

Artificial intelligence for a reduction of false denials in refugee claims*

Hilary Evans Cameron
Faculty of Law
Ryerson University
h.evanscameron@ryerson.ca
CANADA

Avi Goldfarb
Rotman School of Management
University of Toronto and NBER
avi.goldfarb@rotman.utoronto.ca
CANADA

Leah Morris
Munk School of Global Affairs
University of Toronto
leah.morris@rotman.utoronto.ca
CANADA

March 2021

Abstract

Deciding refugee claims is a paradigm case of an inherently uncertain judgment and prediction exercise. Yet refugee status decision-makers may underestimate the uncertainty inherent in their decisions. A feature of recent advances in artificial intelligence (AI) is the ability to make uncertainty visible. By making clear to refugee status decision-makers how uncertain their predictions are, AI and related statistical tools could help to reduce their confidence in their conclusions. As it now stands, this would only hurt claimants, since countries around the world have designed their refugee status determination systems to resolve decision-making uncertainty at the claimant's expense. Increasing uncertainty would therefore contribute to mistaken rejections. If, however, international refugee law was to recognize an obligation under the UN Convention to resolve decision-making doubt in the claimant's favour, as Evans Cameron has recently advocated, then by making uncertainty visible, AI could help to ensure that fewer refugees were wrongly denied the protection they need. Moreover, in many countries around the world, refugee claims are decided in legal proceedings governed by the logic of inductive inference, a form of reasoning that profoundly distorts risk assessment. Under Evans Cameron's proposed risk-assessment model of refugee status decision-making, one based on abductive rather than inductive inference, AI's ability to make uncertainty visible would indeed help to resolve doubt in the claimant's favour, as the drafters of the Convention intended.

I. Introduction

A wide literature in psychology and behavioral economics documents the difficulty people have in understanding probabilities (e.g. Kahneman 2011; Griffin and Tversky 1992; Tversky and Kahneman 1974). This often leads people to be “unjustifiably certain of their beliefs” (Russo and Schoemaker 1992). Deciding refugee claims is a paradigm case of an inherently uncertain judgment and prediction exercise.

* We gratefully acknowledge financial support from SSHRC (grant # 502500) and the Sloan Foundation. This work has benefited from helpful discussions with Ajay Agrawal and Joshua Gans and constructive suggestions from the review team.

Yet refugee status decision-makers may underestimate the uncertainty inherent in their decisions. Indeed, some report having considerable confidence in their decisions (Evans Cameron 2008; Colaiacovo 2018; CBC/Radio Canada 2019).

One feature of artificial intelligence (AI) is its ability to make uncertainty visible. By making clear to refugee status decision-makers how uncertain their predictions are, AI and related statistical tools could help to reduce their confidence in their conclusions. As it now stands, this would only hurt claimants, since countries around the world have designed their refugee status determination systems to resolve decision-making uncertainty at the claimant's expense. Increasing uncertainty would therefore contribute to mistaken rejections. If, however, international refugee law was to recognize an obligation under the UN Convention to resolve decision-making doubt in the claimant's favour, as Evans Cameron (2018) has recently advocated, then by making uncertainty visible, AI could help ensure that fewer refugees were wrongly denied the protection they need.

AI and making uncertainty visible

Artificial intelligence is a tool that can encourage the explicit consideration of uncertainty. While there are many aspects of AI, the technologies that have brought about AI's recent prominence are best understood as advances in machine learning, a form of prediction technology (Agrawal, Gans, and Goldfarb 2018). AI is perhaps the most widely-reported of a cluster of data-focused technologies that have recently gained commercial interest. These statistical tools also include data mining, data science, and natural language processing and together likely represent a general purpose technology (Goldfarb, Taska, and Teodoridis 2021). They already provide decision support in a wide variety of applications from healthcare to entertainment (Topal 2019). In what follows, we will refer to these statistical tools as 'AI' or 'machine prediction', because the implementation of what we describe will often use machine learning, the prediction technology most commonly labeled 'AI' (specifically, natural language processing and image recognition).

Used properly, AI provide both an estimate of the most likely result as well as a measure of uncertainty. Statistically, it provides both a mean and the distribution of likely outcomes around that mean. Given the sparse data and uncertain environment in refugee claims, many machine predictions in this context are likely to generate wide distributions. Put differently, AI, as a type of statistical tool, will make the inherent uncertainty explicit.

Much anxiety about the use of AI in legal proceedings generally—and in immigration and refugee law in particular—emphasizes lack of transparency, accountability, and the potential for bias when decisions are made by algorithms (Desai and Kroll 2011; Mayson 2019; Molnar and Gill 2018). We argue that under theoretically ideal circumstances using AI in refugee claims could increase transparency and accountability while reducing biases that lead to rejected claims. We emphasize a different mechanism than others who have advocated for the use of algorithms and machine learning in the legal system. For example, Kleinberg et al (2019) argue that algorithms create opportunities to detect discrimination and increase the transparency of competing values. Our argument emphasizes that AI algorithms are statistical predictions, and therefore these algorithms provide explicit acknowledgment of uncertainty. If the law resolved decision-making doubt in favour of refugee protection, AI would increase protection of the vulnerable.

We recognize that for this to work, we would need to overcome a major legal challenge, various political constraints, as well as a number of technical, ethical, and administrative challenges, some of the most salient of which are highlighted below. If these challenges could be overcome, AI could facilitate a main recommendation of Evans Cameron (2018): resolving doubt in the claimant's favour.

Artificial intelligence and decision-support

One definition of AI is “a truly intelligent machine that can do everything humans can do” (Pethokoukis 2018). This definition refers to strong AI or artificial general intelligence (AGI), which can outperform humans in most cognitive tasks, reason through problems, as well as invent and improve itself. While there is substantial debate that considers when, if ever, such an AGI might be developed, it is important to recognize that the recent attention to AI in the press and in academia is much narrower and has little relation to AGI (for a discussion see Agrawal, Gans, and Goldfarb 2018). Instead, the current attention to AI is driven by advances in a particular branch of computer science known as machine learning. Related to computational statistics, machine learning is a prediction technology where prediction is defined as the process of filling in missing information.

