatic advancements in computing led to the development of machines capable of learning from "examples, data, and experience" (Royal Society, 2017 at 16). Conceived broadly, machine learning algorithms are computational systems "designed to emulate human intelligence by learning from the surrounding environment" (El Naqa & Murphy, 2015 at 3). Though machine learning has received an abundance of recent of attention in popular media and scholarship, its underlying principles and ideas are nearly as old as computing itself (Royal Society, 2017 at 25). The vision of automated systems capable of learning and operating without significant human intervention was long the dream of futurists and fiction writers. But rapid accelerations in computing power and an accompanying proliferation of accessible personal data have made the development of sophisticated machine learning algorithms possible. Precisely delineating the way machine learning algorithms work is far beyond the scope of this essay. What is important for our present purposes is just that machine learning systems

are characterized by their capacity to make predictions and decisions on the basis of sample data without having been directly programmed to do so (Royal Society, 2017 at 29)**.**

[4] Machine learning systems sit on a spectrum. On one end of the machine learning spectrum are AI systems with predictive properties structured heavily by human programmers. The model operates, in other words, according to some set of instructions or assumptions provided by its developers and informed by human supervisors (LeCun et al, 2015 at 436). On the other end of this spectrum are AI systems intended to principally operate without human instruction or intervention (Beam & Kohane, 2018 at 1317). This latter kind of machine learning system engages in *representation learning*, "a set of methods that allows a machine to be fed with raw data and to automatically discover the representations needed for detection or classification" (LeCun et al, 2015 at 436). The most sophisticated and revolutionary variety of machine learning system are *deep learning models*, in which representational learning operates on multiple layers of representational processing (LeCun et al, 2015 at 436). Deep learning models are constructed of *artificial neural networks*, computing structures inspired by neurons in biological systems (Priddy & Keller, 2005 at 5). They are both notoriously and "stunningly complex" (Beam & Kohane, 2018 at 1317) and "very good at discovering intricate structures in high-dimensional data" (LeCun et al, 2015 at 436). While their range of applications is potentially enormous, deep learning systems have revealed particular promise in medical imaging (Wang et al, 2018 at 293). As I suggest below, they are likely also to have significant impact on administrative decision making.

[5] Deep learning models, owing to their programmatic and operational complexity, encounter what is usually described as a *black box problem*. Broadly stated, this problem refers to the inability of *post hoc* reviewers to determine the reasons for which a deep learning system reaches one decision rather than another (Castelvecchi, 2016 at 21). Deep learning models will often give decisional reasons that are "too complex for us to explicitly understand" (Price, 2017 at 430). A decisional system, for example, might consider a set of factors that, while each conceivably interpretable on their own, bear only an unintuitive or confounding relationship to each other. Deep learning models have the "capacity to learn subtle relationships in data that humans might overlook or cannot recognize" (Selbst & Barocas, 2018 at 1094). Consider a deep learning

system trained in image recognition, and specifically in the recognition of ducks from large and diverse sets of bird images. While a human decision maker might look at the shape of a bird's bill or the shade of a bird's feathers, a deep learning algorithm will likely focus on altogether different factors, such as background landscape, texture contrast, or image subject orientation. The deep learning system will engage with these factors in permutations that might surprise human reviewers. It might consider factors that human reviewers would ignore or consider many more factors than any human reviewer could realistically hold in contemplation (Price, 2017 at 430). This complexity means that deep learning decisional processing is often not just *difficult*, but *impossible* to explain.

[6] This inexplicability problem raises a series of potential challenges in the administration of justice. Persons harmed by the effects of an automated decision may seek redress in the private law. But judges and lawyers unable to assess the reasons for which a deep learning system reached an injury–causing decision face a significant litigation challenge. The reasons for which the offending decision is made, after all, will likely play an essential role in adducing a case in contract or extracontractual obligation (Atkinson et al, 2020 at 18). Similar challenges exist in the public law context. In the following part of this essay, I begin suggesting that inexplicability raises a unique problem for administrative law. Administrative decisionmakers will often owe the subjects of their decisions an obligation of procedural fairness. As I show below, this obligation will under certain conditions include a duty to give reasons, a duty likely challenged by inexplicability.

