Luck of the Draw III: Using AI to Extract Data About Decision-Making in Federal Court Stays of Removal

*Sean Rehaag**

This article examines decision-making in Federal Court of Canada immigration law applications for stays of removal, focusing on how the rates at which stays are granted depend on which justice decides the case. The article deploys a form of computational natural language processing, using a large-language model machine learning process (GPT-3) to extract data from online Federal Court dockets. The article reviews patterns in outcomes in thousands of stay of removal applications identified through this process and reveals a wide range in stay grant rates across many justices. The article argues that the Federal Court should take measures to encourage more consistency in stay decision-making and cautions against relying heavily on stays of removal to ensure that deportation complies with constitutional procedural justice protections. The article is also a demonstration of how machine learning can be used to pursue empirical legal research projects that would have been cost prohibitive or technically challenging only a few years ago—and shows how technology that is increasingly used to enhance the power of the state at the expense of marginalized migrants can instead be used to scrutinize legal decision-making in the immigration law field, hopefully in ways that enhance the rights of migrants. The article also

* Director of the Centre for Refugee Studies, Director of the Refugee Law Laboratory, Associate Professor at Osgoode Hall Law School, York University. This article draws on research supported by the Social Sciences and Humanities Research Council, the Law Foundation of Ontario, and the Digital Research Alliance of Canada. The author is grateful for the able research assistance provided by Krystyna Jones (who assisted with data verification and copy editing), Jacob Danovich (who programmed the web-scraping technology that this study used), Simon Wallace (who provided helpful advice on using machine learning models to extract data from the dockets and feedback on various drafts), and Sarah Dawson (who assisted with statistical analysis). He also appreciates feedback provided by Talia Joundi and Pierre-André Thériault, both of whose ongoing research was the main inspiration for this paper. He is also grateful for helpful feedback on a draft from Jonathan Khan and Petra Molnar, as well as for assistance with statistical analysis from Michael Friendly, through York University Statistical Consulting Services. The time between when the research for this paper was undertaken (November & December 2022) and when the final version of the article was revised for publication (January 2024) saw rapid advances and an explosion of interest in generative artificial intelligence. The article reflects technology available publicly when the research was undertaken and has not been updated with reference to more recently released generative artificial intelligence models.

contributes to the broader field of computational legal research in Canada by making available to other non-commercial researchers the code used for the project, as well as a dataset of several thousand Federal Court dockets that can be used for future research.

Introduction

In Canada's deportation regime, the final procedure available to stop removal once all other recourses have been exhausted is to apply to the Federal Court for a stay of removal. These applications are typically heard on an expedited basis, days or even hours before individuals are scheduled to be put on a plane.

Everyone involved in stays of removal, including lawyers for individuals and the Department of Justice, as well as the Federal Court justice deciding the issue, are confronted with challenging tasks. They have little time to prepare or review materials, hearings tend to be short, and decisions must be made quickly. Often, individuals who are about to be removed are experiencing a crisis—many in immigration detention—with all the attendant difficulties that poses. Levels of stress for counsel can be high, which can make it difficult to secure counsel, as many lawyers are understandably disinclined to take on these cases.[1]

These challenges combine with the enormous stakes of stay of removal decision-making. Individuals applying for stays assert that their removal would result in irreparable harm, sometimes including persecution, torture, or even death.[2] Recent jurisprudence has also placed increasing weight on stay proceedings as a site for ensuring that the deportation regime complies with the *Canadian Charter of Rights and Freedoms* (*Charter*).[3] This means that the stakes are high, not just for individuals, but also for the legal system.

While some immigration and refugee law procedures have attracted substantial scholarly attention, there is comparatively little research on stays of removal. This is likely due to methodological challenges. Until recently, decisions in Federal Court stays of removal were mostly unpublished, which made traditional legal doctrinal analysis difficult. Happily, in 2018, the Federal Court began publishing most stay decisions. Several scholars are in the process of completing research projects on this new body of published stay decisions.[4]

This article offers a quantitative overview of all Federal Court stay of removal decision-making over the past ten years to help provide context for ongoing research on the new body of published stay decisions. It does so through computational legal research methods that build on two prior

---

1. For a discussion of the deportation process and some of the challenges that it causes for those involved, see Kathryn Tomko Dennler & Brianna Garneau, *Deporting Refugees: Hidden Injustice in Canada* (Toronto: Romero House, 2022) at 29–30, online: <romerohouse.org> [perma.cc/XF82-DE4D].

2. As discussed below, to succeed with an application for a stay of removal, an applicant must demonstrate that they face irreparable harm if they are deported. See *infra* note 26 and accompanying text.

3. Some recent jurisprudence in the immigration law setting arguably suggests that section 7 of the *Canadian Charter of Rights and Freedoms*, Part I of the *Constitution Act*, *1982*, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11 [*Charter*] is not engaged in most immigration and refugee procedures on the theory that those procedures do not necessarily directly result in removal—with the moment that the section 7 is engaged being pushed closer and closer to the moment of removal. See generally Colin Grey, "Thinkable: The Charter and Refugee Law After Appulonappa and B010" (2016) 76 SCLR (2nd) 111 at 136–38; Gerald Heckman, "Revisiting the Application of Section 7 of the *Charter* in Immigration and Refugee Protection" (2017) 68 UNBLJ 312 at 312–13; Asha Kaushal, "The Webbing of Public Law: Looking Through Deportation Doctrine" (2022) 59:2 Osgoode Hall LJ 291 at 313.

4. See e.g. Pierre-André Thériault, "2020 stay dockets" (11 November 2022) via e-mail [communicated to author]; Talia Joundi, *Unfortunate but Ordinary: A Study of Federal Court Approaches to Stays of Removal* (LLM Thesis, Osgoode Hall Law School, 2023) [unpublished].

research projects about the luck of the draw in Federal Court decision-making.[5] Specifically, the article leverages a form of computational natural language processing using a large language model machine learning process (GPT-3) to extract data from online Federal Court dockets. It then reviews patterns in outcomes in thousands of stay of removal applications identified through this process.

The patterns in Federal Court stay of removal applications revealed through this research appear to be troubling. From 2012 to 2021, the rates at which stays of removal were granted varied dramatically depending on which justice was assigned to hear the case: some justices deciding large numbers of cases granted stays over 80% of the time, while others granted stays less than 10% of the time. The concern is not just with outliers. Rather, there was a wide range in stay grant rates across many justices. In other words, there appears to be a large unexplained variance in stay of removal grant rates depending on which justice decides the application—similar to the findings of large unexplained variance in rates at which different justices granted leave in refugee law applications for judicial review highlighted in the prior two Luck of the Draw studies.[6]

This finding points to the possibility of inconsistent and arbitrary outcomes in high-stakes decision-making. Absent some kind of reasonable explanation for the variance in stay grant rates across justices (e.g., the Federal Court reveals that cases are assigned to particular justices after some sort of screening that would result in some justices hearing, on average, stronger or weaker claims), the Federal Court should take measures to encourage more consistency in stay decision-making. In the meantime, it is risky for courts to rely heavily on stays of removal to ensure that removal complies with constitutional procedural justice concerns because, in practice, this form of time-pressured decision-making appears to generate results that raise concerns about arbitrary outcomes that vary depending on which justice decides the case.

The article also aims to demonstrate how technological development in the field of artificial intelligence—and specifically large language models accessed through simple application programming interfaces—is now sufficiently accessible that legal scholars with modest coding skills can (and should) pursue empirical research projects that would have been cost-prohibitive or technically challenging only a few years ago. There are many concerns about how such technology is being deployed in asymmetrical and rights-limiting

5. Sean Rehaag, "Judicial Review of Refugee Determinations: The Luck of the Draw?" (2012) 38:1 Queen's LJ 1 [Rehaag, "Luck of the Draw I"]; Sean Rehaag, "Judicial Review of Refugee Determinations (II): Revisiting the Luck of the Draw" (2019) 45:1 Queen's LJ 1 [Rehaag, "Luck of the Draw II"].

6. *Ibid*.

ways, especially in the border control setting where this technology has been used to limit access to asylum, to facilitate removals, and to otherwise enhance the power of the state at the expense of (mostly racialized) migrants. However, this article demonstrates that this same technology can be deployed in ways that shift the object of scrutiny from the movement of marginalized people to flaws in human-made legal decision-making, thereby increasing transparency and potentially enhancing access to rights. As this article argues, the main barrier to such rights-enhancing use of machine learning is that access to bulk legal data is currently restricted largely to commercial actors and lawyers representing clients with deep pockets, such as Department of Justice lawyers. Stay of removal decision-making is an excellent example of this problem: the terms of service of the websites where these decisions are published all prohibit bulk and programmatic access—meaning that if one wants to study stay of removal decision-making at scale, one must find methodologies that do not rely on published decisions. This study does so by extracting data from court metadata (specifically online court dockets) that are made available by the Federal Court on a website that does not prohibit programmatic access. So, in addition to making the code used for this project publicly available to assist non-commercial researchers with other similar projects, another output of this project is a large dataset of hundreds of thousands of Federal Court dockets that is being made available for non-commercial researchers.

The key contributions of this article, then, are: (1) to describe a method of deploying machine learning tools to extract legal data at scale from online court metadata in a context where working programmatically with the text of decisions themselves is impossible, and to share the code used for the research so that it can be used by other researchers; (2) to share the data on hundreds of thousands of Federal Court online dockets for use in other research; and (3) to explore patterns in a form of legal decision-making that is high stakes, high volume, and understudied.

The article begins by briefly setting out the law and process for applications for stays of removal for readers who may be unfamiliar with this area of law. Next, the article describes the research methodology used for this study. Then the article sets out the findings of the study. Finally, the article offers several recommendations and conclusions.

I. Context

Canada's deportation regime involves many different processes through which non-citizens may seek to remain in the country. These include refugee claims, pre-removal risk assessments, Immigration and Refugee Board processes to contest inadmissibility, humanitarian and compassionate applications, requests for temporary resident permits, and requests for administrative deferrals of removal—as well as administrative appeals for some of these procedures and judicial review. While any given non-citizen will likely not have access to all these processes, most (although not all) do have access to some kind of process to challenge their removal. In some of these processes, individuals benefit from not being removable while the process is ongoing, but for others, removal can occur even while the matter is pending.[7]

When all the other processes to which an individual has recourse that prevents or delays removal have been exhausted, the final process available is to apply for a stay of removal from the Federal Court. Stays of removal are a form of interlocutory relief. That is, a motion for a stay of removal is not a freestanding application to remain in Canada. Rather, it is a request for an injunction to delay removal pending the outcome of some other process. For example, a person who is not entitled to an automatic stay of removal might seek a stay of removal pending the determination of a judicial review of a refugee decision or pending a judicial review of an administrative deferral of removal request.

The *Immigration and Refugee Protection Act*[8] (*IRPA*) and the *Federal Courts Act*[9] (*FCA*) provide the Federal Court with jurisdiction over judicial reviews of immigration and refugee law matters. The *IRPA* recognizes that people cannot be removed from Canada in contravention of a Federal Court stay order.[10] The *Immigration and Refugee Protection Regulations* (*IRPR*) also outlines some circumstances in which pending Federal Court procedures automatically stay removal.[11] However, neither the *IRPA* nor the *IRPR* offers explicit guidance about Federal Court applications for stays of removal in circumstances where automatic stays are unavailable. The *FCA* also does not specifically refer to stays of removal. However, the *FCA* does provide broad jurisdiction to the Federal Court over federal administrative law judicial reviews,[12] including the authority to make any interim orders that it considers appropriate pending the disposition of such judicial reviews.[13] The *FCA* also provides the Federal Court general powers to grant "an injunction . . . in all cases in which it appears to the court to be just or convenient to do so".[14]

7. For an overview of these various immigration law procedures, see Jamie Chai Yun Liew & Donald Galloway, *Immigration Law*, 2nd ed (Toronto: Irwin Books, 2015). See also Dennler & Garneau, *supra* note 1.

8. *Immigration and Refugee Protection Act*, SC 2001, c 27, ss 72–75 [*IRPA*].

9. *Federal Courts Act*, RSC 1985, c F-7, ss 17–18 [*FCA*].

10. *IRPA*, *supra* note 8, s 50(a).

11. *Immigration and Refugee Protection Regulations*, SOR/2002-227, s 231 [*IRPR*].

12. *FCA*, *supra* note 9, ss 17–18.

13. *Ibid*, s 18.2.

14. *Ibid*, s 44.

Both the *Federal Court Rules*[15] (*FCR*) and the *Federal Court Citizenship, Immigration and Refugee Protection Rules*[16] are silent about applications for stays of removal—and the latter does not address motions or other forms of interlocutory relief. The *FCR* does, however, address motions for interlocutory relief in detail.[17] Moreover, the Federal Court has issued practice directions that explain how those rules apply specifically to motions for stays of removal.[18]

Generally, applications for stays of removal proceed as follows. Applicants must first file an underlying application for judicial review of an immigration decision or have such a judicial review already in process. Applicants must then generally file and serve a notice of motion for a stay of removal three days prior to when they propose that the Federal Court hear the matter.[19] Where applicants cannot comply with this timeline—such as where applicants were only informed about removal at the last minute and where the need for a stay motion could not have reasonably been anticipated in advance—applicants can apply for urgent motions asking the Court to exercise its discretion to hold a special hearing with less than three days notice.[20] Given the short timelines, the parties are expected to keep the record reasonably brief (i.e., under 100 pages absent exceptional circumstances).[21] Hearings are also typically brief, generally around one hour, though sometimes hearings can be only a few minutes, and sometimes they extend several hours. Justices can either issue orders immediately at the conclusion of the hearing or they can reserve their decisions. Even when decisions are reserved, they are typically issued the same day or the next day. In their orders, justices will usually include brief reasons, though they are not required to do so.[22]

15. *Federal Courts Rules*, SOR/98-106 [*FCR*].

16. *Federal Courts Citizenship, Immigration and Refugee Protection Rules*, SOR/93-22.

17. *FCR*, *supra* note 15, rs 358–70.

18. Federal Court of Canada, *Consolidated Practice Directions, Consolidated Practice Guidelines for Citizenship, Immigration and Refugee Proceedings* (FC, 2022) at 5–8 online: <fct-cf.gc.ca> [perma.cc/3QEA-BK92] [*FC Practice Directions*].

19. *FCR*, *supra* note 15, r 362(1).

20. *Ibid*, rs 35(2), 362(2). For a discussion of when it would be appropriate for the Court to exercise its discretion to hear urgent stay of removal motions, see *FC Practice Directions*, *supra* note 18 at 7–8.

21. *FC Practice Directions*, *supra* note 18 at 6.

22. The *FCR* provide that judges "may" provide reasons for judgment, but they may also simply sign an order (see *FCR*, *supra* note 15, rs 392–93). For a discussion on obligations to provide reasons in Federal Court immigration law judicial reviews, see Sean Rehaag & Pierre-André Thériault, "Judgments v Reasons in Federal Court Refugee Claim Judicial Reviews: A Bad Precedent?" (2022) 45:1 Dal LJ 185 at 188–91.