This new information is given as a probability, or likelihood, that can be used to inform a decision. The prediction itself does not suggest a decision. Consider the decision of whether to take an umbrella when going for a walk (Agrawal, Gans, and Goldfarb 2018, p. 78). A weather forecast might state that there is a 70% chance of rain. Inherent in this prediction is a degree of uncertainty based on the quality of the data and the ability for the prediction model to learn from feedback. This uncertainty is present even if the prediction is stated as a relative certainty. With a 70% chance of rain, local weather forecasters might

claim “it will rain today” and even put 100% in their forecast (Silver 2012). Taking the prediction without uncertainty at face value, most people will take the umbrella.

Now consider what will happen with the accurate prediction with the recognition that there is a 30% chance that it does not rain. In this case, more people might not bother with the umbrella. The decision will depend on their judgment of how unpleasant it is to carry an umbrella relative to how unpleasant it is to get wet. No matter the prediction, such judgment is needed to make a decision. That judgment determines the relative importance of false positives (taking the umbrella when it does not rain) and false negatives (leaving the umbrella but it does rain). The decision involves a trade-off between the false positive of carrying an umbrella unnecessarily and the false negative of getting wet. In the extreme, it is possible to eliminate all false positives by using a constant decision rule of never taking the umbrella or eliminate all false negatives by always taking the umbrella. Using AI for decision support enables more precise decision-making. It is useful because the predictions allow for contingent decision-making, such as if the prediction is more than—say—90% rain, then take the umbrella. Otherwise, leave the umbrella at home. In other words, you need to determine precisely the consequences of the potential outcomes including which outcome would be worse if the prediction was incorrect.

While the decision to carry an umbrella is trivial, it illustrates well-understood ideas from a rich literature in decision theory (Gilboa 2010; Binmore 2011; Peterson 2013). In the context of the refugee claimant decision process, this example demonstrates that decisions require a prediction and an assessment of the relative value of different types of mistakes.

Much of the attention to AI focuses on its role in automating decisions. For automation, the judgment of the consequences of the different outcomes needs to be specified in advance. For example, if rain is higher than 90% certain then take the umbrella. Otherwise, leave it at home. In many decision contexts, such fixed rules are difficult to implement. There might be extenuating factors that a decision maker wishes to consider on a case-by-case basis, or legal obligations to keep a human in the loop. In such cases, AI can be used as a decision support tool in which a human is given the AI’s prediction, and then the human makes the final decision (Jamieson and Goldfarb 2019). We argue that under ideal theoretical circumstances AI could play such a role in the refugee claims process.

AI, like other statistical tools, uses data to generate predictions. As we detail below, this represents both a big challenge and a big opportunity in using AI for decision support in refugee claims. The challenge is that relevant data is often missing. The opportunity is that AI could force the decision-maker to confront

the lack of available information explicitly. As discussed below, if the law were changed to resolve decision-making doubt in the claimant's favour, then having an AI that makes clear the extent of the uncertainty in the evidence should increase the claims accepted.

An example from medicine: The primary purpose of initial triage is to differentiate patients that are critically ill from those that are stable. The Babylon Triage and Diagnostic System is an AI-powered triage and diagnostic system. Building the Babylon system involved a variety of types of training data, including connections between diseases, symptoms, and risk factors from medical experts and prior probabilities in epidemiological data (Razzaki *et al.* 2018). Once the model was built, it was improved through feedback. In particular, the model was given clinical vignettes, or cases, and its performance was evaluated and scrutinized by scientists and doctors. This feedback was then used to further train the AI.

This system has been able to identify the condition modelled by a clinical vignette with accuracy comparable to human doctors based on precision and recall. In brief, the system provided higher precision at the expense of lower recall, and safer triage recommendations (97% versus 93.1%) at the expense of a marginally lower appropriateness (90% versus 90.5%). Thus, under the conditions of a trial analysis, the recommendations of this system were, on average, safer than that of human doctors (Razzaki *et al.* 2018, p. 5-6).

Despite these results, this tool is not a replacement for doctors; it is limited to a diagnostic task within the full decision-making process for the doctor. The system provides predictions that can advise staff and facilitate accurate diagnosis to triage patients safely. A human makes the final decision on the next step for each patient (Razzaki *et al.* 2018).

To integrate into the human decision-making process, the information provided must be easily accessible and interpretable. For instance, there is ample evidence that physicians faced with binary decisions, even those with dedicated statistical training, have poor comprehension of basic statistical measures relevant to healthcare decisions (Jamieson and Goldfarb 2019). Effective AI triage systems must therefore communicate both the prediction and the associated uncertainty inherent in that prediction in ways that are easily interpreted by the human decision-maker.

The communication of uncertainty is fundamental. Uncertainty is high in initial triage. It is conducted before a patient receives diagnostic tests. A well-designed AI triage system should explicitly address this uncertainty. For example, the machine should skew to higher levels of triage severity when uncertainty is

high (Berlyand *et al.* 2018). In this way, a cautionary approach in the emergency department can be taken to sensitize predictions toward outcomes that would be harmful to the patient, such as acute and delayed cardiac complication in patients with chest pain. In the Babylon study, there was considerable disagreement between the subjective ratings of human doctors (Razzaki *et al.* 2018, p. 8). Akin to many problems where uncertainty is present and no ‘gold standard’ exists to achieve accurate decisions, personal assessments, preferences, and error influence human predictions. Machine prediction may be able to assist in these cases by skewing toward cautionary approaches.

Uncertainty and doubt in refugee status determination

In discussing the role of error-preference within the law of fact-finding in *Refugee Law’s Fact-finding Crisis: Truth, Risk, and the Wrong Mistake*, Evans Cameron refers variously to ‘resolving uncertainty’ ‘resolving doubt’ and ‘tipping the balance,’ using these terms interchangeably. The present thought experiment has made clear the need to define these and related concepts more precisely.

In judgment problems that require a decision-maker to choose between two conclusions (to accept or reject an allegation, to accept a theory or its counter-theory), *uncertainty* is the condition of being undecided about which conclusion or conclusions can be drawn from the evidence. *Epistemic paralysis* is the condition of being unable to draw either conclusion from the evidence because the decision-maker is equally uncertain of both conclusions. A legal system will *resolve uncertainty* in favour of a party when its fact-finding structures (its burdens of proof, standards of proof and presumptions) function such that, all else being equal, a greater percentage of errors will be made in that party’s favour than at their expense (Evans Cameron 2018, Chapter 1). ‘All else being equal’ assumes that most decisions are not made under conditions of epistemic paralysis, that only a small minority of decisions will require a tie-breaker. Put another way, legal epistemology assumes that the standard of proof will resolve uncertainty in most cases and the burden of proof exists for the rare cases in which it is required.