2. PROCEDURAL FAIRNESS AND THE DUTY TO GIVE REASONS

[7] Procedural fairness is a cornerstone principle in modern Canadian administrative law (Huscroft, 2013 at 149). The obligation of procedural fairness owed by administrative decision makers to subjects of their decisions has its proximate roots in House of Lords jurisprudence of the 1960s and Supreme Court jurisprudence of the late 1970s. In the 1979 *Nicholson v. Haldimand-Norfolk* decision, Chief Justice Laskin, writing for a 5–4 majority Court, introduced a general obligation of procedural fairness :

> *I am of the opinion that although the appellant clearly cannot claim the procedu-ral protections afforded to a constable with more than eighteen months' service, he cannot be denied any protection. He should be treated "fairly" not arbitrarily. I ac-*

cept, therefore, for present purposes and as a common law principle what Megarry J. accepted in Bates v. Lord Hailsham, at p. 1378, "that in the sphere of the so-called quasi-judicial the rules of natural justice run, and that in the administrative or executive field there is a general duty of fairness" (Nicholson v. Haldimand-Norfolk at 324).

[8] Before *Nicholson*, administrative law was dominated by natural justice, the application of which depended on the characterization of a decision as 'judicial,' 'quasi–judicial,' or administrative (Huscroft, 2013 at 148). There are two foundational natural justice principles : *audi alteram partem*, in which a decision maker must hear both sides of an issue, and *nemo judex in sua causa*, in which a decision maker is forbidden from standing as judge in their own case (Huscroft, 2013 at 148). While natural justice principles were enforced in judicial and quasi–judicial decision making, they did not exist in merely administrative cases (Huscroft, 2013 at 149). Persons subject to an administrative decision before *Nicholson* would need either to convince a court that their matter exemplified a judicial and quasi–judicial character or else be left with no procedural protection (Huscroft, 2013 at 149). As the administrative state expanded in the postwar period, this position became increasingly untenable. Administrative bodies and tribunals empowered to make decisions affecting important individual interests faced no formal procedural constraints. *Nicholson* changed this condition. Following *Nicholson*, the principles of natural justice were gradually subsumed into procedural fairness. Grant Huscroft maintains that there is, in practice, no longer any reason to distinguish between these concepts (Huscroft, 2013 at 150).

[9] Procedural fairness, at minimum, requires that the subjects of administrative decisions have a right to be heard at an independent and impartial hearing (Huscroft, 2013 at 151). Though procedural fairness is presumptively applicable in administrative law, specific statutory language can abrogate an administrative decision maker's procedural fairness obligations (*Kane v. Board of Governors of UBC* at 1113). In the 1999 *Baker v. Canada* decision, the Supreme Court defined the content of the administrative obligation of procedural fairness. It did so in the context of an immigration decision on an order for the deportation of Mavis Baker, a visitor who had overstayed her visa in Canada (*Baker v. Canada*, para 2). Baker sought an exemption from the requirement that she apply for permanent residence from outside of Canada on humanitarian and compassionate grounds (*Baker v. Canada*, para 3). Her application on for

an exemption was denied and Baker appealed, arguing among other things that the Minister of Immigration and Citizenship had not adequately discharged its obligations of procedural fairness (*Baker v. Canada*, para 11). Justice L'Heureux–Dubé's opinion sets out five criteria that determine the structure of an administrative body's procedural fairness obligations. The criteria are:

> 1. *" the nature of the decision being made and the process followed in making it,"*
>
> 2. *" the nature of the statutory scheme and the 'terms of the statute pursuant to which the body operates,'"*
>
> 3. *" the importance of the decision to the individual or individuals affected,"*
>
> 4. *" the legitimate expectations of the person challenging the decision,"*
>
> 5. *and "the choices of procedure made by the agency itself" (Baker v. Canada, paras 23–27).*

