Data-Driven Futures of International Refugee Law

Byrne, William Hamilton; Gammeltoft-Hansen, Thomas; Piccolo, Sebastiano; Moller, Naja Holten; Slaats, Tijs; Katsikouli, Panagiota

Published in:
Journal of Refugee Studies

DOI:
10.1093/jrs/feac069

Publication date:
2023

Document version
Peer reviewed version

Document license:
Unspecified

Citation for published version (APA):
MOBILE Working Paper Series, no. 1, 2023

Data Driven Futures of International Refugee Law

William Hamilton Byrne, Thomas Gammeltoft-Hansen, Sebastiano Piccolo, Naja Holten Møller, Tijs Slaats, Panagiota Katsikouli

Forthcoming in Journal of Refugee Studies

This research is funded by the Danish National Research Foundation Grant no. DNRF169 and conducted under the auspices of the Danish National Research Foundation’s Center of Excellence for Global Mobility Law (MOBILE).

Jan 2023
Abstract:

As refugee law practice enters the world of data, it is time to take stock as to what refugee law research can gain from technological developments. This article provides an outline for a computationally driven research agenda to tackle refugee status determination variations as a recalcitrant puzzle of refugee law. It firstly outlines how the growing field of computational law may be canvassed to conduct legal research in refugee studies at a greater empirical scale than traditional legal methods. It then turns to exemplify the empirical purchase of a data driven approach to refugee law through an analysis of the Danish Refugee Appeal Board’s asylum case law and outlines methods for comparison with datasets from Australia, Canada and the United States. The article concludes by addressing the data politics arising from a turn to digital methods, and how these can be confronted through insights from critical data studies and reflexive research practices.

Author: William Hamilton Byrne, Thomas Gammeltoft-Hansen, Sebastiano Piccolo, Naja Holten Møller, Tijs Slaats, Panagiota Katsikouli
E-mail: William.hamilton.byrne@jur.ku.dk

MOBILE – Center of Excellence for Global Mobility Law – focuses on systematically studying the legal infrastructures of human mobility across geographies, social divides, travel patterns and time.

This research is funded by the Danish National Research Foundation Grant no. DNRF169.
Introduction

Whether we like it or not, refugee law is increasingly entering the world of big data. Across jurisdictions, both governments and refugee-assisting organisations are increasingly embracing forms of artificial intelligence (AI) and digital technologies as part of border controls, predictive migration management, refugee case handling or asylum procedures (Molnar and Gill 2018). The UNCHR and some NGOs, for example, has used forms of AI and machine learning to track and predict refugee and migration flows, and aid in the provision of humanitarian assistance (Casagran et al 2021; Andres et al 2020; Johns 2017). Similarly, a number of countries are currently testing digital tools which assist legal decision-making of e.g. family reunification or asylum applications (Kinchin 2021). Algorithms have further been developed to improve refugee integration, both locally and as part of resettlement procedures (Jones and Teytelboym 2018, 2017; Bansak 2018). The EU Commission thus anticipates that AI driven initiatives in asylum will be leveraged by most EU countries within the next decade, and the EU itself has already spent €341 million on such research since 2007 (EU Commission 2021; Kilpatrick and Jones 2022).

Some scholars have expressed optimism about what AI can mean for alleviating existing unfairness by promoting better decision-making in for instance refugee status determination (Cameron et al. 2022; Laqueur and Copus 2020; Bansak 2018). Others point to the fact that the use of AI in areas such as asylum raises significant ethical (Ajana 2015) and human rights (Land and Aronson 2020) concerns, and warn that the turn to AI may well become one of the next great challenges to international protection of refugees and asylum seekers (e.g. Molnar 2019).

Regardless of where one stands in regard to this debate, the fact remains that both practitioners and scholars of refugee law are likely to be confronted by these empirical developments. Any critical assessment of concrete AI application in refugee law will moreover require a degree of ‘big data literacy’ if we as scholars are to understand the algorithmic manipulations performed on large datasets and weigh the real or potential ethical impacts of data-driven decisions for asylum-seekers and refugees (D’Ignazio and Bhargava 2015). Such a move, however, may also benefit refugee law scholarship in other ways. The past decades has seen as growing field of computational law research embracing data-driven methods as a means to analyze legal text at a greater empirical scale than what is possible using traditional legal methods.