The *error burden* is the cost of uncertainty to a party: the heavier their burden, the more likely it is that they will “pay the price” if they have their allegation or theory rejected on account of the decision-maker’s uncertainty (Gaskins 1992; Scott 2005). Knowledge of this error burden is the judgment discussed above in the context of AI. A system that resolves uncertainty in a party’s favour may yet force that party to carry a heavy error burden. This is why the relevant normative question is not merely *whether* the claimant should ‘pay the price’ but “whether *and to what extent*” they should (Evans Cameron 2018, p. 7). For a start, a system that ensures that a greater percentage of errors favour the claimant (e.g. 51%),

nonetheless allows a high percentage of errors at her expense (49%). Moreover, even if the system resolves uncertainty in a party's favour with a low standard of proof, in practice the party may carry a heavy error burden if what they are expected to establish is too difficult to prove: "Since the standard of proof is the threshold at which the law resolves uncertainty, [a low standard of proof] will not assist the claimant if, at the end of the day, because of an overly heavy burden, the decision-maker has very little uncertainty left to resolve" (Evans Cameron 2018, p. 197). Most fundamentally, the party with the legal onus will always bear the full weight of the error burden in cases of epistemic paralysis. This is, after all, the express function of the legal onus: it breaks a tie. And crucially, while a system can resolve uncertainty in favour of the party that bears the onus, as described below, such a system will deprive that party of more and more of the practical benefit of its error preference as more and more of its decisions are made at the point of epistemic paralysis.

Doubt is broader than uncertainty. Whereas uncertainty is not only an inherent element of the reasoning process but indeed its initial starting point – as a rule, our system requires that every decision-maker begin the judging process "without any notion of who has the better case" (Posner 1999) – doubt encompasses the full range of affective responses to feelings of not knowing. Whereas uncertainty is about indecision, doubt is about a lack of conviction. A civil juror who is prepared to find a defendant liable on the civil law's "balance of probabilities" standard is no longer uncertain in a legal sense. They may nonetheless have reasonable doubts about whether their conclusion is correct.

Whereas the civil law has decided that those doubts should not stand in the way of a finding of liability, in a criminal court not only uncertainty but also doubt must resolve in favour of the accused. A system will *resolve doubt* in the party's favour when that party's claim or theory will in practice pay less of a price for the decision-maker's doubts than the opposing claim or counter-theory. To do this, it must both resolve uncertainty in their favour and sufficiently lighten their error burden. The more its decisions are made under conditions of epistemic paralysis, the harder it will be for a system to resolve doubt in favour of the party that bears the onus.

To see this, imagine what would happen to the parties' respective error burdens if, instead of requiring the Crown to prove guilt against a very high threshold, our criminal legal system required the accused prove their innocence against a very low threshold – to prove, for example, that there is "more than a mere possibility" that they are innocent. Since we assume that the standard of proof governs decision-making in the large majority of cases, in the large majority of cases both systems would resolve uncertainty

squarely in favour of the accused. But if the accused bore the onus, they would nonetheless pay the full price for uncertainty in cases of epistemic paralysis. In any such case, this heavy error burden would have the effect of depriving the accused of the benefit of the low standard of proof. This is why the Supreme Court of Canada made clear that the accused's *Charter*-protected right to be presumed innocent requires that both uncertainty and epistemic paralysis resolve in their favour: the standard of proof must be one that squarely resolves uncertainty in their favour, and the Crown must bear the burden of proof (*R. v Oakes* 1986). Only in this way can the criminal legal system truly resolve doubt in the accused's favour.

The Refugee Convention put in place a very low standard of proof. To qualify as a refugee, a claimant must have a "well founded fear of persecution." In Canada, to meet this standard, a claimant must show that there exists "more than a mere possibility" that they will face persecution if returned home (Evans Cameron 2018, p. 82-85).

On its face, this very low standard of proof – one of the lowest anywhere in law – would suggest that the Convention's drafters intended for decision-makers to resolve their doubts in the claimant's favour. On this point, however, the Convention's drafters do not have the last word: "By allowing state parties to establish their own fact-finding procedures, in accordance with their own legal traditions, the drafters invited them to be part of the conversation about what this obstacle course should look like." (Evans Cameron 2018, p. 177) And when it comes to the design of this obstacle course, Canadian refugee law has not made up its mind. One body of law will often resolve uncertainty in the claimant's favour and it imposes a lighter error burden. The other will often resolve uncertainty at the claimant's expense and it imposes a much heavier error burden. Both, however, fail to resolve doubt in the claimant's favour. This is because both require claimants not only to prove, in the final analysis, that they face 'more than a mere possibility' of persecution, but also, as a preliminary step, that each one of their factual allegations is 'more likely than not' to be true. This intervening civil standard of proof by its very nature resolves doubt against the party that bears the onus (Evans Cameron 2018, Chapter 5), an effect that is further amplified by several unique features of this decision-making context.

A central question in many refugee claims often concerns the possibility of a triggering precondition to the harm that the claimant fears. In a significant body of judgments, the Court finds that this should be treated a question of fact – and so answered against the higher 'balance of probabilities' standard – rather than as part and parcel of the risk analysis itself, and so answered on the lower 'more than a mere possibility' standard. On this reasoning, a gay claimant from a country that persecutes sexual minorities

cannot win his case merely by showing that there is ‘more than a mere possibility’ that his state authorities will discover his orientation. He must prove that it is ‘more likely than not’ that they will discover it. If a claimant can prove that her ex-husband is actively seeking her and will almost certainly kill her if he finds her, it is not enough for her to show that there is ‘more than a mere possibility’ that he will be able to locate her. She must prove on a balance of probabilities that he will. The same reasoning applies to arguments about the ineffectiveness of police protection: the claimant must prove that the police will not adequately protect her. It is not enough for her merely to raise serious doubts about whether they will be willing and able to help her. Evans Cameron (2018, Chapter 8) provides a critique of this reasoning.