When considering whether to provide requested stays of removal, the Court applies the tripartite test for interlocutory injunctions as articulated by the Supreme Court of Canada in *RJR-MacDonald Inc v Canada (Attorney General)*[23] and by the Federal Court of Appeal in the immigration context in *Toth v Canada (Minister of Employment and Immigration)*.[24] The test generally requires applicants for stays of removal to demonstrate: (1) a prima facie case that their underlying application for judicial review will be successful (sometimes referred to as the "serious issue" branch of the test);[25] (2) that if the stay of removal is not granted they will face irreparable harm; and (3) that all things considered the application should be granted because the balance of inconvenience favours the applicant.[26] If the applicant demonstrates that the test is met, the Court can exercise its discretion to grant an injunction delaying or preventing removal.

There is, to date, little published scholarly research about applications for stays of removal or about how the Court applies the tripartite test in such applications.[27] The reason for this is mostly methodological and relates to difficulties in accessing stay of removal decisions.

---

23. 1994 CanLII 117 (SCC) [*RJR*]. See also *R v Canadian Broadcasting Corp*, 2018 SCC 5.

24. 1988 CanLII 1420 (FCA) [*Toth*].

25. See *Wang v Canada (Minister of Citizenship and Immigration)*, 2001 FCT 148 at paras 9, 11 [*Wang*]. Note that where the application for a stay is based on a judicial review of a denied request for a deferral of removal (rather than in connection with a judicial review of a different immigration decision), instead of applying the standard serious issue test, the Court applies a "likelihood of success" test. The reason for this is that the relief sought in the stay is the same relief sought in the underlying judicial review, and thus, the test should be the same at both the stay and the judicial review stages.

26. *RJR*, *supra* note 23 at 314–15; *Toth*, *supra* note 24. For examples of the Federal Court applying this test in the stay of removal context, see e.g. the two most recent stay of removal decisions posted on CanLII that were granted and dismissed as of the time of writing: *Olanipekun v Canada (Citizenship and Immigration)*, 2022 CanLII 109757 (FC) (stay of removal dismissed); *Singh v Canada (Citizenship and Immigration)*, 2022 CanLII 109758 (FC) (stay of removal dismissed); *Tagari v Canada (Citizenship and Immigration)*, 2022 CanLII 105340 (FC) (stay of removal granted); *Ngabo v Canada (Citizenship and Immigration)*, 2022 CanLII 106391 (FC) (stay of removal granted).

27. Andrew D Little, Dominique T Hussey & Shelby Morrison, "Injunctions, Stays and Mandatory Orders: Lessons from Recent Federal Court Decisions" (2021) 51:3 Adv Q 386.

For many years, reasons issued in stay decisions were mostly unpublished.[28] In recent years, the failure to publish most stay of removal decisions generated pushback from members of the immigration and refugee bar.[29] The main concern was that failing to publish stay decisions resulted in asymmetrical and unfair access to jurisprudence.[30] Because the government is a party in all litigation involving stays of removal, Department of Justice lawyers had easy access to all unpublished stay decisions. In contrast, lawyers for individuals could only access the decisions if they knew that the decisions existed and they obtained copies of the decisions from the Federal Court registrar. These decisions could be relied upon as precedent in subsequent cases, which gave Department of Justice lawyers an unfair advantage in litigating stays of removal because they could selectively bring to the Court's attention precedents that supported their positions.[31]

Responding to these concerns, in 2018, the Federal Court issued a notice to the profession indicating that, going forward, the Court will assign a neutral citation to stay of removal decisions that the Court views as having precedential value and post those decisions on the Court's website (leading other legal publishers including CanLII to also publish them), and will also provide all other stay of removal decisions to CanLII for publication.[32] As a result, most stay decisions are now available on CanLII.[33]

While publishing stay of removal decisions on CanLII is a step in the right direction, there are ongoing concerns about fair access to stay of removal decisions. Currently, decisions are not equally accessible to everyone because most are not given a neutral citation, translated, and posted on the Federal Court's website. This is a problem for two reasons.

First, by publishing decisions via CanLII only in the decisions' original language, rather than on the Federal Court's website where the Court would be required to publish the decisions in both official languages, the Court makes it difficult for people who practice primarily in one official language to

28. For example, a search in CanLII's Federal Court database for the term "stay of removal" in cases decided in 2017 (i.e., the year before the change in publication practices by the Federal Court) leads to 39 results. By contrast, a search for the same term for cases in 2019 (i.e., a year after the change in publication practices) leads to 233 results—an increase of over 500%.

29. Rehaag & Thériault, *supra* note 22 at 210.

30. *Ibid*.

31. *Ibid* at 210, 213–16.

32. Federal Court of Canada, *Notice to the Parties and the Profession: Publication of Court Decisions* (FC, 2018) online (pdf): <fct-cf.gc.ca> [perma.cc/7XSX-R4SR].

33. CanLII, "Federal Court – Canada (Federal)" (n.d.), online: <canlii.org/en/ca/fct> [perma.cc/EF6F-F8EX].

access relevant precedents fully. Because most decisions are written in English, this disproportionately impacts lawyers who practice in French.[34]

Second, because of restrictive terms of service on CanLII, researchers cannot access these decisions in bulk.[35] In contrast, the Federal Court's website has no such limitation and allows full non-commercial reproduction of the cases it makes available.[36] The choice to publish stay of removal decisions only on CanLII means that the sorts of computational methods that are described later in this article (which made use of the Federal Court's permissive terms of service) cannot be applied to these decisions. This not only limits research on this important form of decision-making but also mimics, at a bulk level, the problem that led the Court to publish the decisions on CanLII to begin with: because the Department of Justice is involved in all stay applications, it has access to the decisions in bulk without going through CanLII, meaning that it can employ computational research methods on stay of removal decisions while those outside the Department of Justice cannot.[37]

---

34. This practice also raises the troubling possibility that the Federal Court, and indeed other federal courts, could use this reasoning to entirely circumvent their language rights obligations: *Official Languages Act*, RSC 1985, c 31 (4th Supp), s 20. If courts can avoid language rights obligations in stay of removal decisions by providing them without translation to third party publishers, what is to stop them from doing the same with all other cases? For a discussion of concerns over compliance with language rights obligations in the federal courts, see Parliamentary Information and Research Service, *Bilingualism in Canada's Court System: The Role of the Federal Government*, by Marie-Ève Hudon, Publication No 2017-33-E (Ottawa: Library of Parliament, 26 November 2020), online: <lop.parl.ca> [perma.cc/G99N-LZMR].

35. CanLII, "Terms of Use" (last modified 18 October 2023), online: <canlii.org/en/info/terms.html> [perma.cc/P8CN-5LSX] (prohibiting "bulk or systematic downloading of documents, including via programmatic means").

36. Federal Court of Canada, *Important Notices* (last modified 6 November 2020), online: <fct-cf.gc.ca> [perma.cc/X7MD-PXXP] (indicating that, subject to certain requirements such as ensuring accuracy and attribution, information on the Court's website is being made available for "personal and public non-commercial use and may be reproduced, in part or in whole, and by any means, without charge or further permission, unless otherwise specified").

37. For further discussion about why all Federal Court immigration law decisions should be made available in bulk to non-commercial researchers (and about how to best protect privacy while ensuring transparency and fair access), see Jonathan Khan & Sean Rehaag, "Promoting Privacy, Fairness, and the Open Court Principle in Immigration and Refugee Proceedings" (2023) 54:2 Ottawa L Rev 357. For a discussion of Department of Justice efforts to leverage bulk data to build computational tools in the immigration setting, see Petra Molnar & Lex Gill, *Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System* (Toronto: University of Toronto CitizenLab, 2018) at 15–16, online (pdf): <citizenlab.ca> [perma.cc/872P-UCMB].

At any rate, while fair access to these decisions has not been fully achieved, it is nonetheless the case that due to the Federal Court's revised practice of providing its stay decisions to CanLII for publication, it is now much easier to research stay of removal decision-making. Several scholars are currently in the process of studying this body of decisions.[38] One of the aims of this article is to assist with that research by providing a quantitative overview of stay decision-making by applying computational methods to Federal Court dockets involving stays of removal, in the absence of being able to apply such methods to the decisions themselves due to the limited CanLII terms of service.

II. Methodology

This study uses computational legal research methodologies with online Federal Court dockets to gain insight into patterns in Federal Court stay of removal applications. This methodology was selected for several reasons. First, preliminary results from research by another scholar involving interviews with lawyers who prepare stay of removal applications raised concerns about consistency in how the tripartite test for injunctive relief has been applied to stay of removal applications—suggesting that a review of a large number of stay decisions over a long period of time would be helpful in identifying whether those concerns are well founded.[39] Second, if one wishes to study stay decision-making over a long period, one must confront the problem identified in the prior section about how the Federal Court failed to systematically publish these decisions until 2018. So, to understand how stay decision-making occurred before that time, we need to go beyond reviewing published cases. Third is the problem of volume: as demonstrated in more detail below, each year the Federal Court decides hundreds of applications for stays of removal. For the years when these decisions are available, it would be work intensive to manually review each of these decisions to examine patterns in decision-making. Fourth, unfortunately, using computational methods to extract data from published stay decisions as an alternative to manual reviews of published stay decisions faces the problem identified in the prior section: the Federal Court does not publish these decisions directly but has, since 2018, provided them only to CanLII, and CanLII's terms of service prohibit bulk access. As a result, it is not currently possible to apply computational methods to study published stay decisions (at least it is not possible to do so in compliance with the terms of service of CanLII). Fifth, and finally, even if one did somehow access published stay decisions in bulk, they may be of limited use in computational methods because many stay decisions are prepared very quickly and contain virtually no

38. Thériault, *supra* note 4; Joundi, *supra* note 4.

39. Joundi, *supra* note 4.

details—including, for example, decisions which do not even include the country of removal of the applicant, let alone the main factors that were considered in applying the test for a stay.[40]

To address all of these challenges, the current study uses an innovative methodology involving several steps. First, all data publicly available online in the Federal Court docket database was systematically gathered for all Federal Court files over a ten-year period. Next, machine learning language models were created and applied to classify and extract information from the gathered docket entries. Additional code was then applied to infer case-level data using the classifications and extracted information. Data verification was undertaken to ensure the accuracy of the resulting dataset. Finally, statistical analysis was undertaken on the dataset.

The data, including the scraped dockets and human-coded training datasets, as well as the code (other than the web-scraping code) used in this project, has been made available for non-commercial use by other researchers.[41] The code is written in Python, a free, open-source programming language that is comparatively easy to learn and popular among data scientists.[42] The code takes the format of a Jupyter Notebook, an open-source interactive browser-based programming environment used by many data scientists for exploratory research.[43] The Jupyter Notebook for this project includes instructions to assist other researchers who would like to build on this work.

A. Web-Scraping Using Python

The first step in this project was obtaining the full text of the relevant Federal Court online dockets. These online dockets offer a table of contents for each immigration and refugee law application for judicial review in the Federal Court.[44] Appendix A shows an example of a Federal Court docket with a stay application.

40. See e.g. *Jalloh v Canada (Citizenship and Immigration)*, 2023 CanLII 123002 (FC); *Rana v Canada (Citizenship and Immigration)*, 2021 CanLII 134610 (FC); *Linadi v Canada (Immigration, Refugees and Citizenship)*, 2021 FC 1389.

41. Sean Rehaag, "Luck of the Draw III: Code & Data" (last modified 23 February 2024), online: <github.com> [perma.cc/J688-FKW9] [Rehaag, "Luck of the Draw III: Code & Data"].

42. Python Software Foundation, "Python" (n.d.), online: <python.org> [perma.cc/4YNB-DRRV]. For an introduction to Python, see Al Sweigart, *Automate the Boring Stuff: Practical Programming for Total Beginners*, 2nd ed (San Francisco: No Starch Press, 2020), online: <automatetheboringstuff.com> [perma.cc/PN45-CAYB].

43. Project Jupyter, "Jupyter", online: <jupyter.org> [perma.cc/P3LE-8TLN]. See also Cyrille Rossant, *IPython Interactive Computing and Visualization Cookbook*, 2nd ed (Birmingham, UK: Packt, 2018), online: <ipython-books.github.io> [perma.cc/P6DT-YSMR].

44. Federal Court of Canada, *Court Files*, online: <fct-cf.gc.ca> [perma.cc/W5SM-DSV3].

As can be seen in Appendix A, Federal Court online dockets include some structured data, such as fields for how the Federal Court categorized the application (i.e., what is the type of immigration decision being contested in the application for judicial review) and the date and office where the application for judicial review was first filed. In addition, the dockets include a table that lists each step taken and each document filed in the case (including the date the step was taken or the document was filed). Thus, for example, there will be a row in the table when an application for judicial review is filed, another when the applicant perfects the application, another if there is a notice of a motion for a stay of removal, another if either party files materials for use in adjudicating that motion, another if there is a hearing on the motion, another if an order is issued following the hearing of the motion, and so on. The entries in each row are written in natural language, in either French or English.[45] They vary in terms of how they are entered, though most follow specific patterns.

Note that throughout the remainder of this article, a "docket" will refer to all of the information available on the Federal Court online docket webpage for a given case, and a "docket entry" will refer to one row of the table available in a given docket.

Using a custom web-scraping program,[46] all immigration law online Federal Court dockets filed from 2012 to 2022 were downloaded and saved to a database that is up to date as of December 1, 2022. The full dataset involves 87,776 dockets. The database of dockets is being made freely available online for non-commercial use to facilitate further research.[47]

B. Docket and Docket Entry Screening Using Regex

Once the full dataset of dockets was downloaded, the next step was to extract information from the natural language docket entries using a computer program written by the author in Python using Jupyter Notebooks.[48] The

45. Any given docket can include docket entries in either official language. These are not translated into the other language.

46. The web-scraping program was coded by Jacob Danovich, a data scientist working with York University's Refugee Law Laboratory. It is written in Python and uses compute resources provided by the Digital Resource Alliance of Canada. The program was written as part of a Law Foundation of Ontario funded project to assist refugee lawyers in drawing on data from the Federal Court and Immigration and Refugee Board. The program is informed by earlier web-scraping of Federal Court dockets. See e.g. Samuel Norris, "Examiner Inconsistency: Evidence from Refugee Appeals" (2019) Becker Friedman Institute for Economics, Working Paper No 2018-75, online: <papers.ssrn.com> [perma.cc/C2AB-H89S]; Rehaag, "Luck of the Draw II", *supra* note 5.

47. Rehaag, "Luck of the Draw III: Code & Data", *supra* note 41.

48. See *supra* notes 42, 43.

specifically targeted information included (a) whether a case involved a motion for a stay of removal, (b) whether the Federal Court issued an order on the motion, (c) which justice issued the order, and (d) whether the motion was granted or denied. The dates of motions, hearings and orders were also recorded.

To obtain this information, each docket was first screened using Regexes,[49] which is a Python package that helps to search text strings programmatically. Because dockets involving stays of removal will almost always include certain terms (e.g., "stay", "removal", "*sursis*", "*renvoi*"), this process quickly eliminated most dockets that were not relevant to the project because the dockets did not include these terms.

Next, individual docket entries were also screened out or screened in using Regexes if they included or excluded phrases that were relevant to extract targeted information. For example, docket entries involving notices of appearances, application records, confirmation of service, books of authorities, etc., often start with specific phrases which can be easily identified. Such docket entries do not contain useful information for this project, so they were excluded. By contrast, docket entries describing orders and written directions from the Court usually begin with particular terms, so docket entries beginning with those terms of interest for this study could be screened in.