[10] These criteria are each equally relevant in characterizing procedural fairness obligations and no one factor is more important than the others (Huscroft, 2013 at 167). The *Baker* criteria are similarly not intended to be an exhaustive list of factors structuring procedural fairness (Huscroft, 2013 at 167). This position has been reiterated by the Court on several occasions since *Baker* (See *Canada v. Mavi*, para 42). While the five *Baker* criteria help to illuminate the overall structure of procedural fairness obligations in varying administrative contexts, they do not on their own describe what procedural fairness specifically requires. Though the demands of procedural fairness are many, there are roughly a half dozen rules to which administrative decision makers will typically be bound. These include a duty to give notice, a duty of disclosure, and a right to an oral hearing (Huscroft, 2013 at 171–173). In the context of this essay, the most important specific demand of procedural fairness is the duty of administrative decision makers to give reasons. Like procedural fairness itself, the duty to give reasons has its roots in *Baker* (Huscroft, 2013 at 177). Justice L'Heureux–Dubé writes :

> *In my opinion, it is now appropriate to recognize that, in certain circumstances, the duty of procedural fairness will require the provision of a written explanation for a decision. The strong arguments demonstrating the advantages of written reasons suggest that, in cases such as this where the decision has important significance for*

[11] This framing suggests that the duty to give reasons it is not without limit. Justice L'Heureux–Dubé states explicitly that this duty extends only when a decision bears particular importance for an individual or where the administrative framework would mitigate in favour of a necessity for reasons, such as when there exists a statutory right of appeal. While measuring the significance of a decision might appear somewhat nebulous on its face, the existence of a right of appeal has a kind of straightforward resonance. A right of appeal logically demands an account of the reasons for which the decision to be appealed was reached. It will, as Grant Huscroft writes, be "difficult, if not impossible, to determine whether to appeal a particular decision and which sorts of arguments to make on appeal if no explanation is provided for that decision" (Huscroft, 2013 at 177). While the *Baker* decision sets out that administrative decision making will sometimes require reasons, it does not say anything about the quality or form such reasons are expected to take. In the *Newfoundland Nurses* case, Justice Abella confirms this, writing that *Baker* does "not say that reasons were *always* required, and it did not say that the *quality* of those reasons is a question of procedural fairness" (*Newfoundland Nurses*, para 20; emphasis in original). In that same decision, Justice Abella suggests that the threshold for acceptable reasons in cases that require them is quite low. She writes:

> *It strikes me as an unhelpful elaboration on Baker to suggest that alleged deficiencies or flaws in the reasons fall under the category of a breach of the duty of procedural fairness and that they are subject to a correctness review. As Professor Philip Bryden has warned, "courts must be careful not to confuse a finding that a tribunal's reasoning process is inadequately revealed with disagreement over the conclusions reached by the tribunal on the evidence before it" ("Standards of Review and Sufficiency of Reasons: Some Practical Considerations" (2006), 19 C.J.A.L.P. 191, at p. 217) ...*
>
> *It is true that the breach of a duty of procedural fairness is an error in law. Where there are no reasons in circumstances where they are required, there is nothing to review. But where, as here, there are reasons, there is no such breach (Newfoundland Nurses, paras 21–22; emphasis in original).*

[12] Where reasons are required, in other words, even qualitatively deficient reasons would appear to satisfy the obligation set out in *Baker*.

Following *Newfoundland Nurses*, certain commenters noted that the law on the duty to give reasons was wanting for clarity (Huscroft, 2013 at 179). In the *Alberta v. Alberta Teachers' Association* decision, for example, the Supreme Court contemplated that an administrative decision not accompanied by sufficient reasons might need to be returned to the decision maker for additional clarity (*Alberta v. Alberta Teachers' Association*, para 23). Against this backdrop, the most important recent shift in the law of administrative review came in the Supreme Court's landmark 2019 *Canada v. Vavilov* decision. Like *Baker,* the *Vavilov* case centres on an immigration question, whether the Canadian–born child of foreign spies is entitled to Canadian citizenship (*Vavilov*, para 3). *Vavilov*, along with its companion case *Bell Canada*, reframes and answers a lingering and contentious question in the law of administrative review: whether administrative decisions should be generally reviewed by courts on a reasonableness or correctness standards. Following the *Dunsmuir* decision in 2008, the Supreme Court vacillated between reasonableness and correctness, defining and redefining their respective application according to the nature of the administrative decision under review. Though the question of standard of review is probably the dominant academic and judicial concern in the administrative context, it is not strictly our primary concern here. But there is an intimate and essential relationship between the provision of reasons and standards of review.   The Court notes in *Vavilov* that reasonableness review begins necessarily with an engagement with the reasons for which the impugned administrative decision was reached (*Vavilov*, para 84). Written reasons, after all, are "the means by which the decision maker communicates the rationale for its decision" (*Vavilov*, para 84). A reasonable decision, in turn, is one that "is based on an internally coherent and rational chain of analysis" and in which the relevant reasons are "justified in relation to the facts and law that constrain the decision maker" (*Vavilov*, para 85). This intimate relationship between reasons and the reasonableness standard of review, while perhaps unsurprising, underscores the essential and dominant role that the explanation of decisions through written reasons plays in administrative law. Having introduced the duty to give reasons and sketched out its relationship to standard of review determination, I will turn in the next section to the role that deep learning decision making might play in disrupting the adequate discharge this duty.