Moreover, many if not most refugee status decisions are, if not actually made under conditions of epistemic paralysis, at least justifiable on the grounds of epistemic paralysis. It has been widely commented that refugee hearings are a paradigm example of decision-making under conditions of “radical uncertainty” (Luker 2013; Kagan 2003). Refugee status decision-makers face a host of unique fact-finding challenges. There are typically no witnesses in refugee hearings and few if any supporting documents, and adjudicators’ assumptions about how people think and act may be of limited use when they are judging a person from a different culture, of a different gender, who is suffering the aftereffects of trauma and giving evidence through an interpreter. Furthermore, as one decision-maker wrote, “we never know if the assessment was right or wrong” (Care 2001, as cited in Evans Cameron 2018, p. 33).

As a result, whether or not decision-makers are actually uncertain of their findings, in nearly any refugee claim, epistemic paralysis is readily available to a decision-maker who may want to take advantage of the legal onus. This is because a decision-maker can almost always plausibly assert that the evidence is fatally unclear. A large literature explores the politicized nature of refugee status decisions and the potential for credibility assessment, in particular, to be influenced by a decision-maker’s unconscious bias (Hersh 2015; Rehaag, Beaudoin, and Danch 2015; Keith and Holmes 2009; Evans Cameron 2018, Chapter 3) or to be used strategically to create “rejections by design” (Noll 2005; Bohmer and Shuman 2007; Hamlin 2014; Zyfi and Atak 2018). In the current Canadian context, a decision-maker looking to make a negative decision – or nothing but negative decisions (Rehaag 2017) – need only say “I am left in doubt” in order to avail themselves of the system’s permission to resolve that doubt at the claimant’s expense. In conditions of ‘radical uncertainty,’ such doubt is not hard to come by.

When and how AI could be used in refugee status determination

Pre-requisite: A new mode of legal reasoning

This paper suggests that AI and other statistical tools could contribute to reducing mistaken rejections by making visible to decision-makers the uncertainty inherent in their judgments and thereby shaking any undue confidence in their conclusions. But a lack of confidence – doubt – cannot help claimants within a system that resolves doubt at their expense. For this paper’s approach to succeed, the law would have to resolve doubt in the claimant’s favour.

Evans Cameron has argued that the current system, in addition to resolving doubt at the claimant’s expense, in fact does not even allow decision-makers to assess properly whether claimants are at risk. Instead, it requires them to assess whether claimants *have established* that they are at risk, in a legal proceeding governed by the logic of inductive inference. Evans Cameron (2018, Chapter 8) suggests that this form of legal reasoning profoundly distorts risk assessment. She has proposed a risk-assessment model of refugee status decision-making based on a different mode of reasoning, abductive inference or ‘inference to the best explanation’. Such a model would have another advantage: if implemented along with AI, machine prediction could further reduce the risk of mistaken rejections.

Any legal system governed by the logic of inductive inference needs a tie-breaker to resolve equally balanced uncertainty. The onus does this by removing the uncertainty from one half of the equation. If a jury member in a criminal trial is feeling paralyzed – “I’m not certain enough that he did it; but I’m also not certain enough that he didn’t” – the onus says: “Forget your second set of uncertainties, they are not relevant. Since the Crown has the onus, only your first set of uncertainties is relevant. If you are not certain enough that he did it, he is not guilty. Full stop.” In other words, when epistemic paralysis brings the onus into play, one set of uncertainties evaporates.

A central feature of abductive inference is that all doubts remain in play throughout the decision-making process, right up until the moment of decision. In an abductive model, a decision-maker compares the theory being put forward and the most compelling counter-theory and decides which is more persuasive. To give claimants the benefit of the doubt, Evans Cameron’s proposed model would have the decision-maker accept the claimant’s theory unless the most compelling counter-theory is decidedly more persuasive.

On such a model, the claimant retains the burden of proof and “[w]hat she must prove remains the same – that she has a well-founded fear.” What has changed is “how she must prove it...by showing that this theory explains the available evidence at least as plausibly as any counter-theory” (Evans Cameron 2018, p. 439) If in the final analysis the decision-maker is not convinced, the claimant loses. But an abductive model does not require a legal onus as a separate tie-breaking structure because the situation that leads to epistemic paralysis in an inductive model – when the decision-maker is equally uncertain of either conclusion – does not lead to paralysis in this model. If the decision-maker is equally uncertain of either conclusion, the claimant wins.

To see how doubt would resolve differently on this model, imagine that the claimant has convinced the decision-maker that her ex-husband, who wants to kill her, will be able to find her if she returns. The only remaining question is whether the police will protect her. On the law as it stands, the claimant must prove that the police will not protect her. If evidence on this point is scarce, and the claimant cannot show convincingly that the police will not protect her, she loses. Because the claimant bears the legal onus, the only doubts that matter are doubts about the assertion that the claimant is trying to prove, that the police *will not* protect her. Any doubts about the contrary assertion – doubts about whether the police *will* protect her – are legally irrelevant. On an abductive model, all doubts are relevant. If evidence is very scarce, the member may conclude that both assertions are equally uncertain. In which case, the claimant has met her burden of proof.

If refugee status decisions were made using an abductive model of reasoning that resolves doubt squarely in the claimant’s favour, uncertainty about the counter-theory would always be relevant. Put simply, whatever the merits of the claimant’s case, the more uncertain the counter-theory, the greater the claimant’s chances of winning. Under such a model, making uncertainty visible would indeed help to resolve doubt in the claimant’s favour, as the drafters of the Convention intended.

Yet any move to enshrine the Convention’s error preference in law will inevitably meet with strong opposition. The Convention itself is facing existential challenges, and even among those nations, like Canada, that remain committed to its foundational principles, sharp and desperate debates turn on this very question of the wrong mistake. On the one side are those who believe that the international refugee protection system is in crisis because too few people are getting the protection that they need (e.g. McAdam 2017; Simeon 2003; Dauvergne 2008; Lewis 2012). On the other are those who believe that it is in crisis because too many people are getting status that they do not deserve. Among the latter are those

who fear that mistaken grants cost a host state financially and politically and endanger its security; encourage false claims and reward liars; allow its citizens to be played for fools; and poison the environment for other would-be immigrants and genuine refugees (Stoffman 2008; Francis 2002; critiqued in Evans Cameron 2019). These concerns have dogged the international refugee protection project from the start and are gaining momentum world-wide. Against this backdrop, advocating for a system that resolves doubt in the claimant's favour will face significant socio-political barriers. But it remains both a legal and an ethical obligation.