Taken together, this Regex screening produced a dataset of 7,045 screened in dockets that included 188,584 docket entries, of which 25,069 docket entries were screened-in for further review.

C. Docket Entry Categorization and Data Extraction Using GPT-3

Next, machine learning tools were applied to each screened in docket entry to identify which docket entries involved a notice of motion for a stay of removal and which involved orders relating to stays of removal. The docket entries categorized as involving stay of removal orders were subject to further machine learning tools to identify the justice and the outcome.

The machine learning models employed for this project were built on OpenAI's GPT-3 platform.[50]

OpenAI began as a non-profit organization devoted to research on artificial intelligence "unconstrained by a need to generate financial return".[51] It aimed to become "a leading research institution which can prioritize a good

49. Python Software Foundation, "re — Regular expression operations" (23 February 2024), online: <docs.python.org> [perma.cc/UGP6-S9GC].

50. Tom B Brown et al, "Language Models are Few-Shot Learners" in *Advances in Neural Information Processing Systems 33* (NeurIPS, 2020) 1877. See also OpenAI, "Introduction", online: <beta.openai.com> [perma.cc/64UB-5Z73].

51. OpenAI, "Introducing OpenAI" (11 December 2015), online: <openai.com> [perma.cc/EAA5-RA4A] [OpenAI, "Introducing OpenAI"].

outcome for all over its own self-interest" and was committed to freely sharing "papers, blog posts, . . . code, and . . . patents . . . with the world".[52] OpenAI has undertaken research on artificial intelligence in many areas,[53] but it is most well known for its work on Generative Pre-trained Transformers (GPTs).

GPTs are machine learning models using neural networks—specifically transformers—that are pre-trained on large quantities of text from the Internet. The initial training is unsupervised, meaning that the system does not use data labelled by human beings and then tries to match that labelling. Instead, the task that model is trained on is to predict (or calculate the probability of) the next word or sequence of text after any given sequence of text in the massive dataset of text it uses. This form of training makes GPTs particularly well-suited to generating predicted sequences of words based on an inputted prompt. GPTs can also be fine-tuned—that is, given examples of pairs of prompts and desired completions and then asked to generate similar responses to new prompts.[54]

OpenAI released early versions of its GPT models on an open-source basis.[55] For example, GPT-2, released in February 2019, is, at the time of publication in 2024, still among the most frequently used open-source large language models.[56] However, in March 2019, OpenAI changed its corporate structure and began developing products through a new for-profit company, OpenAI LP.[57] It also received substantial infusions of cash and technical resources from other tech companies.[58]

Shortly after creating this new corporate structure, OpenAI announced that it had built GPT-3,[59] which was also the first model that OpenAI chose

52. *Ibid.*

53. OpenAI, "Introducing Whisper" (21 September 2022), online: <openai.com> [perma.cc/6DH9-EBCE] (detailing an open-source automated speech recognition tool developed by OpenAI); OpenAI, "DALL·E API Now Available in Public Beta" (3 November 2022), online: <openai.com> [perma.cc/EXM8-DBH2] (detailing a text to imagine machine learning system now available through a paid application programming interface).

54. Brown et al, *supra* note 50. See also Alec Radford et al, "Language Models are Unsupervised Multitask Learners" (2019), online (pdf): <cdn.openai.com> [perma.cc/F2ZL-ABVY].

55. OpenAI, "GPT-2: 1.5B Release" (5 November 2019), online: <openai.com> [perma.cc/J59L-NAPQ]. See also OpenAI, "GPT-2", online: <github.com> [perma.cc/P8TW-V4P4]; OpenAI, "GPT-2", online: <huggingface.co> [perma.cc/FL4B-89RW].

56. Hugging Face, "Models" (last visited 26 February 2024), online: <huggingface.co> [perma.cc/KEY9-3M5H] (listing GPT-2 as the third-most frequently downloaded model as of 26 February 2024, with more than 18.5 million downloads).

57. Greg Brockman, Ilya Sutskever & OpenAI, "OpenAI LP" (11 March 2019), online (pdf): <openai.com> [perma.cc/A4WU-QAHU].

58. Greg Brockman, "Microsoft invests in and partners with OpenAI to support us building beneficial AGI" (22 July 2019), online (blog): <openai.com> [perma.cc/AY3B-ZJK3].

59. Brown et al, *supra* note 50.

not to release on an open-access basis, citing concerns about security, potential misuse, and the need to pursue commercialization strategies to afford ongoing development costs.[60] Instead, OpenAI charges users to access GPT-3 through an application program interface (API),[61] initially restricted to invited users but now more broadly available to paid users in many, but not all, countries.[62]

GPT-3 generated a great deal of attention[63]—some would say "hype"[64]—when it was released. At the time, it was the largest language model ever trained, generating unexpectedly good text sequences in response to prompts. GPT-3, especially in subsequent updated versions that have been released more recently, excels at generating plausible responses to prompts without further fine-tuning on specific tasks. In particular, GPT-3 does a good job at "zero-shot" learning, which means it can provide plausible responses to a prompt without being provided examples of the types of responses sought. It also excels at "few-shot" learning, which enables it to adjust its outputs based on examples of desired responses to similar prompts.[65] GPT-3's zero-shot learning capabilities have attracted the most public attention: there are, for instance, examples of people using GPT-3 to generate newspaper op-eds,[66] write law journal

60. Greg Brockman et al, "OpenAI API" (11 June 2020), online (blog): <openai.com> [perma.cc/F84C-MHBZ]. See also OpenAI, "Pricing", online: <openai.com> [perma.cc/B9P7-ED4B] [OpenAI, "Pricing"].

61. *Ibid*.

62. See e.g. OpenAI, "OpenAI's API Now Available with No Waitlist" (18 November 2021), online (blog): <openai.com> [perma.cc/C282-FZDR]. See also OpenAI, "Supported countries and territories", online: <beta.openai.com> [perma.cc/43BB-ZW62] (noting that as of the time of writing, the OpenAI API is available in 156 countries, out of 195 countries in the world).

63. See e.g. Cade Metz, "Meet GPT-3. It Has Learned To Code (and Blog and Argue)", *The New York Times* (24 November 2020), online: <nytimes.com> [perma.cc/M236-YPS4]; Tom Simonite, "Did a Person Write This Headline, or a Machine?", *Wired* (22 July 2020), online: <wired.com> [perma.cc/JM9Z-AW7S]; Matthew Hutson, "Robo-writers: the rise and risks of language-generating AI" (2021) 591 Nature 22, DOI: <10.1038/d41586-021-00530-0>; Prasenjit Mitra, "A language generation program's ability to write articles, produce code and compose poetry has wowed scientists", *The Conversation* (23 September 2020), online: <theconversation.com> [perma.cc/LMG5-RUMB]; Will Douglas Heaven, "OpenAI's new language generator GPT-3 is shockingly good—and completely mindless", *MIT Technology Review* (20 July 2020), online: <technologyreview.com> [perma.cc/4WUX-7VBW].

64. See e.g. Rob Toews, "GPT-3 Is Amazing—And Overhyped", *Forbes* (19 July 2020), online: <forbes.com> [perma.cc/345E-DMVQ]; Sam Altman, "The GPT-3 hype is way too much" (19 July 2020), online: <twitter.com> [perma.cc/CP3B-88QU] (a Twitter post from OpenAI's CEO downplaying the capabilities of GPT-3).

65. Brown et al, *supra* note 50 at 6.

66. GPT-3, "A robot wrote this entire article. Are you scared yet, human?", *The Guardian* (8 September 2020), online: <theguardian.com> [perma.cc/YS83-CD6S].

articles,[67] successfully pass multiple-choice sections of several bar exams,[68] and prepare student essays (as well as grades and instructor feedback).[69] There is even a recent example of an application to assist a person with intellectual disabilities to prepare better business emails that was apparently built without any code in 15 minutes, using zero-shot prompt instruction.[70] GPT-3's few-shot performance, as well as its performance when fine-tuned on a relatively small number of examples, is also full of potential, in part because working with either process is simple. Recent examples of applications built using fine-tuned GPT-3 in this way include simple homemade mental health chatbots[71] and an AI assistant to help scholars automate literature reviews.[72]

GPT-3 has also generated a great deal of critique. Some have raised concerns about OpenAI's turn from open-access and non-profit research that benefits the broader community to a profit-seeking organization that no longer shares code and training data for its highest-impact projects.[73] Others have pointed out that due to biases in online materials used to train GPT-3, completions display many biases, including racial biases, religious biases, and

67. Benjamin Alarie & Arthur Cockfield, "Will Machines Replace Us? Machine-Authored Texts and the Future of Scholarship" (2021) 3:2 L Tech & Humans 5.

68. Michael J Bommarito II & Daniel Martin Katz, "GPT takes the Bar Exam" (3 January 2023) Working Paper, online: <papers.ssrn.com> [perma.cc/EPG5-Z5Z5].

69. Stephen Marche, "The College Essay Is Dead", *The Atlantic* (6 December 2022), online: <theatlantic.com> [perma.cc/FJT5-MD6X] (discussing ChatGPT, an offshoot of GPT-3).

70. See Danny Richman, "GPT-3 Business Email Generator" (1 December 2022), online: <seotraininglondon.org> [perma.cc/NVY5-FKZW]. See also Danny Richman, "I mentor a young lad with poor literacy skills who is starting a landscaping business" (1 December 2022), online: <twitter.com> [perma.cc/R3LF-XC2C].

71. Amogh Agastya, "Fine-tuning GPT-3 Using Python to Create a Virtual Mental Health Assistant Bot" (11 September 2022), online: <betterprogramming.pub> [perma.cc/TZ26-PL4V].

72. Elicit, "The AI Research Assistant", online: <elicit.org> [perma.cc/8AQJ-9VTV]. See also Ought, "How to use Elicit for topics that have lots of research" (8 March 2022), online (video): <youtube.com> [perma.cc/K3WZ-MHC8].

73. Karen Hao, "The messy, secretive reality behind OpenAI's bid to save the world", *MIT Technology Review* (17 February 2020), online: <technologyreview.com> [perma.cc/7SZN-KWRR].

gender-based biases.[74] Observers have documented that GPT-3 frequently hallucinates or makes up facts, often in plausible-sounding and difficult-to-detect ways[75]—and it can also be induced to produce disinformation intentionally.[76] Still others have raised concerns about large language models in general, suggesting that the practice of ingesting ever-expanding amounts of non-curated online text and running that text through neural networks with increasingly large numbers of parameters can lead to serious harms. These harms flow in part from the way that large language models produce plausible-sounding (but not necessarily truthful) text, in part from the way the models reflect back the biases embedded in their training data, and in part from the way that human beings react to seemly coherent text. As several people close to the development of some of these models put it, "the mix of human biases and seemingly coherent language heightens the potential for automation bias, deliberate misuse, and amplification of a hegemonic worldview".[77] There are also significant concerns about the environmental impact of large language models,[78] and about whether the use of text from the Internet complies with copyright law and with licences or terms of service—particularly where models are used for commercial purposes.[79]

Despite these real and ongoing concerns, this project used GPT-3 to build its machine-learning models. The main reason for doing so is because GPT-3 is very easy to use—and because it works well in bilingual French-English

74. Brown et al, *supra* note 50 at 36–39; Abubakar Abid, Maheen Farooqi & James Zou, "Persistent Anti-Muslim Bias in Large Language Models" (delivered at the 2021 AAAI/ACM Conference on AI, Ethics, and Society) 298, online: <arxiv.org> [perma.cc/3J7V-RMLK]; Li Lucy & David Bamman, "Gender and Representation Bias in GPT-3 Generated Stories" (delivered at the Third Workshop on Narrative Understanding) 48, DOI: <10.18653/v1/2021.nuse-1.5>; Amy B Cyphert, "A Human Being Wrote This Law Review Article: GPT-3 and the Practice of Law" (2021) 55:1 UC Davis L Rev 401 at 413.

75. Tianyu Liu et al, "A Token-level Reference-free Hallucination Detection Benchmark for Free-form Text Generation" (delivered at the 60th Annual Meeting of the Association for Computational Linguistics) 6723, online: <arxiv.org> [perma.cc/LV7L-6BT6].

76. Ben Buchanan et al, *Truth, Lies, and Automation: How Language Models Could Change Disinformation* (Washington, DC: Centre for Security and Emerging Technology, 2021), DOI: <10.51593/2021CA003>.

77. Emily Blender et al, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" (2021) (delivered at the 2021 ACM Conference on Fairness, Accountability, and Transparency) 610 at 618, DOI: <10.1145/3442188.3445922>.

78. *Ibid* at 612–13.

79. Cade Metz, "Lawsuit Takes Aim at the Way A.I. Is Built", *The New York Times* (23 November 2022), online: <nytimes.com> [perma.cc/M388-SSNA] (noting that Codex, an offshoot of GPT-3 that assists with coding, has been the subject of a class action lawsuit for breach of copyright).

settings, switching back and forth with ease.[80] Unlike many other state-of-the-art multi-lingual machine learning models, researchers with limited coding skills can easily engage with GPT-3, and they can do so without needing access to powerful cloud-based computer clusters.[81] In fact, it is possible to use GPT-3 through a browser without any coding skills at all: OpenAI has a web-browser interface where users can input prompts and get responses and can adjust various settings. Some legal scholars have reported that using this browser-based environment with carefully crafted prompts can produce usable legal writing.[82] However, where GPT-3 becomes particularly useful for legal researchers is that it only requires basic coding skills and small amounts of labelled data to train highly accurate, fine-tuned models and to apply those models to large legal datasets through the easy-to-use OpenAI API.[83]

To fine-tune GPT-3 on a particular task, all that one needs to do is prepare a file[84] with matching pairs of prompts and completions and send that file to the OpenAI API, selecting which base model[85] to use for training. Then, to use the fine-tuned model, one just sends the API a single command in Python with a new prompt, the name of the trained model to be used, and, if desired, values for a small number of settings. The API will send back the response, along with some additional data.[86]

80. Brown et al, *supra* note 50 at 14–16.

81. See Alexander Borzunov et al, "PETALS: Collaborative Inference and Fine-tuning of Large Models" (2 March 2023), DOI: <10.48550/arXiv.2209.01188>. Because large language models have billions of parameters, fine-tuning a model requires a huge number of calculations that typically require cloud computing systems with multiple high-cost accelerators. While many scholars have access to national cloud computing research clusters and while attempts are currently being made to democratize access to these models through innovative distributed computing systems, sophisticated technical knowledge is currently needed at work with these systems.

82. Cyphert, *supra* note 74 at 420.

83. OpenAI, "Introducing OpenAI", *supra* note 51.

84. The file is in JavaScript Object Notation (JSON) format, which is a human-readable data format that stores information in a series of pairs (or nested pairs) that is commonly used to transfer data online.