3. REASONS AND MACHINE LEARNING IN ADMINISTRATIVE LAW

[13] In the first part of this essay, I outlined how certain machine learning decisional systems, deep learning algorithms in particular, produce inexplicable decisions. These systems provide conclusions from an assessment of factors but are incapable of explaining why they have reached one decision rather than another. In the second part of the essay, I outlined the duty of administrative decision makers in certain contexts to provide written reasons explaining their decision. This obligation stems from the general obligation of procedural fairness set out in the *Nicholson* decision and, as the *Vavilov* case specifies, it bears an essential relationship to a dominant concern in administrative law: whether a reasonableness or correctness standard of review should be applied. In this final part, I try to connect these positions. The concern, in broad strokes, is the following. As state entities and administrative decision makers begin to increasingly adopt artificially intelligent systems to assist their decision making, there will inevitably arise instances in which the reasons justifying certain conclusions cannot be known and human reviewers will be unable to explain how the automated system reached the conclusion that it did. This is a problem for administrative law. On one hand, administrative law is fundamentally committed to the doctrine of procedural fairness, which under the right conditions produces a duty on the part of administrative decision makers to give a written account of their reasons. On the other hand, inexplicable algorithmic decision making is being increasingly used to make decision making in a diverse set of domains, including in the administrative state, fairer and more efficient. Whether these positions can be reconciled ought to be a serious concern for administrative decision makers. Either inexplicable decision making will need to be abandoned or the obligation to provide written reasons will need to be compromised.

[14] Inexplicable decision making, in other words, is in clear conceptual tension with the obligation to give written reasons. Persons subject to automated administrative decisions might be incapable of knowing why a decision affecting their interests has been reached. The decisional system will by design provide no explanation and, as we saw above, this might

limit the capacity of individuals to challenge the system's conclusions. It is difficult, after all, to make out a coherent case against a decision for which no justification can be given. For especially important decisions, or for decisions that are accompanied by a right of appeal, courts are likely to interpret the absence of reasons as a breach of procedural fairness. And while the *Vavilov* decision specifies that written reasons "must not be assessed against a standard of perfection" and that administrative decision makers should not be held in their reasons to the same standard of completeness and rigour to which we might hold a judge or a lawyer (*Vavilov*, para 91), the unavailability of reasons in even a narrow set of administrative contexts could prove highly disruptive to the law of administrative review. These systems will almost certainly be used in the medium term by administrative agencies in ways that implicate important individual interests. They will do so, at least some of the time, in ways that make the provision of reasons impossible. Even if the use of inexplicable algorithmic systems is limited, this set of conditions strikes at the very heart of the law of administrative review and potentially limits the ability of the legal system to operate as a check on administrative authority. Judicial reasonableness review, after all, *requires* written reasons. Courts will be limited in their capacity to ensure procedural fairness when the reasons justifying decisions are shrouded necessarily in mystery.