How AI would work:

The role of AI is to make doubt explicit. Poor quality data results in imprecise predictions. Well-executed AI systems, like other well-executed statistical decision processes, make this imprecision precise. This well-defined imprecision provides the opportunity for AI to resolve doubt in a claimant's favour within an abductive model of reasoning.

There are significant legal and normative constraints on the ability to use AI to predict whether a particular claimant will come to harm. In Canada, and in other common law jurisdictions, expert opinion is only admissible in a court proceeding in limited circumstances. The Canadian Supreme Court has made clear that while experts may provide an informed opinion that will assist the decision-maker to understand the evidence before them, they must "not be permitted to usurp the functions of the trier of fact" and may only give testimony "in effect of the ultimate issue" in a case under tightly conscribed conditions (*R v Mohan*, [1994] SCJ No 36 at para 24; *R v Bingley*, [2017] SCJ No 12 at para 48; see e.g. Houle & Peterson (2018)). Although administrative tribunals are typically not bound by technical rules of evidence, the concerns that gave rise to these restrictions at common law would be as relevant in a tribunal context. Moreover, in a tribunal that adjudicates matters involving fundamental human rights, such concerns would arguably impose constitutional constraints on the admission of evidence that risked 'usurping the function' of the decision-maker (See e.g. Stewart (2008)).

An AI system would therefore not be positioned to advise an adjudicator directly about the likelihood that a claimant faces a serious risk. It could advise on intermediate issues arising in a claim, however, if it assumes that the claimant's factual allegations in the case are true. For example, an AI could predict how the police in a given country will respond to an appeal for help or how a government will react to a certain kind of activism. This would likely involve some combination of natural language processing and standard statistics.

To be able to function, such an AI will require data. Broadly, there are two key types of data. First, there is the input data used at the moment of a prediction in order to generate that prediction. For example, in order to generate a prediction on the likelihood that a particular police force responds adequately to a particular type of appeal for help, the input data would be all the information available about the specific appeal for help and police force in a given case. This data will be available as it will form part of the evidence on record in that claim. Second, there is the training data used to build the predictive algorithm, for example information about similar appeals for help to that particular police force, and whether that police force responded adequately. In many cases, such data does not exist. Training data will therefore be limited and of lower quality than in many other machine prediction contexts: it is difficult to know the outcomes of information that is relevant to refugee cases.

This lack of outcome data represents the main challenge for AI in refugee law, but also the main opportunity. Given the large number of relevant variables in the individual details of any particular case, the training data will usually be sparse and incomplete and the AI system's predictions therefore will lack confidence. It is this transparent inaccuracy that gives AI systems an advantage over human-only systems. A well-executed AI system that is built on limited data will make the inherent uncertainty in this decision-making exercise explicit. It would acknowledge the uncertainty associated with small samples and missing data (Manski 2013). Limited and biased data means that predictions are highly uncertain, whether those predictions are made by machines or humans. Imprecision does not mean zero information. It is difficult, but not impossible, to know relevant information in refugee cases with confidence. This is the key insight in Manski's framework. Missing and biased data can still improve decision-making, so long as the uncertainty is recognized and the data contain some relevant information.

As noted above, this lack of information and feedback is well-recognized by the human adjudicators. Nevertheless, this poor-quality data also results in poor predictions from humans (Evans Cameron 2018; Jamieson and Goldfarb 2019).

In recent years, some Canadian adjudicators have accepted all, or nearly all, of the claims that they have heard. Others have rejected every one. And the same adjudicator will sometimes decide very similar cases differently. One recently reached opposite conclusions in two hearings held hours apart, on the same package of evidence, for members of the same family, who feared the same people, for the same reasons (Evans Cameron 2018, p. 2-3).

This variance is not unique to Canada or to refugee claims. In a wide variety of contexts, decisions by legal adjudicators have been found to be highly variable (Bambauer 2018; Danziger, Levav, and Avnaim-Pesso 2011; Rehaag 2018; Storey 2013). Even when provided with detailed and reliable information, decision-makers may be unable to digest the details of a specific case.

The Canadian Immigration and Refugee Board has a Research Directorate that produces comprehensive publicly accessible information by country. However, this information is organized under broad headings and can be upwards of 7000 or more pages (Government of Canada 2018, p. 70). In these cases, it is easy to understand how individuals may leverage heuristics to make quicker decisions given the severe time constraints faced by the adjudicators. A large literature explores biases that arise from such heuristics, and some of that literature proposes the use of algorithms and other technology to generate consistency (Sunstein *et al.* 2001; Kleinberg *et al.* 2019).

Machines can be better than humans at the prediction task because AI aggregates data across many cases, and then accounts for the lack of reliable outcome data. This means that, when faced with poor data, a well-designed AI will generate a prediction that recognizes the high level of uncertainty. The opportunity for AI in refugee claims is similar to the opportunity in triage, discussed above. It has potential to be a decision-support tool that enables a decision-maker to be cautious in the face of uncertainty.

With little data available, the machine will likely make inaccurate point predictions (Smith 2016). In contrast to humans, however, it is possible for the machine to have an explicit measure of the inaccuracy of the predictions. In other words, tools exist for specifying the uncertainty and making effective decisions, even in the presence of missing or biased data (Manski 2013). This recognition of the distribution of possible outcomes generates an opportunity for machine prediction when data are sparse. For example the AI would explicitly note that there is a great deal of uncertainty as to whether the police will respond to an appeal. As previously mentioned, humans often overweight salient information and fail to properly account for uncertainty (Tversky and Kahneman 1974). The formal specification of uncertainty creates an opportunity for AI to improve decision-making that is distinct from arguments for reduced bias or increased fairness that rely on rich data and accurate predictions.

As noted above, AI will be useful in generating predictions about the threat to the claimants assuming that the facts of the case are true. The AI would provide decision support in the form of a prediction of the likelihood that—taking the facts of a particular case as given—those facts would support a relevant legal inference, as discussed below.