85. OpenAI has several different base models that are starting points for fine-tuning. A GPT base model is a complex algorithm that has been developed through training on vast amounts of data where the task that the algorithm is optimized for is predicting the next word in a sequence of text. In fine-tuning, the process starts with this algorithm and then adjusts values (or weights) within that algorithm to optimize for a new task with new data (i.e., to optimize for predicting a desired output for a given input). The main differences between base models offered by OpenAI relates to the size of the model (i.e., the number of parameters in the algorithm) and the amount (and recency) of data that the base model was trained on. Generally speaking, the larger the model the more accurate it is and the more expensive it is to use.

86. *Ibid*.

To give a simple example, one could fine-tune GPT-3 by sending the API a file with the following docket entry prompts and expected completions:

> {"prompt": "Order rendered by The Honourable Mr. Justice John Norris at Toronto on 27-AUG-2019 dismissing the stay of execution doc.3", "completion": "Norris"
>
> "prompt": "Ordonnance rendu(e) par Monsieur le juge Scott à Montréal le 28-MAI-2012 accordant la demande de sursis d'exécution Décision déposée le 28-MAI-2012", "completion": "Scott"
>
> "prompt": "Copy of Direction of the Court (Grammond, J.) dated 17-SEP-2019 'These proceedings are held in abeyance until a case management conference is held in these matters.'", "completion": "Grammond"
>
> "prompt": "Ottawa 28-JAN-2022 BEFORE The Honourable Madam Justice Roussel Language: E Before the Court: Motion Doc. No. 3 on behalf of Applicant Result of Hearing: Matter reserved held by way of Conference Call Duration per day: 28-JAN-2022 from 09:03 to 10:08 Courtroom : Ottawa (Zoom) Court Registrar: Beatriz Winter Total Duration: 1h 5min Appearances: Dotun Davies representing Applicant Rachel Beaupre representing Respondent Minutes of Hearing entered in Vol. 399 page(s) 25 - 27 Abstract of Hearing placed on file", "completion": "Roussel"
>
> "prompt": "Oral directions of the presiding judge dated 13-JUL-2020 directing 'The Minister's motion for a stay of release will be heard on Friday, July 17, 2020 at 10:00am Eastern Time.' received on 13-JUL-2020", "completion": "none"
>
> "prompt": "Ordonnance rendu(e) par Madame la juge Elizabeth Walker à Ottawa le 28-JAN-2019 rejetant la requête demandant le sursis interlocutoire de la Décision Décision déposée le 28-JAN-2019 Pris en considération par la Cour

avec comparution en personne inscrit(e) dans le livre J. & O., volume 811 page(s) 495 - 498 Copie de l'ordonnance envoyé(e) à toutes les parties Lettres placées au dossier.", "completion": "Walker"}

Then, once the system was fine-tuned, if the API was asked to apply the fine-tuned model to a new docket entry prompt, "Copy of Order dated 12-JAN-2017 rendered by Chief Justice Crampton concerning Prothonotary Milczynski is assigned as CMJ. placed on file. Original filed on Court File No. IMM-5388-15", the system would return the completion: "Crampton". In other words, the system would learn from a small number of examples that its task was to extract the family name of a judge in the prompt, if one is available. Unlike simple Regex searches, it would be able to do so regardless of the precise way the name of the judge was recorded, and it would not need a list of expected judge names because the system can infer this from context.

This relatively simple example of extracting the name of the judge from a docket entry was selected to help readers who are not familiar with machine learning get a sense for how fine-tuning works. However, it is important to understand that the fine-tuned models used in this project were necessary because of the nature of the data available. Published court decisions typically follow a fairly well-structured format, and it is generally possible to programmatically extract some key datapoints from decisions using simple code, including, for example, the name of the judge and sometimes even the outcome of the decision, which are often found in specific locations in published decisions and which may be recorded using standardized terms. By contrast, because (for reasons described above) this study had to go beyond published decisions and instead used metadata available in online court dockets, the task of extracting the desired datapoints was considerably more complex. That is because natural language entries in the court dockets are less structured than published court decisions and because dockets frequently involve multiple processes. For example, a court docket may include an application for a stay of removal, a motion for an extension for a deadline, an application for leave for judicial review, and a decision on the merits of a judicial review. Typically, a different judge would decide each of these matters. Information about these processes is not recorded in a standardized way, and often, to understand a given docket entry, they must be read against context available in several prior docket entries.

At any rate, for this project, a series of fine-tuned models similar to the model for extracting the name of a judge described above were built to categorize and extract information from docket entries. Thus, for example, to create a model to categorize the outcome described in court order docket entries, a file was created that had a sample of hundreds of docket entries that the Regex process described in the prior section identified as involving a court order. The author manually reviewed those docket entries and added a field for whether the order granted the requested stay, denied the requested stay, or the order was unclear. A file with pairs of the text of the docket entries (the prompt) and the text of the desired completion (e.g., "granted", "dismissed", and "other") was then sent to GPT-3 for fine-tuning. That produced a model in which all docket entries identified as involving court orders could be sent through the API, and the system sent back the appropriate completion. A similar process was used to create and apply models to extract other information that was needed (e.g., classifying whether a notice of a motion involved a stay application, etc.). The author then followed an iterative process: applying the model to new docket entries; verifying whether the model was working properly, and if not, providing additional examples of relevant dockets with the right completions; then fine-tuning again; and testing the new model again, until the model performed well on new dockets.

It should be noted that there are other machine learning models that would likely do a better—and more efficient—job at these tasks than GPT-3. These tasks could be considered fairly straightforward classification or summarization tasks, and there are other models, including comparatively small open-source models, that excel at such tasks.[87] However, one of the aims of this project was to demonstrate to other legal researchers who might have limited coding skills what can be achieved through simple API-based machine learning systems—especially as we can anticipate that further low code or no code interfaces using similar principles are likely to be developed that will make research of this kind even more accessible for legal scholars. Also, one of the advantages of the relative simplicity of the classification and summarization tasks needed for this research is that it was possible to use the smallest and least expensive version of GPT-3, which made the research inexpensive: training each model cost under $0.25 and applying all the models cost under $10.[88] Because some researchers with more advanced skills might

87. Some tasks are especially simple, such as extracting the name of a judge from a docket entry. Some tasks require more semantic and contextual understanding, such as summarizing the outcome of an order or determining whether a motion involves a stay of removal. There is also some complexity related to the multilingual nature of the dockets. It is likely that relatively small open-source models that can be loaded easily on a single computer could complete these tasks, though more fine-tuning data would likely be needed to achieve comparable levels of accuracy.

88. The project used the smallest GPT-3 model: "Ada", which at the time of writing costs $0.0004 / 1,000 tokens for fine-tuning and $0.0016 / 1,000 tokens for inference (there are typically one or two tokens per work). That is about 1% of the cost of using the most performant, and most expensive, "Davinci" model. For pricing details, see OpenAI, "Pricing", *supra* note 60.

understandably prefer to work with open-source models, the dataset and the training data used for this article have been made available online.[89]

D. Docket Level Logic Using Pandas and Final Dataset

Once the machine learning tools were applied to categorize and summarize the relevant docket entries, logic was applied to the dataset of dockets using Pandas (a commonly used Python package that assists with manipulating tabular data[90]) to produce summaries at the docket level. Specifically, logic was applied to the data about docket entries extracted through the machine learning tools in order to work out (a) whether a given docket involved a motion for a stay of removal, (b) the outcome if available, (c) the justice if available, (d) the dates of the motion and the outcome, (e) the city where the application for judicial review was first filed, and (f) the category of the judicial review identified by the Federal Court in the docket (for example, is the application for judicial review about an IRB refugee decision, a Pre-Removal Risk Assessment, etc.).

Taken together, this produced a dataset of 6,161 dockets involving motions for stays of removal from 2012 to 2022 (to December 1 in 2022), of which 4,717 resulted in a hearing that led to an order on the stay motion, for which the name of the justice issuing the order was available.

E. Data Verification

Data verification was then undertaken on the resulting database. To ensure that most stays of removal were included, the database was compared with the dataset for one year's worth of stay of removal decisions published on CanLII that another scholar is manually reviewing for another project.[91] The automated process for this project identified 96 out of 98, or 98.0%, of the manually identified decisions, so we can be confident that the automated process is

89. Rehaag, "Luck of the Draw III: Code & Data", *supra* note 41.

90. Pandas, online: <pandas.pydata.org> [perma.cc/9GT9-S9R8]. See also Wes McKinney, *Python for Data Analysis: Data Wrangling with Pandas, NumPy and Jupyter*, 3rd ed (Sebastopol, Cal: O'Reilly, 2022), online: <wesmckinney.com> [perma.cc/3DXE-J4DJ].

91. Thériault, *supra* note 4.

accurately identifying most stay of removal cases.[92] Interestingly, despite the Federal Court's formal policy of providing all stay orders to CanLII for publication, the automated process identified many cases that were properly coded based on the information in the dockets but where there was no corresponding published CanLII decision. It is not clear why some decisions are not being provided to CanLII.[93]

To verify that the remaining datapoints gathered on the included decisions were accurate, a research assistant manually verified 200 coded dockets.[94] According to this data verification process, the data was accurate in 198 of these 200 dockets. This 99% accuracy rate is excellent for research of this kind.[95]

92. The two cases that were not properly identified included one where there was no formal motion for a stay of removal (instead, the case proceeded based on an informal request). The second case involved a complex docket with several intertwined and overlapping steps that would have been difficult for a human research assistant to correctly code. Interestingly, the automated process was more accurate than the manual human process: the automated process correctly identified 4 cases that are published on CanLII that had been missed in the manual human process.

93. See Rehaag & Thériault, *supra* note 22 and accompanying text.

94. The datapoints verified are (a) through (d) as described in the prior section. Datapoints (e) and (f) were not manually verified, because they tabular structured data extracted directly from the dockets and thus perfectly reflect the information in the dockets.

95. One pattern that was not characterized as an error is that the automated process sometimes characterized cases where there may have been a stay of removal hearing and order as not involving a stay of removal where the docket did not include a formal motion for a stay of removal, but instead proceeded based on informal requests (sometimes in letters, emails, or faxes). These were not treated as errors because the system was intentionally trained to categorize cases in this way. That was done because, often, dockets with informal requests do not provide sufficient information about the type of motion requested. As such there was a worry that including these cases would add an element of arbitrariness to the findings (e.g., whether a case was included or excluded from the analysis would depend on choices by the registrar in how to describe the request, which might introduce biases). Because the system was coding these cases as expected, these cases were not characterized as errors. At any rate, based on the data verification that compares the automated process with human data gathering from CanLII, cases that proceed in this way are rare (i.e., under 1%).

F. Data Analysis

Finally, once the summary data on dockets was produced, statistical analysis was undertaken to discern patterns in stay applications and to produce the tables and charts for this article. This was done in Python using Pandas,[96] Matplotlib,[97] and Statsmodels.[98]

G. Limitations

Before moving on to examine the findings of the study, it is worth taking a moment to highlight some of the limitations of the methodology used. One limitation is that the accuracy of the underlying data was not verified—and in some cases, likely could not be verified. That is, no effort was made to go beyond the information in online court dockets. If this information was entered inaccurately into the dockets by the Federal Court Registrar, that inaccurate information would be reflected in the statistics used for this study. While the study did not attempt to estimate the frequency of any such inaccuracies, one might reasonably assume that errors (particularly if these are typographic errors in the dockets) are likely randomly distributed and, thus, should not impact the overall patterns. Another limitation of the study is that there are many factors that likely correlate with outcomes in stay decision-making that could not be extracted from the dockets alone. The facts of cases, the reasoning offered by courts, the reasoning offered by the underlying decision-maker, the quality of materials prepared, and the like, all undoubtedly affect outcomes in stay decision-making. This study, therefore, suffers from a common limitation in quantitative work of this kind: convenience bias—that is, researching independent variables that are relatively easy to study and ignoring other independent variables that might impact dependent variables. Because of this limitation, this study should be supplemented with other research that uses other methodologies to examine stay decision-making. Some suggestions for future research are set out in the concluding section of this article.

96. See Pandas, *supra* note 90.

97. Matplotlib, "Visualization with Python", online: <matplotlib.org> [perma.cc/Q47H-4UUK]. See also Jake VanderPlas, *Python Data Science Handbook* (Sebastopol, Cal: O'Reilly, 2017), online: <github.com> [perma.cc/U6T2-BXJW].

98. Statsmodels, "statsmodels 0.13.5: statistical models, hypothesis tests, and data exploration", online: <statsmodels.org> [perma.cc/6XXK-VSPH]. See also Skipper Seabold & Josef Perktold, "Statsmodels: Econometric and Statistical Modeling with Python" (delivered at the 9th Python in Science Conference) 92, online: <conference.scipy.org> [perma.cc/8KWV-QK8D].

III. Findings

According to the dataset produced through the methodology described in the prior section, from 2012 to 2022, there were 6,161 motions for stays of removal in immigration proceedings in Canada's Federal Court, of which 4,717 involved a hearing and an order made by the Court granting or dismissing the stay. The remaining motions were either discontinued by the applicant, or the Court declined to hear the application. The former frequently arises because the issue has become moot. That can occur for many different reasons. For example, applicants may discontinue motions because Canada Border Services Agency has agreed to voluntarily defer removal, because some other immigration process has been successful and they no longer face removal, because applicants have already been removed and now the issue is moot, or because applicants no longer oppose removal. The Court can also decline to hear the motion for the same reasons—as well as for several additional reasons. That includes instances where a justice views the application as an abuse of process, where the application was brought at the last minute without any explanation about why it was necessary to do so, where the applicant does not have "clean hands", or where the application is on its face clearly unfounded. Frequently, no reasons are provided when the Court declines to hear the application, but sometimes lengthy reasons are set out in the online court dockets.

The statistics in the remainder of the article involve the 4,717 motions for applications that resulted in a hearing and a court order granting or dismissing the motion.

A. Outcomes in Stay of Removal Motions, Overall and by Year

Table 1 and Charts 1 and 2 break down the 4,717 outcomes in motions for stay of removal applications from 2012 to 2022 that resulted in a hearing and a court order by the year of the order.

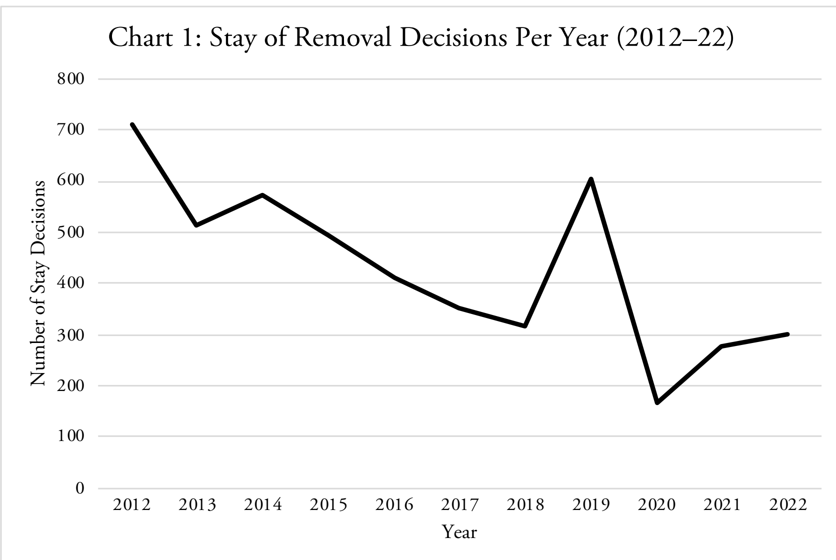
As can be seen in Table 1, the overall success rate in such motions is 33.8%. This rate fluctuates from year to year, from a low of 29.9% in 2012 to a high of 38.0% in 2015.