[15] This is not a merely abstract challenge, for administrative decision makers are already beginning to adopt algorithmic systems in their work. Consider, for example, algorithmic decision making applied in the context of immigration law. Immigration, Refugees, and Citizenship Canada (IRCC) has been in the process of developing an automated predictive analytics system "to automate activities currently conducted by immigration officials and to support the evaluation of immigrant and visitor applications" (Molnar & Gill, 2018 at 14). In particular, the IRCC has suggested that algorithmic systems are presently being used in to identify immigration fraud and to provide recommendations concerning the acceptance or refusal of certain visa applications (Molnar & Gill, 2018 at 14). It is unclear, at present, precisely how widespread these practices are within IRCC, but the agency appears interested in making further use of algorithmic processing to expedite and simplify the review of immigration applications. These practices raise significant questions about discrimination and human rights (Molnar & Gill, 2018 at 10), including whether the algorithms in use are being trained on appropriately representative datasets. I want to suggest that the potentially inexplicable

nature of IRCC decisions founded on algorithmic processing ought to be a similarly pressing concern. To be sure, the precise technical nature of IRCC's most recent algorithmic innovations is unknown, and little easily accessible public information on whether decisional outcomes are likely to be explainable yet exists. But the trends are relatively clear. Algorithmic decision making is becoming more common in the administrative state and inexplicable, deep learning is becoming more common in algorithmic decision making. It is likely only a matter of time before inexplicable decision making proliferates in public administration. It may be there already.

[16] In beginning to address concerns surrounding the application of automated decision making by state bodies, Canada's federal public service set out a series of guardrails in the Treasury Board Secretariat (TBS) 2019 *Directive on Automated Decision-Making* (TBS, 2019). Federal agencies and departments using automated systems to recommend or make administrative decisions are required by the Directive to (1) notify decision subjects that a decision will be made in whole or in part by an automated decision system and (2) provide a meaningful explanation to affected individuals of how and why the decision was made (TBS, 2019 at 6.2.1–6.2.3). The TBS Directive applies to externally sourced systems, tools, and statistical models used by federal agencies and bodies (TBS, 2019 at 5.2). Narrow exceptions include automated systems in the national security context or those used by agents of Parliament, such the Office of the Auditor General and the Office of the Chief Electoral Officer (TBS, 2019 at 5.4 & 9.1.1). While TBS rules represent a significant development in the regulation of automated decision making in the federal government, the Directive is likely not to be an effective mechanism for ensuring that persons subject to administrative decisions receive the explanations to which they are entitled. Though the TBS Directive requires a "meaningful explanation" of how and why automated decisions are made, it does not clearly demarcate how compliance with this standard ought to be assessed. There is a significant gulf between the normative position that explanations ought to be given and the tangible mechanisms through which this is actualized. The Directive does not, for example, define what constitutes a meaningful explanation. What might be a meaningful and adequate explanation from the perspective of the subject of an administrative decision, after all, might be insufficient for the purposes of administrative law and the decision subject's right to an explanation. TBS rules, more troublingly, also operate on an assumption

that automated decision systems are conceptionally *capable* of admitting of the kinds of explanations that would produce decisional transparency. But as I outlined above, inexplicable decision systems are not just difficult to satisfactorily explain, they are impossible to explain. To the extent that even a system's initial programmer will be unable to understand how and why a particular automated decision was made, the TBS Directive may be requiring more than it can reasonably be thought to deliver.

[17] We may thus interpret the Directive's demand for explanation in one of two ways. On the one hand, it could be that TBS rules implicitly prohibit the use of inexplicable decision systems in the administrative context. If state agencies and departments are required to give explanations, and such explanations are sometimes unavailable, we might conclude that inexplicable systems are prohibited by necessary implication. But this conclusion would appear to be in tension with signals from within the administrative state, including within IRCC, that express a strong interest in making more widespread use of cutting edge automated decision making. On the other hand, it could be that TBS rules require that automated decision making be supplemented with human review such that reasons and parameters for a decision could be inferred and communicated to the decision subject. But this interpretation would generate significant inefficiency in the decision making process, for it would effectively require that some set of decisions made by state actors would be made twice: once by an algorithm and again by a human reviewer.

[18] Arguments for *post hoc* written reasons that are designed to satisfy procedural fairness obligations also face another, potentially deeper objection. These reasons will not be, and perhaps cannot be, the *actual* reasons for which an inexplicable deep learning model reached some decisional output. Reasons accompanying an automated decisional output may be contrived to fit a given conclusion, drafted not in a manner that exposes an artificially intelligent system's real internal processing, but that expresses a human decision maker's presumptions about how the conclusion might or ought to have been reached. As artificially intelligent decision making systems become increasingly common fixtures in the administrative state, it may be that human decision makers will tend to defer to their automated judgment. To the degree that human decision makers perform a supervisory function, they may work to contrive justificatory reasons in support of automated decisions rather than

carefully scrutinize them. This, of course, raises a broader concern about human psychology and of the potential inclination of human decision makers to trust machine enabled processing. While this tendency is not firmly established in the deep learning algorithmic context, its possibility raises serious concerns about *who*, exactly, is making decisions in administrative settings and what the reasons justifying these decisions really signify.