Importantly, the AI would not identify high levels of uncertainty in all cases. A well-executed AI could affirm that some predictions in fact involve very little uncertainty. In two recent cases, for example, the Board rejected claims by US citizens with delusional disorders fearing persecution by various US government agencies (MB5-05930 and MB5-05913). In another case in which one of the authors was involved, a claimant from Germany feared persecution by the administration at her son's school and claimed that the German police would not assist her. Leaving aside other legal considerations, the adjudicators in these cases would have had very little doubt about their findings that these claimants did not face a serious risk. Similarly, an AI trained on information about the practices of the relevant US government agencies and the responsiveness of the German police to reports of criminality would likely make the same prediction as the adjudicators and with a comparably high level of confidence: there is a sufficiently high degree of certainty attached to the prediction about how these actors would act.

In other cases, adjudicators would have comparably high levels of confidence in their decisions to grant a claim. Provided they accept the claimant's identity, an adjudicator could quite confidently grant the claim of a Kurdish political activist from Turkey, for example, or an LGBTQ activist from Yemen, or a young man from a persecuted minority in Sudan. Indeed, in 2019, the Board accepted 97% of all of the claims, on all grounds, brought by citizens of these countries (Rehaag, 2019). An AI could be trained on relevant data to reach a similarly confident prediction on a number of intermediate issues arising in such claims.

Very many cases, however, will fall in between and will leave much room for doubt. In the experience of one of the authors of this paper, who represented refugee claimants for a decade, there is often conflicting or inadequate information about how a police force responds to appeals for protection from domestic violence. Similarly, there may be little information to help establish whether a claimant's profile is sufficient to bring them to the attention of the agents of persecution. Some governments tolerate no dissent of any kind, while others crack down primarily on their most active opponents. Often no information is available that would make clear how much of a thorn in the government's side a person must be before there is enough of a risk that they will be targeted. Furthermore, some time will have passed since the claimant left their country, and it is often unclear whether an agent of persecution who posed a real risk to the claimant in the past would still be interested in harming them. In some cases, claimants may have the option of relocating to another area of their country, and it may be unclear whether the proposed area is beyond their persecutor's reach.

A well-designed AI would make such uncertainty explicit. Even when the actor is a known quantity, such as a government, guerrilla group, crime syndicate, or police force, relevant data may be partial or conflicting. Such uncertainty about the particular context is common. To work as intended, an AI for refugee claims would provide predictions and the associated uncertainty for only a small subset of all refugee claimants. The uncertainty would be high for most predictions but the machine prediction would still provide information by showing low uncertainty in some cases.

Analogously, Manski (2019, p. 56-57) emphasizes the usefulness of clinical trial data for some types of patients even when the outcomes of many trial participants are missing for reasons that may affect the results in unknown ways. As long as the information is precise enough for one patient group, a treatment can be recommended for that group though perhaps not for the others. Put differently, in the refugee status determination context, as long as there is some useful information, we see the lack of data as an opportunity not a challenge. Currently, human decision makers make predictions and act on them with conviction using no more data, and often less data, than what would be available to a machine. The opportunity for a decision support tool is to make this uncertainty transparent. Some cases would generate precise predictions about the context of the case from which the adjudicator would conclude that the case ought to be rejected. These would be cases in which adjudicators would very likely have reached the same conclusions from their own reading of the evidence. In many other cases, designing a machine that can explicitly communicate with human decision-makers how unreliable a prediction could be may help decision-makers revise the types of convictions that lead to harming an individual if false. In providing transparency by emphasizing the level of uncertainty, machine prediction could reduce bias that occurs in the form of overconfidence from decision-makers.

In these cases, although the predictions will likely be imprecise, an indication of uncertainty can skew decision-making toward a cautionary approach. Effective decision-making in the presence of uncertainty requires explicit acknowledgment of the wrong mistake. A false positive, accepting a claim that is not well-founded, has different consequences from a false negative, rejecting the claim of a Convention refugee. Evans Cameron emphasizes that in refugee status decision-making, there is a wrong mistake: false positives are less consequential than false negatives. Effective use of AI for refugee claims will require adjudicators to have a precise understanding of how much less consequential false positives are compared to false negatives. While it is possible to have a constant decision-rule to accept every claim, AI and related statistical tools can recognize the uncertainty inherent in the predictions of harm while still enabling the rejection of claims in the rare circumstances when predictions are precise.

As refugee claims are complex situations, decision-making should remain a human task. Machine predictions can aid decision-makers by providing an understanding of the uncertainty inherent in predictions of harm.

Challenges to the use of AI in refugee status determination

Despite this opportunity, there are a number of challenges that need to be overcome before an AI could be used as a decision-support tool in refugee claims. Applying recently developed AI tools to refugee status determination systems will require fundamentally rethinking the legal structures that govern this kind of decision-making, as well as overcoming other legal, technical, social, and regulatory challenges. It is only if these challenges could be solved that AI would improve decision-making in the refugee claims process in Canada and elsewhere.

The above argument has emphasized the challenge posed by legal structures. As it stands, in Canada and in many other countries around the world, AI and related statistical tools will not help. Unless the law that governs refugee status decision-making decisively resolves doubt in the claimant's favour, an AI that makes uncertainty clear will only increase the likelihood that decision-makers will make the "wrong mistake" and refuse a claim that should have been accepted. And as noted above, in the current socio-political climate, any move to implement a system that honours the Convention's error preference will face significant barriers.

In the context of our argument, the biggest technical challenge will be to build a user interface that clearly communicates the level of uncertainty. The opportunity lies in the recognition that both humans and machines lack data about many outcomes in the refugee context. The user interface needs to communicate to decision-makers the level of uncertainty. Furthermore, the decision-makers need to receive enough training so that they can use the decision support tool properly.

Many of the other challenges have been discussed elsewhere, including challenges in gathering data, bias, and transparency (Harcourt 2006; Hannah-Moffat 2013; Starr 2014; Chouldechova 2016; Garcia 2017; Berk *et al.* 2018; Huq 2019; Pettitt, Townhead, and Huber 2008). It is important to collect enough data so that the AI can provide confident predictions where such confidence is warranted. Otherwise, the AI would not be necessary. It would be more effective to recognize the uncertainty and accept every claim. Even though we expect that machine predictions will reduce the number of rejected claims and that there will be a human making the final decision, the AI will nevertheless influence decisions. Therefore, a variety

of administrative law concerns would need to be considered, including procedural fairness, the right to be heard, the right to a fair, impartial, and independent decision-maker, right to reasons, right of appeal, and substantive review (Molnar and Gill 2018, p. 47-54).