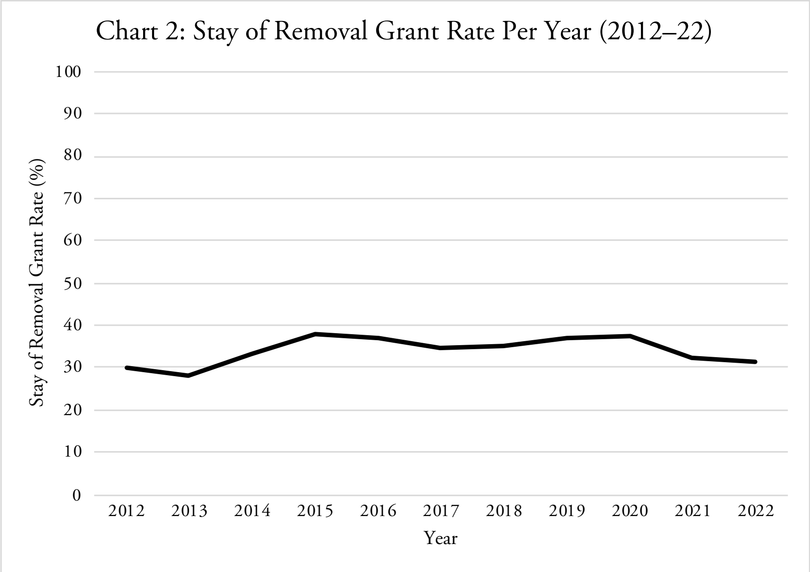
The number of stay applications made each year also varies significantly. There were 712 such decisions in 2012, compared with only 166 decisions in 2020 due to the global COVID-19 pandemic. Overall, however, the trend is towards a smaller number of cases per year.

Table 1: Federal Court Stays of Removal by Year Decided (2012–22)*

Year	Dismissed	Granted	Total	Grant Rate (%)
2012	499	213	712	29.9
2013	369	146	515	28.3
2014	383	191	574	33.3
2015	305	187	492	38.0
2016	257	152	409	37.2
2017	229	123	352	34.9
2018	205	112	317	35.3
2019	379	224	603	37.1
2020	104	62	116	37.3
2021	187	89	276	32.2
2022	207	94	301	31.2
Total	**3,124**	**1,593**	**4,717**	**33.8**

* (To December 1, 2022)





B. Outcomes in Stay of Removal Motions, by Judge Deciding the Motion

Table 2 and Chart 3 set out the study's main findings: stay of removal grant rates from 2012 to 2022 appear to vary dramatically depending on which justice decides the motion.

For example, applicants whose stay motions were heard by Justices Near (2.6%), Gascon (7.8%), and McVeigh (9.0%) were much less likely to succeed than applicants whose stay motions were heard by Justices Campbell (79.2%), O'Keefe (68.8%), and Ahmed (63.9%). All heard more than thirty motions.

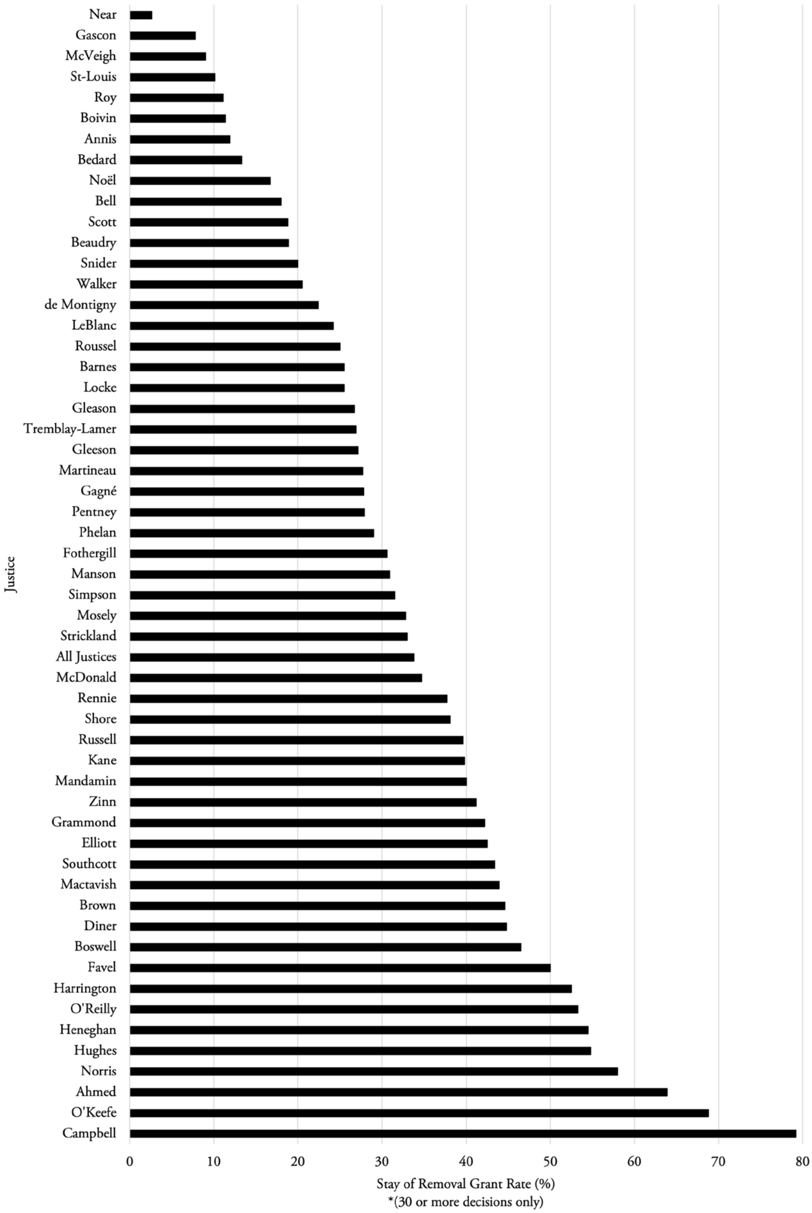
Table 2. Federal Court Stays of Removal by Justice (2012-22)*

Year	Dismissed	Granted	Total	Grant Rate (%)
Campbell	16	61	77	79.2
O'Keefe	15	33	48	68.8
Ahmed	22	39	61	63.9
Norris	21	29	50	58.0
Hughes	28	34	62	54.8
Heneghan	111	133	244	54.5
O'Reilly	35	40	75	53.3
Harringston	28	31	59	52.5
Favel	27	27	54	50.0
Boswell	54	47	101	46.5
Diner	58	47	105	44.8
Brown	62	50	112	44.6
Mactavish	74	58	132	43.9
Southcott	56	43	99	43.4
Elliott	77	57	134	42.5
Grammond	26	19	45	42.2
Zinn	77	54	131	41.2
Mandamin	21	14	35	40.0
Kane	65	43	108	39.8
Russell	99	65	164	39.6
Shore	73	45	118	38.1
Rennie	33	20	53	37.7
McDonald	64	34	98	34.7
Strickland	69	34	103	33.0
Mosely	78	38	116	32.8
Simpson	76	35	111	31.5
Manson	65	29	94	30.9
Fothergill	59	26	85	30.6
Phelan	49	20	69	29.0

Pentney	44	17	61	27.9
Gagné	52	20	72	27.8
Martineau	81	31	112	27.7
Gleeson	70	26	96	27.1
Tremblay-Lamer	38	14	52	26.9
Gleason	66	24	90	26.7
Locke	38	13	51	25.5
Barnes	76	26	102	25.5
Roussel	48	16	64	25.0
LeBlanc	47	15	62	24.2
de Montigny	97	28	125	22.4
Walker	35	9	44	20.5
Snider	28	7	35	20.0
Beaudry	43	10	53	18.9
Scott	26	6	32	18.8
Bell	50	11	61	18.0
Noël	65	13	78	16.7
Bedard	26	4	30	13.3
Annis	118	16	134	11.9
Boivin	70	9	**79**	**11.4**
Roy	96	12	108	11.1
St-Louis	71	8	79	10.1
McVeigh	71	7	78	9.0
Gascon	59	5	64	7.8
Near	37	1	38	2.6
All Justices	**3,124**	**1,593**	**4,717**	**33.8**

*(30 or more decisions only. To December 1, 2022.)

Chart 3: Stay of Removal Grant Rate by Justice*



To put the point starkly: from 2012 to 2022, applicants for stays of removal were approximately thirty times as likely to succeed before one Federal Court justice than another Federal Court justice. Moreover, the remarkable divergence in stay of removal grant rates is apparent not just in a handful of outlier justices. Rather, as can be seen in Chart 3, it appears that Federal Court justices are not clustered around an average grant rate but are instead distributed quite evenly across the full range of rates.

C. Outcomes in Stay of Removal Motions, by City Where Judicial Review Was Filed

Feedback provided on an earlier draft of this article suggested that it would be worth looking to see whether there was a difference in stay of removal grant rates depending on the city where the underlying application for judicial review was filed.[99]

99. The author is grateful to Simon Wallace for making this suggestion.

Table 3: Federal Court Stays of Removal by City Where Judicial Review Filed (2012–22)*

City	Dismissed	Granted	Total	Grant Rate (%)
Winnipeg	62	12	74	16.2
Montréal	641	149	790	18.9
Calgary	79	24	103	23.3
Edmonton	42	13	55	23.6
Vancouver	187	84	271	31.0
Ottawa	142	65	207	31.4
Toronto	1,948	1,239	3,187	38.9
All Cities	**3,124**	**1,593**	**4,717**	**33.8**

*(30 or more cases only. To December 1, 2022.)

As can be seen in Table 3, there are significant variations in stay of removal grant rates based on the city where the underlying judicial review was filed, ranging from 16.2% for Winnipeg to 38.9% for Toronto. The large majority of stay of removal applications were filed in Toronto (67.5% of stay applications) and Montreal (16.7% of stay applications). As such, the differences between outcomes in motions for stays of removal in cases where the underlying application for judicial review was filed in either city is particularly interesting: 38.9% for Toronto compared to 18.9% for Montreal.

D. Outcomes in Stay of Removal Motions, by Type of Underlying Application

Feedback on an earlier draft also recommended looking to see whether outcomes in motions for stays for removal varied depending on the type of underlying application for judicial review.[100] Table 4 breaks down outcomes in motions for stays of removal on this basis. As can be seen in the table, applications for stays of removal appear to be more likely to succeed where the underlying judicial review relates to a risk on return than when they relate to other immigration proceedings. For example, judicial reviews of the various categories of refugee claimants (the Federal Court categories for refugee-claim judicial reviews have shifted over time) and pre-removal risk assessments result in successful stay applications 46.0% of the time and 37.1% of the time, respectively. By contrast, cases categorized as "Other Arising in Canada", which would include judicial reviews of humanitarian and compassionate applications as well as judicial

100. Again, the author is grateful to Simon Wallace for the suggestion.

reviews of requests for administrative deferrals, were only successful 31.9% of the time. The "Other Arising in Canada" category does not disaggregate further between various types of applications, which is unfortunate both because it combines rather distinct types of applications and because this category reflects the majority of applications.

Table 4: Federal Court Stays of Removal by Type of Application Filed (2012–22)*

Case Type Assigned by Federal Court	Dismissed	Granted	Total	Grant Rate (%)
Immigration — Application for Leave and Judicial Review — Immigration Review Board — Immigration Division	27	9	36	25.0
Immigration — Application for Leave and Judicial Review — Other Arising in Canada	2,201	1,029	3,230	31.9
Immigration — Application for Leave and Judicial Review — Immigration Review Board — Refugee Protection Division**	30	16	46	34.8
Immigration — Application for Leave and Judicial Review — Immigration Review Board — Pre-Removal Risk Assessment	712	420	1,132	37.1
Immigration — Application for Leave and Judicial Review — Immigration Review Board — Refugee**	112	105	217	48.4
All Types	**3,124**	**1,593**	**4,717**	**33.8**

*(30 or more cases only. To December 1, 2022.)

**The categories assigned by the Federal Court are inconsistently named and have changed over time. The category "Refugee" was a catchall used between 2012 and 2018 for all judicial reviews from the Immigration Review Board involving refugee claimants. In 2019, it was replaced by the categories "Refugee Appeal Division", which is excluded from this table because it had fewer than 30 stay applications, and "Refugee Protection Division", which indicates that the claimant was not eligible to appeal to the Refugee Appeal Division.

E. Variance in Judge Stay Grant Rates Is Not Fully Attributable to City, Year, and Case Type

One question that arises when reviewing the prior tables and charts is whether the variance in stay grant rates across justices is reflective of patterns in case assignment. For example, suppose a justice mostly heard cases from a particular city where stay applications are on average weaker, mostly heard case types that are least likely to be successful, or only served on the Court during periods where stay grant rates were low. In such circumstances, if that justice had a low stay grant rate, that might be attributable largely to patterns in the types of cases they heard.

To see whether this might explain the variance observed in stay grant rates across justices, the data was filtered to try to limit the impact of these factors. This was achieved by looking at the stay grant rates of justices only in cases filed in the city with the largest number of cases filed (Toronto), only in cases involving the most common category of application ("Immigration — Application for Leave and Judicial Review — Immigration Review Board — Other Arising in Canada"), and only in a period of five years where stay grant rates were fairly consistent, that is, the rates did not vary by more than around 3% (2016–20).

Table 5: Federal Court Stays of Removal Decisions in Toronto, Case Type "Other Arising in Canada", by Justice (2012–22)*

Justice	Dismissed	Granted	Total	Grant Rate (%)
Heneghan	27	36	63	57.1
Southcott	17	17	34	50.0
Russell	18	15	33	45.5
Boswell	30	23	53	43.4
Elliott	34	26	60	43.3
Diner	25	17	42	40.5
McDonald	34	17	51	33.3
Brown	22	10	32	31.2
Simpson	29	12	41	29.3
Strickland	28	11	39	28.2
Gleeson	36	13	49	26.5
Annis	32	5	37	13.5
All Justices	**581**	**379**	**960**	**39.5**

*(30 or more decisions only. To December 1, 2022.)

As can be seen in Table 5, even when looking at this subset of cases from a single city, from a time period with relatively flat stay grant rates, and reflecting a single category of the underlying application, there is a large variation in stay grant rates across justices. For example, the stay grant rates for Justices Annis (13.5%), Gleeson (26.5%), and Strickland (28.2%) were much lower in such cases than for Justices Russell (45.5%), Southcott (50.0%), and Heneghan (57.1%). Moreover, if we drop the filter for the number of cases decided per justice lower, to twenty cases rather than thirty cases, the figures are even more extreme. For example, Justice Campbell decided twenty-four cases from this subset with a 100.0% grant rate, compared to Justice Gascon, who decided 21 cases from this subset with a 9.5% grant rate.