[19] So, neither possible interpretation of the TBS Directive's explanation provisions is especially compelling. That subjects have a right to a meaningful explanation of decisions affecting them appears to simply restate the dichotomy outlined above: either inexplicable decision making should be abandoned in light of the duty to give reasons or the duty to give reasons should be compromised in light of inexplicable decisions. I do not think either of these approaches are particularly attractive. Reasons are not mere incidental artefacts in decisional processes; they are at the heart of procedural fairness and are a primary constraint on the arbitrary exercise of state power. At the same time, automated, algorithmic decision making is likely to soon become an irresistible force in an increasingly complex administrative state. Algorithmic decision making, considered in its most favourable light, promises to make administrative processes more efficient and equitable. These are objectives to which decision makers are right to aspire. How then, can the inexplicable nature of certain algorithmic decisions be reconciled with the fundamental obligation to provide written reasons? I do not propose to set out a complete response here, only to outline in broad terms some of the principles that ought to feature in an eventual solution. For one thing, I think it is essential that administrative decision makers resist the possible temptation to replace conventional specific written reasons with rote or canned explanations that purport to capture an algorithm's internal processing. An obligation to give reasons might prompt human decision makers to supplement algorithmic processes with contrived estimations about how an algorithm might have weighed a set of factors, even if the actual algorithmic process is programmatically unknowable. It strikes me as a potentially serious breach of basic fairness to intentionally provide the subjects of decisions manufactured explanations of an outcome affecting their interests. In the absence of a better model, it may be preferable that subjects are informed that an inexplicable system was used in their matter than for a human administrator to make up reasons that do not reflect any actual decisional process. For another thing, and

notwithstanding the above, it might be permissible under the right conditions to replace the requirement for specific written reasons with a requirement that individuals subject to an inexplicable decision making process are provided an explanation in broad terms of how the relevant algorithmic system works. Some administrative decisions are not likely to raise substantially important individual interests. In such cases, it might be appropriate to forego reasons provided by a human administrator in favour of information outlining that an inexplicable algorithm was used, and that the inexplicable algorithm produces decisions falling within a certain range of possible outcomes.

[20] Neither of these views lend themselves to a complete and coherent strategy for addressing inexplicable algorithmic decision making in administrative law. This is a highly multifaceted issue, a full accounting of which is well beyond the scope of this paper. For now, I think the following observations might help to structure our thinking: (1) the duty of administrative decision makers to give reasons should not be compromised solely in pursuit of more efficient algorithmic decisions, (2) administrative decision makers should not provide contrived explanations of inexplicable algorithmic decisions, and (3) it may in certain circumstances be appropriate to allow explanations of an algorithm's essential functions and range of possible decisions to satisfy the obligation of administrative decision makers to give written reasons. All of this provides much for courts and administrative agencies to consider. And while this essay does not provide a cohesive response, I do want to suggest that the problem is one of significant magnitude. Automated decision making prompts us to think differently about the role reason giving plays in assuring fairness in administrative law. Doing so is of the highest importance, for holding the administrative state to account for its decisions, ensuring that its decisions are procedurally fair, vitally requires that we are able to understand why and how its decisions were made. As more decisions in public administration become inexplicable, courts and administrators will need to carefully consider the demands of the law of procedural fairness in this rapidly shifting context.