In this article, we seek to outline how computationally driven legal methods may be applied to gain new insights into one of the most persistent and foundational issues of refugee law - the gross
disparities of refugee status determination for similar asylum claims within and across jurisdiction. Despite decades of regional harmonization and a growing international jurisprudence, UNHCR states that even within the relatively standardized EU ‘recognition rates for the same asylum-seeking populations can vary from between 1% to over 50%’ (Foster 2012). In their seminal study, Refugee Roulette, Ramji-Nogales et al (2009) found significant variance between US judges within the same court, including one judge that was 18 times more likely to grant relief than his colleagues. Sean Rehaag has similarly described asylum adjudication in Canada as the ‘luck of the draw’ with one judge 50 times more likely to decide in favour of asylum applicants than other judges (Rehaag 2012). The issue of outcome discrepancy is thus not only cross-jurisdictional, but also institutionally endemic to refugee law.

Refugee law – and perhaps refugee studies at large – is a field that has been engendered to focus on ‘crisis’ over mundane problems (Charlesworth, 2002). Yet, such an orientation easily risks sideling or even normalizing outcome disparities in asylum decision-making as a recalcitrant problem for refugee law research. Existing research has suggested different explanations for this phenomenon, including the lack of compulsory guidance at the international level on as to how states should implement asylum procedures (Liodden 2020), or the predominantly factual and normative nature of asylum assessment, which inevitably results in some inconsistency between decision makers (Legomsky 2007). Social scientific research further indicates that asylum decision making can be significantly affected by both internal subjective dynamics (e.g. Noll 2005) or external pressures such as political events (Neumayer 2005). However, it is also an area where doctrinal legal and social scientific research seemingly reaches its end point: a problem so intractable and of such scale and scope – one of many of international migration’s ‘ordinary’ crises (Allen et al 2018) – that makes it difficult to document and unpack through traditional legal and qualitative methods.

We posit that computational legal methods hold significant potential to help overcome at least some of these limitations by making legal research more scalable. Computational legal research treats law as data: the treaties, case law, legislation, and administrative decisions that make up refugee law is also text that can be processed and analysed systematically with algorithms (Alschner 2021). Advances in the power of computation have enhanced but also made available on a wider basis the ability to analyse large bodies of data. Recent years have seen a growing embrace of empirical methods in refugee law scholarship, which has provided fertile ground for greater interdisciplinary integration (Byrne and Gammeltoft-Hansen 2020). NGO’s are experimenting with large datasets to analyse protection challenges (e.g. Kjærum 2020; Thomas et al 2019), and similar inroads are being
made in refugee law research (see e.g. Rehaag 2019; Byrne, Gammeltoft-Hansen et al 2022; Ghezelbash et al (forthcoming), Chen and Eagel 2017). However, the wider benefits of computational methods in refugee law research arguably still remains underexplored and is rarely reflected in more mainstream refugee law scholarship.

Building on these pioneering works, this article seeks to make a broader argument for what computational legal research can offer refugee law. We exemplify this via a case study of decision making at the Danish Refugee Appeals Board (‘RAB’), showing how large-scale legal analysis enhances our ability to focus in on patterns that emerge from the data (see also Drohout et al 2022). Addressing a key limitation of existing work in this area, we secondly discuss methods for comparing asylum adjudication data across jurisdictions, drawing on previous studies from Australia, Canada and the United States.

At the same time, however, we argue against perceiving computational methods as a panacea or displacing the role of more qualitative forms of assessment. The quantification of legal decisions can also reinforce human biases that inhere within the scholar’s selection criteria, but also more fundamentally within the legal text itself (see also Baden et al 2021). Equally important, we thus underscore the importance of closely engaging insights from critical data science in terms of understanding the limitations and risks of datafying the legal hermeneutic through a computational approach to refugee law research.

The analysis proceeds as follows. Part I sets the frame for analysis with an introduction to the field of computational legal methods in context of a growing empirical turn in refugee law scholarship, and canvasses some of its principal implications as a field that manages twin demands of knowledge production and advocacy. Part II provides a case study of the application of computational legal research by showing how these methods can provide new insights into asylum determination based on a dataset of Danish case law, by exposing the discrepancies that arise in rates of acceptance for specific groups of claims based on religious conversion, and sexual orientation. Additionally, through the analysis of individual panel recognition rates we provide tools from statistics to understand and measure outcome variation. Part III concludes reflexively by raising some of the key normative challenges arising from the incipient use of artificial intelligence in asylum decision making, and outlines some key remedial measures that keep the law central to analysis, but confronts its harms through a turn to socio-technical data ethics.
Computational Refugee Law: Between Theory and Practice

Understanding the nature and scope of computational legal research is a necessary starting point for any discussion of its possible application to refugee law. This is particularly important since refugee law scholarship draws from a range of international, domestic, and comparative forms of legal research. An application of computational methods in this area furthermore takes on necessary and specific features, which impose acute sensitivities for any refugee law researcher.