Statistical analysis, set out in Appendix B, was undertaken to explore whether the differences in grant rates across justices could be explained by patterns in case assignment related to the city where the case was filed, the year of decision, or the underlying case type. To this end, cases were filtered out where the deciding justice decided fewer than thirty cases, there were fewer than thirty cases filed in the city where the case was filed, or there were fewer than thirty cases of the same case type filed. This left a dataset of 4,468 cases. Chi-square tests and binary logistic regression were then applied to the dataset.

The statistical analysis confirms that even when holding other factors constant, the identity of the justice deciding the case remains a statistically significant predictor of stay outcomes. The statistical analysis also confirms that mean stay grant rates by city filed, by year of decision, and by case type are all statistically significant predictors of stay outcomes, and that the impact of the mean grant rate of each justice appears to be slightly larger than the impact of cities (at the 95% confidence level).

IV. Discussion & Conclusions

Three main sets of implications flow from the findings of this study. The first relates to variance in grant rates in motions for stays of removal across justices, and the second relates to variance in grant rates across cities where the application for a stay of removal was filed. The third set of implications involves more general lessons from this study about legal research involving computational methodologies, including the use of machine learning tools in legal research.

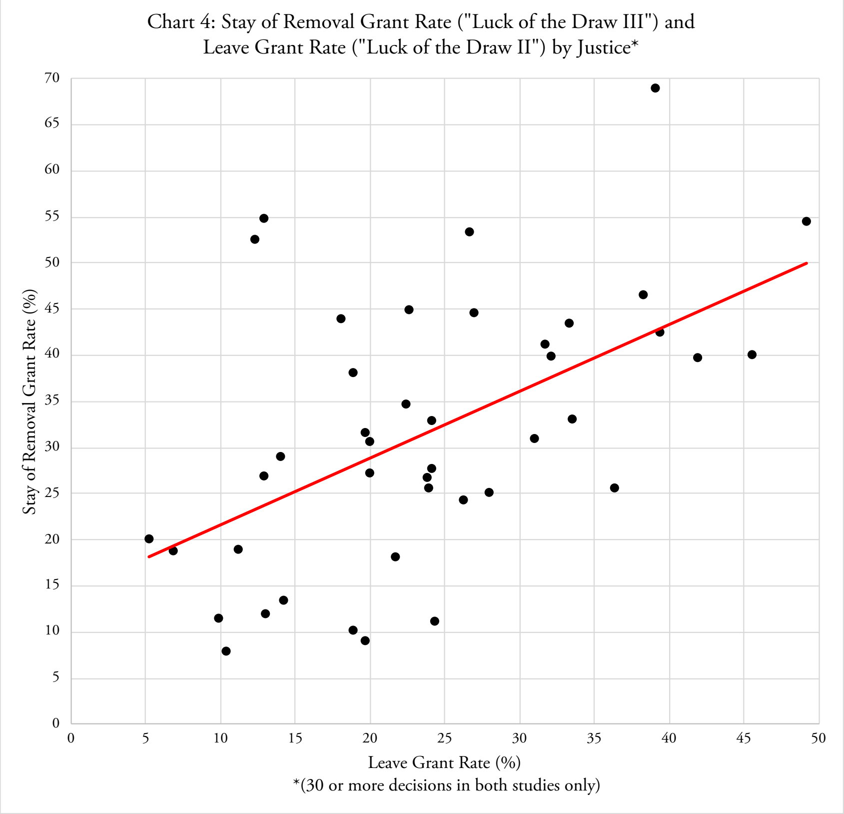
A. Variance in Stay Grant Rates Across Judges

The main finding of this study is that there appears to be a large unexplained variance in stay of removal grant rates depending on which justice decides the application.

Some of the observed variance appears to be correlated to other factors in ways that might be expected. For instance, it seems reasonable that variations in stay grant rates might fluctuate over time, both in responses to legal changes and changes in circumstances. As an extreme example, consider how stay of removal decision-making might have understandably shifted during the initial stages of the global COVID-19 pandemic. In light of the temporal differences in stay of removal grant rates found in this study, it makes sense that a justice deciding stays during only a portion of the period of the study may have different stay grant rates than a justice deciding stays during a different period (e.g., a justice was appointed or retired at the midpoint of the study). Similar points could be made about differences in grant rates depending on other factors examined in this study. If patterns in case assignment reflect these factors (e.g., if a given type of application is likely to be stronger or weaker,[101] and if a justice is mostly assigned that type of application), then variations in justice grant rates would be expected and would not raise concerns.

101. For example, the legal test applied to stay motions can vary depending on the underlying application (i.e., a judicial review of a deferral request vs other types of judicial reviews), which may make stay motions involving particular types of applications stronger or weaker. See *Wang*, *supra* note 25.

However, while this study found that these factors do partly account for variance in grant rates, they do not fully explain the variance. Significant variance in stay of removal rates across justices persists when one controls for these factors.



Moreover, some of the patterns observed in this study track directly onto findings from prior empirical research demonstrating similar correlations between who serves as the deciding justices and the likelihood that an applicant for judicial review of a refugee determination would be granted leave prompted responses from the Federal Court—including the second Luck of the Draw study.[102] Consider for example that, for justices who decided at least thirty cases in both the second Luck of the Draw study and the current study, where the justice was among the ten justices with the highest stay grant rates, eight had above

102. See generally Rehaag, "Luck of the Draw II", *supra* note 5.

average leave grant rates, and where the justice was among the ten justices with the lowest stay grant rates, all had below average leave grant rates. Chart 4 plots the leave grant rates and stay grant rates of all justices covered by both studies who decided at least thirty cases in each. As can be seen, there is a clear (but imperfect) correlation between the leave and stay grant rates identified in the two studies.

The Chief Justice of the Federal Court acknowledged that the prior research on leave grant rates raised "an issue that we perceive as being a troublesome one because it does have a fairness dimension to it".[103] The Federal Court also institutionally responded to this prior research by committing in a strategic planning document to make efforts to identify measures to address the variance in judicial outcomes.[104] Some research suggests that these efforts were successful in increasing consistency across justices.[105] The Court should similarly make efforts to decrease variance in stay of removal grant rates identified in this article—while, of course, preserving judicial independence. One way that the Federal Court could try to achieve greater convergence in stay of removal decision-making would be for justices to meet and discuss hypothetical case studies with the aim of generating increased consensus on appropriate stay outcomes.[106]

The central finding of the current study also points to the need for further research into whether there are particular aspects of stay of removal decision-making that might account for divergent stay grant rates across justices. The study did not examine reasons that justices offer for granting or denying stays. If the methodological challenges noted earlier in the article could be overcome (e.g., non-publication of stay decisions prior to 2018, prohibitions on bulk access to decisions published after 2018, volume that makes manual review challenging, many decisions that include few if any details, etc.), however, it would be instructive to undertake empirical examinations of those reasons. For example, it would be worth trying to understand the role of country of origin in stay outcomes—and considering whether patterns in case assignment in terms of country of origin might partly account for some of the variance in stay grant rates observed across justices. It would also be worth studying whether there are divergent interpretations of particular aspects of the test for stays of removal. If there are such differences, then appellate-level intervention to resolve the differences might assist in encouraging more

103. The Lawyers Weekly, "Federal Court C.J. Paul Crampton Addresses Immigration Bar's Concerns" (22 April 2012) at 00h:02m:29s, online (video): <youtube.com> [perma.cc/YL9F-864S].

104. Federal Court of Canada, *Strategic Plan* (2014-2019) at part I(A)(v)(a)(3), online (pdf): <fct-cf.gc.ca> [perma.cc/GF5D-F27L].

105. William Bryne et al, "Data-Driven Futures of International Refugee Law" (2023) J Refugee Studies at 14–15, DOI: <10.1093/jrs/feac069>.

106. See Rehaag, "Luck of the Draw II", *supra* note 5 at 3–5.

consistent decision-making. It would also be worth considering whether some form of legislative intervention may be appropriate. Given the high stakes, the volume, and the apparent divergence in approaches across justices in stay of removal decision-making, articulating legislative substantive and procedural norms might also be worth considering.

Continuing with the theme of the high stakes of this area of decision-making, in the absence of a reasonable explanation for the high levels of variance in stay grant rates across justices identified in this study—or of successful efforts to reduce that variance—the Court should reconsider a line of jurisprudence that places increased constitutional weight on stay of removal proceedings. This line of jurisprudence has held that immigration proceedings generally do not need to comply with procedural justice norms mandated by section 7 of the *Charter* on the theory that the right to life, liberty, and security of the person is not engaged in such proceedings.[107] The reasoning is that the proceedings do not necessarily result in removal because of the availability of subsequent procedures to block removal, and thus constitutional arguments about risks that a person might face on removal are premature until the person involved is actually facing removal. Ultimately, this reasoning can result in delaying scrutiny for *Charter* compliance until the last possible stage, that is, when the Federal Court hears stays of removal.[108] This line of jurisprudence is problematic for many reasons, including that the last-minute and time-pressured nature of stay of removal litigation does not lend itself to careful constitutional analysis.[109] However, even if, in theory, stays of removal could be effective sites for assessing *Charter* compliance, the variance in stay grant rates across justices identified in this study suggests that it would be risky to rely heavily on this form of decision-making to ensure that deportations comply with the *Charter*.

Finally, the finding that stay grant rates appear to vary across justices also raises interesting questions about consistency in legal decision-making more generally. There has been an explosion of empirical scholarship using

107. *Charter*, *supra* note 3, s 7.

108. Grey, *supra* note 3 at 136–38; Heckman, *supra* note 3 at 312–13; Kaushal, *supra* note 3 at 313.

109. One might draw links to jurisprudence about courts hesitating to determine complex constitutional matters in the absence of a well-developed record. See e.g. *Mackay v Manitoba*, 1989 CanLII 26 at 361 (SCC) (holding that *Charter* jurisprudence should not proceed in a "factual vacuum" because cases decided in the absence of a carefully prepared record tested through an adversarial process may result in "ill-considered opinions").

computational methods to detect patterns in legal decision-making,[110] and an increasingly sophisticated body of literature about what inferences (if any) can be drawn from the observed patterns in decision-making.[111] This article does not attempt to directly speak to that growing literature. The article's aim is instead more modest: showing how one new computational methodology is increasingly accessible for use by legal scholars who are interested in gathering data about legal decision-making, and then using those tools to examine an area of law that is difficult to study using conventional legal research methods. However, the hope is that scholars who are interested in exploring more sophisticated statistical analysis or in supplementing this research with methods that consider other factors will be able to draw on the data amassed in this study—and thus all the data and code used for this project is being made publicly available. Indeed, this is a pattern that we saw with the prior Luck of the Draw studies, where other scholars used the data initially gathered in those studies to undertake more sophisticated statistical analysis,[112] to pursue comparative research,[113] and to conduct new studies that helped further contextualize the findings of the research.[114]

110. See e.g. Michael Livermore & Daniel N Rockmore, *Law as Data: Computation, Text, and the Future of Legal Analysis* (Santa Fe, N Mex: SFI Press, 2019). For examples from the past year in the immigration and refugee law context alone, see e.g. Bryne et al, *supra* note 105; Daniel Ghezelbash & Keyvan Dorostkar, "Understanding the Politics of Refugee Law and Policy Making: Interdisciplinary and Empirical Approaches" (2023) J Refugee Studies, DOI: <10.1093/jrs/fead039>; Keyvan Dorostkar, Daniel Ghezelbash & Shannon Walsh, "A Data Driven Approach to Evaluating and Improving Judicial Decision-Making: Statistical Analysis of the Judicial Review of Refugee Cases in Australia" (2022) 45 UNSWLJ 1085; Hilary Evans Cameron, Avi Goldfarb & Leah Morris, "Artificial Intelligence for a Reduction of False Denials in Refugee Claims" (2022) 35:1 J Refugee Studies 493; Claire Barale, "Empowering Refugee Claimants and their Lawyers: Using Machine Learning to Examine Decision-Making in Refugee Law" (21 September 2023), online (pdf): <arxiv.org> [perma.cc/9SDC-KE3E].

111. See e.g. Daniel E Ho & Kevin M Quinn, "How Not to Lie with Judicial Votes: Misconceptions, Measurement, and Models" (2010) 98: 3 Cal L Rev 813; Daniel Kahneman, Olivier Sibony & Cass R Sunstein, *Noise: A Flaw in Human Judgment* (New York: Little, Brown Spark, 2021); Dane Thorley, "Randomness Pre-Considered: Recognizing and Accounting for 'De-Randomizing' Events When Utilizing Random Judicial Assignment" (2020) 17:2 J Empirical Leg Stud 342.

112. See e.g. Norris, *supra* note 46.

113. See e.g. Bryne et al, *supra* note 105.

114. See e.g. Jamie Liew et al, "Not Just the Luck of the Draw? Exploring Competency of Counsel and Other Qualitative Factors in Federal Court Refugee Leave Determinations (2005-2010)" (2021) 37:1 Refuge 61.

B. Variance in Stay Grant Rates Across Cities

A second key finding of this study is that there appears to be significant variance in stay of removal grant rates depending on the city where the underlying application for judicial review was first filed, even when controlling for other factors, such as the identity of the decision-maker. Of particular note is the large difference in stay grant rates between applications filed in Canada's largest two cities, with stay grant rates being substantially higher in Toronto than in Montreal.

It is not clear what might be causing this variance, especially because the Federal Court is a national court and, as such, Federal Court justices frequently travel to various cities (though language abilities may influence where the preponderance of cases decided by particular justices originate). One concerning possible factor is quality of counsel. Prior research has identified that quality of counsel appears to be an important driver of outcomes in immigration and refugee law proceedings.[115] Moreover, prior research has raised specific concerns about limitations on legal aid available for immigration and refugee law in Quebec and how that has made it difficult for many non-citizens to access counsel in that province.[116] So, it is possible that varying levels of access to quality counsel in different provinces might be the cause of different stay of removal grant rates. However, there could be other explanations as well. For example, suppose that Canada Border Services Agency staff in Montreal were more likely to grant compelling requests for stays of removal than similar staff in Toronto. This would impact the types of applications for judicial stays of removal in the two cities, and we would expect higher judicial stay of removal grant rates in Toronto than in Montreal. Or suppose that there are distinct groups of non-citizens who end up facing removal in Toronto and Montreal (for example, because of linguistic factors or because of patterns in migration). And suppose that the distinct groups of non-citizens raise different sorts of questions in terms of stay of removal applications. In such circumstances, differences in stay of removal grant rates across the two cities might be understandable.

115. Sean Rehaag, "The Role of Counsel in Canada's Refugee Determination System: An Empirical Assessment" (2011) 49:1 Osgoode Hall LJ 71; Craig Damian Smith, Sean Rehaag & Trevor Farrow, *Access to Justice for Refugees: How Legal Aid and Quality of Counsel Impact Fairness and Efficiency in Canada's Asylum System* (Toronto: Canada Excellence Research Chair in Migration and Integration, Centre for Refugee Studies, Canadian Forum on Civil Justice, 2021), DOI: <10.2139/ssrn.3980954>; Stephanie J Silverman & Petra Molnar, "Everyday Injustices: Barriers to Access to Justice for Immigration Detainees in Canada" (2016) 35:1 Refugee Survey Q 109; Jennifer Bond & David Wiseman, "Shortchanging Justice: The Arbitrary Relationship Between Refugee System Reform and Federal Legal Aid Funding" (2014) 91:3 Can Bar Rev 583; Liew et al, *supra* note 114.