Finally, even if these technical, legal, and administrative issues could be overcome, a change in the system that led to a substantial decrease in the number of rejected claims could potentially lead to other challenges, including political pressure, an increase in the number of refugee claimants, and a decrease in public trust in the system.

Conclusion

While there are a number of legal, socio-political, technical, ethical, and administrative challenges to overcome, a well-designed machine prediction tool could make it clear that in many refugee cases there is insufficient information to know much for sure. This AI tool—like other statistical tools—would inform the decision-maker when there are strong reasons to doubt their predictions about important aspects of a claim. Combined with a change in law that mandated that the decision-maker should resolve their doubts in the claimant's favour, this recognition of uncertainty should lead to a reduction in wrongly rejected claims.

References

- AGRAWAL, A., GANS, J., and GOLDFARB, A.** (2018) *Prediction Machines* (Boston: Harvard Business Review Press).
- BAMBAUER, J.** (2018). ‘Machine Influencers and Decision Makers’. Available at <https://uascience.org/lectures/machine-influencers-and-decision-makers/>
- BERK, R., HEIDARI, H., JABBARI, S., KEARNS, M., and ROTH, A.** (2018) ‘Fairness in Criminal Justice Risk Assessments’. *Sociological Methods & Research*.
- BERLYAND, Y., RAJA, A.S., DORNER, S.C., PRABHAKAR, A.M., SONIS, J.D., GOTTUMUKKALA, R.V., SUCCI, M.D.S and YUN, B.J.** (2018) ‘How Artificial Intelligence Could Transform Emergency Department Operations’. *The American Journal of Emergency Medicine* 36 (8): 1515-1517.
- BINMORE, K.** (2011) *Rational Decisions* (Princeton: Princeton University Press).
- BOHMER, C. and SHUMAN, A.** (2007) *Rejecting Refugees: Political Asylum in the 21st Century*, (London: Routledge).
- CARE, G.** (2001) ‘The Refugee Determination Process: A Judge’s View’ cited in Evans Cameron 2018, p. 33.
- CBC/RADIO CANADA.** (2019) “It’s a Crapshoot”: Asylum Seekers Fret over Fateful Day at Canada’s Immigration Board | CBC Radio.” CBCnews. April 26, 2019. Available at <https://www.cbc.ca/radio/outintheopen/asylum-seekers-1.5095969/it-s-a-crapshoot-asylum-seekers-fret-over-fateful-day-at-canada-s-immigration-board-1.5112314>.

- CHOULDECHOVA, A.** (2016) 'Fair prediction with disparate impact: a study of bias in recidivism prediction instruments'. *BIG DATA* 5(2): 153-163.
- COLAIACOVO, I.** (2018) 'Not Just the Facts: Adjudicator Bias and Decisions of the Immigration and Refugee Board of Canada (2006-2011)'. *Journal on Migration and Human Security* 1(4): 122-147.
- DANZIGER, S., LEVAV, J., and AVNAIM-PESSO, L.** (2011) 'Extraneous factors in judicial decisions'. *Proceedings of the National Academy of Sciences* 108(17): 6889-6892.
- DAUVERGNE, C.** (2008) *Making People Illegal: What Globalization Means for Migration and Law* (Cambridge: Cambridge University Press).
- DESAI, D.R. and KROLL, J.A.** (2011) 'Trust but Verify: A Guide to Algorithms and the Law'. *Harvard Journal of Law and Technology* 31(1): 1-64.
- EVANS CAMERON, H.** (2008) "Risk Theory and 'Subjective Fear': The Role of Risk Perception, Assessment, and Management in Refugee Status Determinations". *International Journal of Refugee Law* 20(4): 567-585.
- EVANS CAMERON, H.** (2018) *Refugee Law's Fact-finding Crisis: Truth, Risk, and the Wrong Mistake* (Cambridge: Cambridge University Press).
- EVANS CAMERON, H.** (2019) 'The Battle for the Wrong Mistake: Error Preference and Risk Salience in Canadian Refugee Status Decision-making'. *42 Dalhousie Law Journal* 1
- FRANCIS, D.** (2002) *Immigration: The Economic Case* (Toronto: Key Porter Books) 116-123.
- GARCIA, M.** (2017) 'Racist in the Machine: The Disturbing Implications of Algorithmic Bias'. *World Policy Journal* 33(4): 111-117.
- GASKINS, R.H.** (1992) *Burdens of Proof in Modern Discourse* (New Haven: Yale University Press).
- GILBOA, I.** (2010) *Making Better Decisions: Decision Theory in Practice* (West Sussex: Wiley-Blackwell).
- GOLDFARB, A., TASKA, B., and TEODORIDIS, F.** (2021) 'Could machine learning be a general purpose technology? A comparison of emerging technologies using data from online job postings'. Working paper, University of Toronto. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3468822
- GOVERNMENT OF CANADA** (2018) 'Report of the Independent Review of the Immigration and Refugee Board: A Systems Management Approach to Asylum'.
- GRIFFIN, D. and TVERSKY, A.** (1992) 'The weighing of evidence and the determinants of confidence'. *Cognitive Psychology* 24(3): 411-435.
- HAMLIN, R.** (2014) *Let Me Be a Refugee: Administrative Justice and the Politics of Asylum in the United States, Canada, and Australia* (New York: Oxford University Press).
- HANNAH-MOFFAT, K.** (2013) 'Actuarial Sentencing: An "Unsettled" Proposition'. *Justice Quarterly* 30(2): 270-296.
- HARCOURT, B.E.** (2006) *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (Chicago: University of Chicago Press).
- HERSH, N.** (2015) 'Challenges to Assessing Same-Sex Relationships under Refugee Law in Canada'. *McGill Law Journal* 60(3): 381-594.
- HOULE, F. and PETERSON, C.** (2018) Hors de tout doute raisonnable: La méthodologie et l'adéquation empirique comme fondements de l'épistémologie du droit de la preuve" (Montréal: Les Éditions Thémis)
- HUQ, A.** (2019) 'Racial equity in algorithmic criminal justice'. *Duke Law Journal* 68(6): 1043-1134.
- JAMIESON, T. and GOLDFARB, A.** (2019) 'Clinical Considerations When Applying Machine Learning to Decision-Support Tasks versus Automation'. *BMJ Quality & Safety* 28: 778-81.
- KAGAN, M.** (2003) 'Is Truth in the Eye of the Beholder - Objective Credibility Assessment in Refugee Status Determination'. *Georgetown Immigration Law Journal* 17(3): 367-415.
- KAHNEMAN, D.** (2011) *Thinking Fast and Slow* (New York: Farrar, Straus and Giroux).
- KEITH, L.C. and HOLMES, J.S.** (2009) 'A Rare Examination of Typically Unobservable Factors in US Asylum Decisions'. *Journal of Refugee Studies* 22(2): 224-241.