CONCLUSION

[21] This essay argues that the increasing use of deep learning decisional systems by administrative agencies is likely to challenge our conventional conception of procedural fairness obligations in administrative law. In the

first part of the essay, I show how deep learning systems make decisions that are inexplicable. The reasons for which these decisions make one decision rather than another are not just complicated, they are often *unknowable*. In the second part of the essay, I describe the obligations of procedural fairness to which administrative bodies are bound. These obligations include, sometimes, a duty to give written reasons explaining why a decision has been made. In the third part of the essay, I tie together these conditions and argue that the use of deep learning algorithmic decision making processes in administrative contexts is likely to frustrate the adequate discharge of obligations of procedural fairness. I suggest that this is not a small problem. Procedural fairness and written reasons are one of a small number of formal safety valves that constrain the state power wielded by administrative bodies. As these bodies increasingly draw on automated systems to facilitate their decision making mandates, it will be critical that the law responds. Scholars and legislators should ward against the imposition of decisional processes that cannot be explained to the people whose interests are subject to them. Reason giving is an important control on administrative authority, one that the allure of decision making efficiency should perhaps not prompt us to abandon.

Bibliography

Articles

Katie Atkinson, Trevor Bench–Capon & Danushka Bollegala, "Explanation in AI and law: Past, present and future" (2020) 289 Artificial Intelligence 1.

Andrew L Beam & Isaac S Kohane, "Big Data and Machine Learning in Health Care" (2018) 319:13 JAMA 1317.

Government of Canada, *Directive on Automated Decision-Making* (Ottawa: Her Majesty the Queen in Right of Canada, represented by the President of the Treasury Board, 2019).

Davide Castelvecchi, "Can we open the black box of AI?" (2016) 538:21 Nature 21.

Thomas Grote & Philipp Berens, "On the ethics of algorithmic decision-making in healthcare" (2020) 46 J Med Ethics 205.

Grant Huscraft, "From Natural Justice to Fairness: Thresholds, Content, and the Role of Judicial Review" in Colleen M Flood & Lorne Sossin, eds, *Administrative Law in Context*, 2d ed (Toronto: Emond, 2013) 147–184.

Pauline T Kim, "Data-Driven Discrimination at Work" (2017) 58:3 William & Mary Law Review 857.

Yann LeCun et al, "Deep learning" (2015) 521 Nature 436.

Tom CW Lin, "Artificial Intelligence, Finance, and the Law" (2019) 88 Fordham Law Review 531.

Petra Molnar & Lex Gill, Bots at the Gate: A Human Rights Analysis of Automated Decision Making in Canada's Immigration and Refugee System (Toronto: Citizen Lab, 2018).

Issam El Naqa & Martin J. Murphy, "What Is Machine Learning?" in Issam El Naqa, Ruijiang Li & Martin J. Murphy, *Machine Learning in Radiation Oncology* (Zurich: Springer International Publishing, 2015) 3–11.

W Nicholson Price II, "Artificial Intelligence in Health Care: Applications and Legal Issues" (2017) University of Michigan Public Law Review No. 599.

Keven E Priddy & Paul L Keller, *Artificial Neural Networks: An Introduction* (Bellingham: International Society for Optical Engineering, 2005).

Royal Society, Machine learning: the power and promise of computers that learn by example (London: Royal Society, 2017).

Andrew D Selbst & Solon Barocas, "The Intuitive Appeal of Explainable Machines" (2018) 87 Fordham Law Review 1085.

Eric Topol, "High-performance medicine: the convergence of human and artificial intelligence" (2019) 25 Nature Med 44.

Treasury Board Secretariat, *Directive on Automated Decision-Making* (Ottawa: Her Majesty the Queen in Right of Canada, 2019).

Fei Wang et al, "Deep Learning in Medicine—Promise, Progress, and Challenges" (2018) 179:3 JAMA Internal Medicine 293.

Jurisprudence

Alberta (Information and Privacy Commissioner) v. Alberta Teachers' Association, 2011 SCC 61, [2011] 3 SCR 654.

Baker v. Canada (Minister of Citizenship and Immigration), [1999] 2 SCR 817, 174 DLR (4th) 193.

Canada (Attorney–General) v. Mavi, 2011 SCC 30, [2011] 2 SCR 504.

Canada (Minister of Citizenship and Immigration) v. Vavilov, 2019 SCC 65, 441 DLR (4th) 1.

Ewert v. Canada, 2018 SCC 30, [2018] 2 SCR 165.

Kane v. Board of Governors of UBC, [1980] 1 SCR 1105, 110 DLR (3d) 311.

Nicholson v. Haldimand-Norfolk Regional Police Commissioners, [1979] 1 SCR 311, 88 DLR (3d) 671.

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