116. Smith, Rehaag & Farrow, *supra* note 115 at 18, 24.

In this context, further research investigating the cause of the differences in stay of removal grant rates in different cities would be warranted. One particularly promising direction for research would be to examine random samples of materials filed in support of stay of removal applications in both Montreal and Toronto to attempt to discern whether quality of counsel might be driving stay of removal grant rates.[117] Legal aid programs across Canada that are interested in ensuring that the services they fund are of high quality, as well as law societies across Canada that are responsible for ensuring that their members meet their professional responsibilities, would also do well to look into whether the findings of this study should prompt further investigations.

C. New Computational Legal Research Tools and Bulk Access to Court Materials

A final set of implications from this study relate to legal research involving computational methodologies, specifically the use of machine learning tools to study high-volume forms of legal decision-making. This research project demonstrates that we are now at the point where legal scholars with modest coding skills can easily use machine learning and large language models hosted on cloud infrastructures to extract useful information from large quantities of legal text efficiently. This presents numerous research possibilities that would have been cost prohibitive and time intensive using human research assistants.

Consider the resources that would have been needed to conduct the present study using human research assistants. Based on the author's prior experience working with law student research assistants to review Federal Court dockets, it takes well-trained research assistants approximately one minute on average to manually review an online Federal Court docket and to extract the information used for this project.[118] It is difficult to get accuracy rates over 95% with human research assistants, who understandably get tired, misread dockets, make copy-paste errors, and other typographic errors. To achieve accuracy rates similar to those obtained by the automated process employed in this project using human research assistants would likely require double coding, that is, having two research assistants code each case, with differences resolved by the author. Based on the above estimates, double coding all 87,776 dockets that were used for this project would take almost 3,000 hours (equivalent to $75,000 at $25/hour), not including training time for research assistants or the time needed to resolve discrepancies. By contrast, the time invested into programming the automated process for this project (including fine-tuning and testing machine learning models) was less than forty hours once the online dockets were downloaded.

117. For an example of research using similar methods, see Liew et al, *supra* note 114.

118. Rehaag, "Luck of the Draw I", *supra* note 5; Rehaag, "Luck of the Draw II", *supra* note 5.

Additionally, using human research assistants produces a static dataset, with substantial costs for new data collection. On the other hand, automated processes can be adapted and run again with minimal effort. That means, for example, that re-applying the automated process after another year of Federal Court decision-making would be straightforward. It also means that if it is discovered that additional datapoints would be helpful in the data analysis stage, the process can be quickly modified. In fact, that occurred during this project. After the author shared initial draft results with other scholars, some suggested additional datapoints, including breaking down statistics on stays of removal based on the city where the judicial review was filed and the type of underlying application for judicial review. Extracting these datapoints programmatically took only a few moments, whereas manually gathering these new datapoints would have taken hundreds of hours.

Of course, some of the benefits of automated data gathering could have been achieved without using machine learning models. As was done in the first Luck of the Draw study, scraping online dockets and applying simple filters on tabular data or categorization using programmatic string searches for specific terms could have dramatically reduced the number of cases that human research assistants would need to review.[119] Alternatively, as was done in the second Luck of the Draw study, it would have been possible to create complex logics using multiple nested string searches that, through extensive trial and error measured against large human-coded datasets, could extract and parse the data needed for this project with reasonable accuracy.[120] While these alternative approaches could have been pursued, the main advantage of the approach taken in this iteration of the Luck of the Draw research was that, by leveraging simple-to-use but powerful machine learning models accessed through an API, the research project was completed quickly and with only small amounts of labelled data. This approach also allows other legal researchers with limited coding skills to replicate this research and to undertake similar types of research on other projects.

Given the time and financial savings, as well as the flexibility for iteratively engaging with data gathering and the simplified interfaces for accessing powerful machine learning processes, it seems likely that researchers will increasingly turn to these methodologies to study legal decision-making.

Currently, the main obstacle to conducting this type of research is access to bulk legal data. As previously mentioned in the methodology section of this article, legal publishers in Canada, including CanLII, prohibit bulk or programmatic access to most of their data. As a result, the main way researchers can currently obtain bulk legal data is by downloading it directly from court

119. For an example of such an approach, see Rehaag, "Luck of the Draw I", *supra* note 5.

120. For examples of projects that used such an approach, see Rehaag, "Luck of the Draw II", *supra* note 5; Norris, *supra* note 46.

and tribunal websites whose terms of service allow non-commercial reproduction. Currently, that can only be done by downloading data for individual cases, one by one.[121]

Conducting systematic web scraping of data on a case-by-case basis, especially for large volumes of cases that need to be kept up to date, requires more advanced coding skills than the simple process of submitting text to machine learning APIs described in this article. Downloading and maintaining an always up-to-date version of the data was the most complex part of this project. To build a system that would not only take a snapshot of Federal Court dockets but also ensure that the data was kept current required building a cloud-based system. This involved setting up cloud infrastructure and security, designing a database, implementing automated error handling, and tackling many other technical challenges typically faced by companies working on production rather than university-based scholars conducting computational legal research. Confronting these challenges was viable for this project because of funding from the Law Foundation of Ontario for another project using the same data, which made it possible to hire staff with relevant expertise.

While privacy concerns are sometimes cited as justification for why data of this sort should not be easily available in bulk, the current situation is not privacy-protective.[122] As this article illustrates, bulk legal data can be obtained if one has sufficient time, skills, and resources. Furthermore, large quantities of bulk legal data are already in the hands of corporations, including commercial legal publishers, as well as repeat litigators such as Department of Justice lawyers. In other words, if there are privacy reasons why this data should be protected, the current approach of posting data online in inconvenient formats for bulk access does not safeguard that privacy. Instead, it creates an unfair asymmetry in access to bulk legal data, where commercial actors, parties with deep pockets who can afford to purchase services from those commercial actors, and powerful repeat litigators like the Department of Justice, can build and use legal research tools that leverage machine learning on large legal datasets, while most scholars, journalists, community activists, and lawyers representing low-income and marginalized groups cannot easily do so.[123]

For these reasons, the current project is making all the data collected available for use by other non-commercial open-access researchers. However, a better solution would be for all legal data that is already being made available online by courts and tribunals to be shared through APIs in easily accessible bulk formats, possibly in combination with efforts to protect privacy by redacting the names and other identifying information in all publicly available legal data (i.e., both in new bulk formats and existing online formats). There

121. See Khan & Rehaag, *supra* note 37 and accompanying text.

122. *Ibid*.

123. *Ibid*.

are indications that some courts are interested in pursuing bulk access. The Federal Court, for example, has identified making legal data more accessible as a key priority in its most recent Strategic Plan.[124] Another example is that CanLII recently issued a call for participants to help build an API for bulk access to British Columbia Supreme Court and Court of Appeal decisions.[125]

These efforts to provide increased access to bulk legal data should continue, as should discussions about how to better protect privacy in a world where court documents are increasingly available online. It is also important to have ongoing conversations about how to encourage the development of legal technology that enhances social justice, protects rights, and increases transparency in legal decision-making. Unfortunately, new technologies all too frequently end up reinforcing the power of already powerful actors, often at the expense of marginalized groups. There have been particularly compelling critiques about discrimination and human rights violations resulting from machine learning and automated decision-making used against the groups affected by the decision-making studied in this article,[126] who are disproportionately racialized[127] and who face other intersecting forms of vulnerability.[128] These critiques include concerns about the exacerbation of historical racism through new technologies based on biased datasets,[129] worries

124. Federal Court of Canada, *Strategic Plan 2020-2025* (2021), online (pdf): <fct-cf.gc.ca> [perma.cc/WSD6-EMJ6].

125. CanLII, "Call for Participants" (8 July 2022), online: <blog.canlii.org> [perma.cc/4ZQS-WHZN].

126. For examples of these critiques, see e.g. Molnar & Gill, *supra* note 37; Petra Molnar, "Surveillance Sovereignty: migration management technologies and the politics of privatization" in Idil Atak and Graham Hudson, eds, *Migration, Security, and Resistance: Global and Local Perspectives* (Abingdon, UK: Routledge, 2022) 66.

127. Sharryn J Aiken, "From Slavery to Expulsion: Racism, Canadian Immigration Law and the Unfulfilled Promise of Modern Constitutionalism" in Vijay Agnew, ed, *Interrogating Race and Racism* (Toronto: University of Toronto Press, 2007) 55 at 63; Robyn Maynard, "Black Life and Death Across the U.S.–Canada Border: Border Violence, Black Fugitive Belonging, and a Turtle Island View of Black Liberation" (2019) 5:1–2 Critical Ethnic Studies 124.

128. Tanya Aberman, "Forced-Voluntary Return: An Intersectional Approach to Exploring 'Voluntary' Return in Toronto, Canada" (2022) 5:1 Migration & Society 13; Ameil J Joseph, *Deportation and the Confluence of Violence Within Forensic Mental Health and Immigration Systems* (New York: Palgrave Macmillan, 2015).

129. See *supra* note 74.

about technocolonialism and data colonialism,[130] and alarm at the rush by states to deploy new technologies to control the movement of people across borders in ways that are insensitive to the serious harms this can produce.[131]

Those of us interested in deploying legal technologies in rights-enhancing ways must pay careful attention to these critiques, and we should participate in efforts to push back against rights-limiting uses of new legal technologies. At the same time, however, as this paper has demonstrated, there are opportunities for legal scholars to use some of these same technologies to expose problems in human decision-making—and hopefully to improve that decision-making.[132] Courts, tribunals, and other legal institutions should welcome this research and, to this end, should help facilitate fair and equal access to bulk legal data.

130. Mirca Madianou, "Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises" (2019) 5:3 Soc Media & Society; Koen Leurs, "On Data and Care in Migration Contexts" in Marie Sandberg et al, eds, *Research Methodologies and Ethical Challenges in Digital Migration Studies Caring For (Big) Data?* (Cham, Switzerland: Palgrave Macmillan, 2022) 221 at 222–24, DOI: <10.1007/978-3-030-81226-3>.

131. Petra Molnar, *Technological Testing Grounds: Migration Management Experiments and Reflections from the Ground Up* (Brussels: European Digital Rights, 2020), online (pdf): <edri.org> [perma.cc/3V7U-9F7W].

132. For other examples of attempting to use AI to enhance rights, fairness, and transparency in legal decision-making involving the marginalized, see e.g. Kristen Bell et al, "The Recon Approach: A New Direction for Machine Learning in Criminal Law" (2021) 36 BTLJ 821; Hilary Evans Cameron, Avi Goldfarb & Leah Morris, "Artificial Intelligence for a Reduction of False Denials in Refugee Claims" (2022) 35:1 J Refugee Studies 493; Cameron, Goldfarb & Morris, *supra* note 110.

Appendix A: Example of Online Federal Court Docket

Recorded Entry Information : IMM-5039-21 ×

Type : Immigration Matters

Type of Action : Immigration Matters

Nature of Proceeding : Imm - Appl. for leave & jud. review - Pre-removal risk assessment

Filing Date : 2021-07-28

Office : Calgary **Language :** English

Recorded Entry Summary Information

Doc	Date Filed	Office	Recorded Entry Summary	Download
null	2021-08-27	Calgary	Memorandum to file from Calgary Office dated 27-AUG-2021 Re: Notice of Discontinuance and proof of service sent to SRO, Ottawa Immigration section placed on file.	
18	2021-08-27	Calgary	Affidavit of service of Jyoti Lal sworn on 27-AUG-2021 on behalf of Applicant confirming service of Notice of Discontinuance upon Respondent by e-mail on 27-AUG-2021 with Exhibits A filed on 27-AUG-2021	
17	2021-08-27	Calgary	Notice of discontinuance on behalf of the applicant filed on 27-AUG-2021	
null	2021-08-03	Ottawa	Acknowledgment of Receipt received from both parties with respect to Order of the Court (Favel, J.) dated 03-AUG-2021 placed on file on 03-AUG-2021	
16	2021-08-03	Ottawa	Order dated 03-AUG-2021 rendered by The Honourable Mr. Justice Favel Matter considered with personal appearance The Court's decision is with regard to Motion Doc. No. 2 Result: granted Filed on 03-AUG-2021 copies sent to parties entered in J. & O. Book, volume 909 page(s) 67 - 69 Interlocutory Decision	
null	2021-08-03	Ottawa	Ottawa 03-AUG-2021 BEFORE The Honourable Mr. Justice Favel Language: E Before the Court: Motion Doc. No. 2 on behalf of Applicant Result of Hearing: Matter reserved held by way of Conference Call in chambers Duration per day: 03-AUG-2021 from 04:00 to 04:56 Courtroom : Judge's Chambers - Ottawa Court Registrar: Jean-Loup Lhérisson Pimentel Total Duration: 56mins Appearances: Ms. Lucinda Wong 403-260-1579 representing Applicant Ms. Galina Bining 780-495-8337 representing Respondent Comments: DARS hand held recorder used Minutes of Hearing entered in Vol. 394 page(s) 59 - 61 Abstract of Hearing placed on file	
15	2021-07-30	Ottawa	Solicitor's certificate of service on behalf of Galina Bining confirming service of Respondent's motion (Docs. 12-14) upon Applicant by email on 30-JUL-2021 filed on 30-JUL-2021	
14	2021-07-30	Ottawa	Motion Record in response to Motion Doc. No. 2 containing the following original document(s): 12 13 Number of copies received: 1 on behalf of Respondent filed on 30-JUL-2021	
13	2021-07-30	Ottawa	Memorandum of fact and law contained within a Motion Record on behalf of Respondent filed on 30-JUL-2021	
12	2021-07-30	Ottawa	Affidavit of Gulnaz Mirza sworn on 30-JUL-2021 contained within a Motion Record on behalf of Respondent in opposition to Motion Doc. No. 2 with Exhibits A-B filed on 30-JUL-2021	
11	2021-07-30	Calgary	Affidavit of Rosanna Delgado on behalf of the applicant sworn on 30-JUL-2021 confirming service of Applicant's Book of Authorities on the respondent by email on 30-JUL-2021 with attached exhibit(s) A filed on 30-JUL-2021	
null	2021-07-30	Calgary	Book of Authorities consisting of 1 volume(s) on behalf of the applicant received on 30-JUL-2021	
10	2021-07-29	Edmonton	Solicitor's certificate of service on behalf of Camille N. Audain confirming service of Doc. No. 9 upon Applicant by telecopier on 29-JUL-2021 filed on 29-JUL-2021	