- KLEINBERG, J., LUDWIG, J., MULLAINATHAN, S., and SUNSTEIN, C.R.** (2019) 'Discrimination in the Age of Algorithms'. *Journal of Legal Analysis* 10: 113–174.
- LEWIS, C.** (2012) *UNHCR and International Refugee Law: From Treaties to Innovation* (London: Routledge).
- LUKER, T.** (2013) 'Decision Making Conditioned by Radical Uncertainty: Credibility Assessment at the Australian Refugee Review Tribunal'. *International Journal of Refugee Law* 25(3): 502-534.
- MANSKI, C.F.** (2013) *Public Policy in an Uncertain World*. (Cambridge MA: Harvard University Press).
- MANSKI, C.F.** (2019) *Patient Care Under Uncertainty*. (Princeton: Princeton University Press).
- MAYSON, S.G.** (2019) 'Bias In, Bias Out'. *Yale Law Journal* 128(8): 2122-2473.
- McADAM, J.** (2017). 'The Enduring Relevance of the 1951 Refugee Convention'. (2017) 29 *International Journal of Refugee Law* 1
- MOLNAR, P. and GILL, L.** (2018) *Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System* (Toronto: The Citizen Lab).
- NOLL, G.** (2005) 'Introduction: Re-Mapping Evidentiary Assessment in Asylum Procedures' in Gregor Noll, ed, *Proof, Evidentiary Assessment and Credibility in Asylum Procedures* (Boston: Martinus Nijhoff Publishers).
- PETERSON, M.** (2013) *An Introduction to Decision Theory* (Cambridge UK: Cambridge University Press).
- PETHOKOUKIS, J.** (2018) "Nobel laureate Daniel Kahneman on AI: 'It's very difficult to imagine that with sufficient data there will remain things that only humans can do'." *American Enterprise Institute*, accessed July 20, 2019 from <http://www.aei.org/publication/nobel-laureate-daniel-kahneman-on-a-i-its-very-difficult-to-imagine-that-with-sufficient-data-there-will-remain-things-that-only-humans-can-do/>
- PETTITT, J., TOWNHEAD, L., and HUBER, S.** (2008) 'The Use of COI in the Refugee Status Determination Process in the UK: Looking Back, Reaching Forward'. *Refuge: Canada's Journal on Refugees* 25(2): 182-194.
- POSNER, R.A.** (1999) *An Economic Approach to the Law of Evidence* (Chicago: Chicago Unbound).
- R v OAKES** [1986, para 32] 1 SCR 103.
- RAZZAKI, S., BAKER, A., PEROV, Y., MIDDLETON, K., BAXTER, J., MULLARKEY, D., SANGAR, D., TALIERCO, M., BUTT, M., MAJEEED, A., et al.** (2018) 'A Comparative Study of Artificial Intelligence and Human Doctors for the Purpose of Triage and Diagnosis'. *Babylon Health*: 1-15.
- REHAAG, S., BEAUDOIN, J., and DANCH, J.** (2015) 'No Refuge: Hungarian Romani Refugee Claimants in Canada'. *Osgoode Hall Law Journal* 52(3): 705-771.
- REHAAG, S.** (2017) "'I simply do not believe': A case study of credibility determinations in Canadian Refugee Adjudication". *Windsor Review of Legal and Social Issues* 38: 38-70.
- REHAAG, S.** (2018). 'Judicial Review of Refugee Determinations (II): Revisiting the Luck of the Draw'. *Queen's Law Journal*: 1-23.
- RUSSO, J.E. and SCHOEMAKER, P.J.** (1992) 'Managing Overconfidence'. *Sloan Management Review* 33(2): 7-17.
- SCOTT, D.N.** (2005) 'Shifting the Burden of Proof: The Precautionary Principle and Its Potential for the Democratization of Risk' in Law Commission of Canada, ed, *Law & Risk* (Vancouver: UBC Press).
- STEWART, H.** (2008) 'Section 7 of the Charter and the Common Law Rules of Evidence', 40 *SCLR* (2d) 415 – 437
- SILVER, N.** (2012) *The Signal and the Noise: Why so many predictions fail—but some don't* (New York: The Penguin Press).
- SIMEON, J.C.** (2003) 'Introduction: searching for ways to enhance the UNHCR's capacity to supervise international refugee law' in James C. Simeon (ed.) *The UNHCR and the Supervision of International Refugee Law* (Cambridge: Cambridge University Press).
- SMITH, R.E.** (2016) 'Idealizations of uncertainty, and lessons from artificial intelligence'. *Economics* 10(7): 1-40.

- STARR, S.B.** (2014) 'Evidence-based sentencing and the scientific rationalization of discrimination'. *Stanford Law Review* 66(4): 803–872.
- STOFFMAN, D.** (2008) 'Truths and Myths about Immigration' in Alexander Moens & Martin Collacott (eds.) *Immigration Policy and the Terrorist Threat in Canada and the United States* (Fraser Institute: www.thefraserinstitute.org).
- STOREY, H.** (2013) 'Consistency in Refugee Decision-making: A Judicial Perspective'. *Refugee Survey Quarterly* 32(4): 112-125.
- SUNSTEIN, C.R., KAHNEMAN, D., RITOV, L., and SCHKADE, D.** (2001) 'Predictably Incoherent Judgments'. *Chicago Unbound* 54(131): 1153-1215.
- TOPAL, E.** (2019) *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (New York: Basic Books).
- TVERSKY, A. and KAHNEMAN, D.** (1974) 'Judgment under Uncertainty: Heuristics and Biases'. *Science* 185(4157): 1124-1131.
- ZYFI, J. and ATAK, I.** (2018) 'Playing with lives under the guise of fair play: the safe country of origin policy in the EU and Canada'. *International Journal of Migration and Border Studies* 4(4): 345-365.