9	2021-07-29	Edmonton	Notice of appearance on behalf of the respondent filed on 29-JUL-2021	
8	2021-07-29	Calgary	Affidavit of Rosanna Delgado on behalf of the applicant sworn on 29-JUL-2021 confirming service of Consent To Electronic Service on the respondent by email on 29-JUL-2021 with attached exhibit(s) A filed on 29-JUL-2021	
7	2021-07-29	Calgary	Consent to electronic service of all documents not required to be served personally on behalf of the applicant filed on 29-JUL-2021	
null	2021-07-29	Ottawa	Acknowledgment of Receipt received from the parties with respect to Oral direction of the court placed on file on 29-JUL-2021	
null	2021-07-29	Ottawa	Oral directions of the presiding judge dated 29-JUL-2021 directing "The Applicants' motion will be heard on Tuesday, August 3, 2021 at 4:00pm EDT (2:00pm MDT) by teleconference for a duration of one (1) hour. The Respondent's responding motion record will be filed with the Court by 4:00pm EDT (2:00pm MDT) on Monday, August 2, 2021." received on 29-JUL-2021	
null	2021-07-29	Ottawa	Communication to the Court from the Registry dated 29-JUL-2021 re: Applicants' motion requesting a stay of the removal scheduled to take place on August 5, 2021 was sent to the Office of the Judicial Administrator.	
null	2021-07-28	Calgary	Letter from Counsel for the Applicants dated 28-JUL-2021 seeking an urgent hearing for a stay motion for applicants removal scheduled for 05-AUG-2021 at 11:55 AM received on 28-JUL-2021	
6	2021-07-28	Calgary	Affidavit of Rosanna Delgado on behalf of the applicant sworn on 28-JUL-2021 confirming service of Applicants' motion Record on the respondent by email on 28-JUL-2021 with attached exhibit(s) A filed on 28-JUL-2021	
5	2021-07-28	Calgary	Motion Record containing the following original document(s): 2 3 4 Number of copies received: 1 on behalf of Applicant filed on 28-JUL-2021	
4	2021-07-28	Calgary	Written representations on behalf of the applicant in support of Motion, Doc 2 filed on 28-JUL-2021	
3	2021-07-28	Calgary	Affidavit of Daya Singh contained within a Motion Record on behalf of the applicant sworn on 28-JUL-2021 in support of Motion, Doc 2 with attached exhibit(s) A-C filed on 28-JUL-2021	
2	2021-07-28	Calgary	Notice of Motion contained within a Motion Record on behalf of Applicant returnable (but no hearing date indicated at this time) with leave for short notice requested for an Order to stay the imminent removal of the Applicants (set for 05-AUG-2021 at 11:55 AM) filed on 28-JUL-2021	
null	2021-07-28	Calgary	Acknowledgment of Receipt received from Respondent with respect to Doc. 1 placed on file on 28-JUL-2021	
null	2021-07-28	Calgary	Pursuant to the Practice Direction and Order dated January 18, 2021, service of the ALJR (Doc 1) was effected by the Registry via email placed on file on 28-JUL-2021	
1	2021-07-28	Calgary	Application for leave and judicial review against a decision Canadian Border Services Agency in Calgary, Canada. Decision dated 12-JUL-2021 and communicated on 25-JUL-2021. File #s: 11-1070-1518, 11-1070-1514 and 11-1068-8318. filed on 28-JUL-2021 Written reasons received by the Applicant Tariff fee of $50.00 received	

Appendix B: Statistical Analysis

To simplify the statistical analysis undertaken for this article, cases were first filtered out where they involved a judge who decided fewer than thirty cases, a case type with fewer than thirty cases, a year with fewer than thirty cases, or a city of application with fewer than thirty cases. This left 4,468 cases (out of 4,717 in the full dataset).

For all statistical analysis, the outcomes in applications were coded as 0 (stay denied) or 1 (stay granted) and the stay outcome date was the year when the stay was decided.

Chi-square tests were conducted to determine whether the identity of the judge, the type of application, the year of application, and the city of application are statistically significant predictors of outcomes.

As can be seen in the table below, the judge, the city where the application was filed, the case type, and the stay outcome date are all statistically significant, with the stay judge and the city where the application was filed showing particularly strong associations given their high Chi-square values and low p-values.

Model 1: Chi-Squared Test			
Variable	**Chi2**	**Degrees of Freedom**	**p**
stay_judge	468.776	53	0.000
city	122.087	6	0.000
case_type	29.219	4	0.000
stay_outcome_date	25.823	10	0.004

Binary logistic regression was also used to explore whether patterns in case assigned might explain the apparent variations in stay of removal grant rates observed across judges in the data. Specifically, several models were used to examine whether either the identity of the judge or the average stay of removal grant rate of the judge deciding the case remained statistically significant when controlling for other factors considered in this article (type of application, year of application, and city of application).

The first binary logistic regression model uses the average (mean) stay of removal grant rates expressed as a percentage for each of the variables (judge, case type, outcome date, and city where the application was filed) as predictors and whether the application was granted as the outcome. For example, if the judge deciding a given case had a 25% average grant rate, then the value of the judge_mean_in_percent variable would be 25.

As can be seen in the table below, the average stay of removal grant rates for judges, case types, outcome dates, and city where the application was filed are all statistically significant predictors of outcomes. Another way of thinking about this is that the table confirms that the average stay of removal grant rates for judges remains statistically significant even when one controls for the average stay of removal grant rates for types of cases, outcome dates, and the city where applications were filed. The confidence intervals for the odds ratios also confirm that the average stay of removal grant rates for judges correlates more strongly with outcomes than the average stay of removal grant rates for cities of application (at the 95% confidence level).

Model 2: Binary Logistic Regression (means)

	Coefficient	Std Err	Z	P Value	Odds Ratio	95% CI Lower	95% CI Upper
Intercept	-5.984	0.466	-12.848	0.000	0.003	0.001	0.006
judge_mean_in_percent	0.045	0.002	18.496	0.000	1.046	1.041	1.051
type_mean_in_percent	0.043	0.008	5.247	0.000	1.044	1.027	1.061
year_mean_in_percent	0.042	0.010	4.098	0.000	1.042	1.022	1.063
city_mean_in_percent	0.025	0.005	5.138	0.000	1.025	1.015	1.035

To see whether the average stay of removal grant rate of judges remains statistically significant when we control for the other variables considered in this article in a more robust way, another binary logistic regression model was run. That model continued to use the average stay of removal grant rates expressed as a percentage for judges, but it used dummy variables for each of the other variables. That means, for example, that a case decided in 2014 would have a value of 1 for stay_outcome_date: 2014, and a value of 0 for all the other stay_outcome_dates.

As can be seen in the table below, the average stay of removal grant rate of the judge deciding the case (expressed as a percentage) remains statistically significant when controlling for these other factors.

Model 3: Binary Logistic Regression (categorical + judge mean)							
Variable	**Coefficient**	**Std Err**	**z**	**P Value**	**Odds Ratio**	**95% CI Lower**	**95% CI Upper**
Intercept	-3.524	0.502	-7.026	0.000	0.029	0.011	0.079
case_type: Imm - Appl. for leave & jud. review - IRB - Refugee	0.942	0.438	2.152	0.031	2.564	1.088	6.045
case_type: Imm - Appl. for leave & jud. review - IRB - Refugee Protection Div.	0.207	0.538	0.385	0.700	1.230	0.429	3.529
case_type: Imm - Appl. for leave & jud. review - Other Arising in Canada	0.297	0.412	0.719	0.472	1.345	0.600	3.018
case_type: Imm - Appl. for leave & jud. review - Pre-removal risk assessment	0.559	0.416	1.344	0.179	1.749	0.774	3.953
stay_outcome_date: 2013	0.138	0.139	0.996	0.319	1.148	0.875	1.507
stay_outcome_date: 2014	0.342	0.133	2.579	0.010	1.408	1.086	1.827
stay_outcome_date: 2015	0.416	0.137	3.038	0.002	1.516	1.159	1.983
stay_outcome_date: 2016	0.479	0.142	3.378	0.001	1.614	1.222	2.130
stay_outcome_date: 2017	0.338	0.152	2.230	0.026	1.402	1.042	1.887
stay_outcome_date: 2018	0.382	0.156	2.443	0.015	1.465	1.078	1.990
stay_outcome_date: 2019	0.342	0.130	2.637	0.008	1.408	1.092	1.816
stay_outcome_date: 2020	0.519	0.196	2.649	0.008	1.681	1.145	2.468
stay_outcome_date: 2021	0.257	0.172	1.496	0.135	1.294	0.923	1.813
stay_outcome_date: 2022	-0.056	0.180	-0.311	0.756	0.946	0.664	1.346
city: Edmonton	-0.142	0.439	-0.324	0.746	0.867	0.367	2.052
city: Montréal	0.239	0.273	0.874	0.382	1.270	0.743	2.168
city: Ottawa	0.545	0.301	1.813	0.070	1.725	0.957	3.111
city: Toronto	0.642	0.257	2.499	0.012	1.901	1.149	3.146
city: Vancouver	0.361	0.294	1.227	0.220	1.435	0.806	2.553
city: Winnipeg	-0.307	0.417	-0.738	0.461	0.735	0.325	1.664
judge_mean_0_to_1	4.695	0.253	18.592	0.000	109.423	66.703	179.503

For readers who are interested in whether specific factors might be statistically significant, the table below sets out a full binary logistic regression model using dummy variables for all categorical variables. In addition, for readers who would like to run their own statistical tests—including perhaps multilevel models or additional interaction variables—the data used for this article is being made publicly available.[133]

133. Rehaag, "Luck of the Draw III: Code & Data", *supra* note 41.

Model 4: Binary Logistic Regression (categorical)					
Variable	**Coefficient**	**Std Err**	**Z**	**P Value**	**Odds Ratio**
Intercept	-0.828	0.563	-1.470	0.141	0.437
stay_judge: Annis	-2.690	0.390	-6.892	0.000	0.068
stay_judge: Barnes	-1.608	0.367	-4.385	0.000	0.200
stay_judge: Beaudry	-1.851	0.463	-3.998	0.000	0.157
stay_judge: Bedard	-1.950	0.622	-3.134	0.002	0.142
stay_judge: Bell	-2.005	0.438	-4.575	0.000	0.135
stay_judge: Boivin	-2.232	0.467	-4.781	0.000	0.107
stay_judge: Boswell	-1.034	0.347	-2.976	0.003	0.356
stay_judge: Brown	-0.936	0.340	-2.756	0.006	0.392
stay_judge: Campbell	0.724	0.404	1.793	0.073	2.063
stay_judge: Diner	-0.964	0.344	-2.806	0.005	0.381
stay_judge: Elliott	-1.013	0.329	-3.081	0.002	0.363
stay_judge: Favel	-0.608	0.402	-1.515	0.130	0.544
stay_judge: Fothergill	-1.518	0.369	-4.110	0.000	0.219
stay_judge: Gagné	-1.300	0.395	-3.286	0.001	0.273
stay_judge: Gascon	-2.987	0.550	-5.428	0.000	0.050
stay_judge: Gleason	-1.459	0.381	-3.835	0.000	0.232
stay_judge: Gleeson	-1.740	0.364	-4.778	0.000	0.176
stay_judge: Grammond	-0.732	0.420	-1.742	0.081	0.481
stay_judge: Harrington	-0.193	0.398	-0.484	0.628	0.825
stay_judge: Heneghan	-0.435	0.310	-1.400	0.161	0.647
stay_judge: Hughes	-0.375	0.390	-0.962	0.336	0.687
stay_judge: Kane	-1.152	0.348	-3.311	0.001	0.316
stay_judge: LeBlanc	-1.745	0.417	-4.180	0.000	0.175
stay_judge: Locke	-1.758	0.438	-4.016	0.000	0.172
stay_judge: Mactavish	-0.873	0.340	-2.568	0.010	0.418
stay_judge: Mandamin	-0.877	0.458	-1.915	0.055	0.416
stay_judge: Manson	-1.461	0.362	-4.040	0.000	0.232
stay_judge: Martineau	-1.339	0.365	-3.666	0.000	0.262
stay_judge: McDonald	-1.392	0.353	-3.940	0.000	0.249
stay_judge: McVeigh	-3.059	0.489	-6.261	0.000	0.047
stay_judge: Mosley	-1.114	0.353	-3.157	0.002	0.328
stay_judge: Near	-3.940	1.059	-3.722	0.000	0.019
stay_judge: Norris	-0.302	0.405	-0.746	0.456	0.739
stay_judge: Noël	-1.759	0.433	-4.067	0.000	0.172
stay_judge: O'Keefe	0.524	0.444	1.179	0.239	1.688
stay_judge: O'Reilly	-0.373	0.372	-1.001	0.317	0.689
stay_judge: Pentney	-1.323	0.404	-3.271	0.001	0.266
stay_judge: Phelan	-1.630	0.395	-4.126	0.000	0.196
stay_judge: Rennie	-0.966	0.416	-2.323	0.020	0.381
stay_judge: Roussel	-1.677	0.415	-4.042	0.000	0.187
stay_judge: Roy	-2.500	0.421	-5.942	0.000	0.082
stay_judge: Russell	-1.032	0.329	-3.134	0.002	0.356
stay_judge: Scott	-1.623	0.551	-2.945	0.003	0.197
stay_judge: Shore	-0.762	0.350	-2.175	0.030	0.467
stay_judge: Simpson	-1.484	0.351	-4.233	0.000	0.227
stay_judge: Snider	-1.740	0.520	-3.348	0.001	0.175
stay_judge: Southcott	-0.924	0.349	-2.649	0.008	0.397
stay_judge: St-Louis	-2.703	0.474	-5.699	0.000	0.067
stay_judge: Strickland	-1.423	0.354	-4.022	0.000	0.241
stay_judge: Tremblay-Lamer	-1.252	0.437	-2.866	0.004	0.286
stay_judge: Walker	-2.011	0.469	-4.284	0.000	0.134
stay_judge: Zinn	-0.903	0.336	-2.685	0.007	0.405
stay_judge: de Montigny	-1.433	0.369	-3.883	0.000	0.239
case_type: Imm - Appl. for leave & jud. review - IRB - Refugee	1.025	0.439	2.335	0.020	2.787
case_type: Imm - Appl. for leave & jud. review - IRB - Refugee Protection Div.	0.279	0.539	0.517	0.605	1.322
case_type: Imm - Appl. for leave & jud. review - Other Arising in Canada	0.363	0.413	0.878	0.380	1.437
case_type: Imm - Appl. for leave & jud. review - Pre-removal risk assessment	0.631	0.417	1.514	0.130	1.879
stay_outcome_date: 2013	0.159	0.142	1.124	0.261	1.173
stay_outcome_date: 2014	0.401	0.138	2.906	0.004	1.494
stay_outcome_date: 2015	0.552	0.150	3.688	0.000	1.736
stay_outcome_date: 2016	0.653	0.161	4.047	0.000	1.921
stay_outcome_date: 2017	0.505	0.169	2.987	0.003	1.658
stay_outcome_date: 2018	0.522	0.175	2.989	0.003	1.686
stay_outcome_date: 2019	0.523	0.153	3.431	0.001	1.688
stay_outcome_date: 2020	0.682	0.214	3.179	0.001	1.977
stay_outcome_date: 2021	0.406	0.194	2.094	0.036	1.500
stay_outcome_date: 2022	0.099	0.202	0.487	0.626	1.104
city: Edmonton	-0.106	0.442	-0.240	0.810	0.899
city: Montréal	0.100	0.280	0.358	0.720	1.106
city: Ottawa	0.503	0.303	1.660	0.097	1.654
city: Toronto	0.697	0.259	2.693	0.007	2.008
city: Vancouver	0.356	0.297	1.199	0.230	1.428
city: Winnipeg	-0.278	0.418	-0.664	0.506	0.758