At its most basic, computational legal research treats law as data. As such, computational methods have found particular traction as part of legal text analytics, the process of ‘deriving of substantively meaningful insight from some sort of legal data’ (Ashley 2017: 5). Some legal scholars, including within refugee law, have resisted such a perspective as epistemologically reductionist (Legomsky 2007). Others, however, have embraced the new opportunities such methods facilitate by enabling research tasks that would simply not be possible without the processing power of computation.

Computational legal research aims to identify patterns in legal materials by treating legal texts or outcomes as data and subjecting this to algorithmic analysis. In principle, no forms of ‘text’ are off limits for this line of research (Ashley 2017: 9), though the ability to compile quality data is a necessary and often difficult prerequisite (Sambasivan et al 2021). The majority of this research has thus far taken the form of converting written legal texts, such as judicial decisions, statutes, and international treaties, into a machine-readable form and analysing the texts on the basis of its elements of discourse. Whilst the application can focus on legal doctrinal questions, it can also extend to issues concerning the practice of law and the nature of the legal system, subjects traditionally more associated with socio-legal and philosophical analysis. (Whalen 2020: 3). Consequently, computational methods are now receiving broader purchase as part of a wider ‘information revolution’ (Kobayshi and Ribstein 2011) in legal research in several areas traditionally dominated by doctrinal legal methods (Olsen and Sadl 2017).

From a scholarly perspective, interest in computational legal research may be seen as closely related to the broader ‘empirical turn’ in legal scholarship, which has sought to ‘provide scholars with tools to gain new facts, see existing ideas through a different lens, and engage in a more nuanced analysis of international law…’ (Franck 2008). For some, it represents an integral part of a theoretical agenda to reform the legal discipline in line with more ‘scientific’ methods (Madsen and Holtermann 2015). For others, computational methods have been embraced from a more pragmatic perspective, as simply helping lawyers find ‘what works’ (Schaffer 2015). As Olsen and Sadl (2017)
point out, computational legal research does not necessarily stand in contrast to more traditional and practice-oriented legal methods, but can enhance ‘the validity, reliability, and transparency’ of doctrinal research whilst preserving its unique capacity for conceptual analysis and attention to legal detail.

Beyond its theoretical heritage, the computational turn in law is ultimately a product of and responsive to profound changes in the way we receive and process legal information. Law is becoming a ‘big data phenomenon’ as much of our activity has shifted to the online world (Alschner 2021). Refugee law scholarship has for quite some time absorbed perspectives from other disciplines, including political science, sociology, anthropology, psychology, and forensic science (Byrne and Gammeltoft-Hansen 2020). International refugee law has also long been the subject of close text analytics by both legal and social science scholars, and its sources are increasingly available in machine-readable forms. It frequently draws from large corpora of both domestic and international law (see e.g. Hathaway 2021), and its scholars perform a unique role in the absence of a central court at the international level in translating, linking, and synthesizing case law’ across national legal traditions (Byrne and Gammeltoft-Hansen 2020, see also Ulen 2003). As such, computational methods may significantly enhance individual scholars’ ability to analyse and compare domestic implementation of international refugee law on a wide scale basis, and thus enhance the rigor of both empirical and doctrinal research, but also bring to the fore the mutual interdependence of these research traditions (Davies 2020).

More generally, computational legal research can broaden empirical analysis in a way that permits greater ability to ‘test, validate, and refine existing frameworks’ and ‘provide empirical evidence to validate hunches and prove legal intuitions correct’ (Olsen and Sadl 2017). Computational methods render legal analysis scalable and empowers researchers to study international law with greater ‘depth and breadth’ and gain ‘a more stable and comprehensive quantitative basis’ for incisive analysis of a research question (Alschner 2021). In this way, observing variation – or lack thereof – in large data sets provides a navigation point and justification for more directed qualitative research. However, it can also be a form of knowledge in itself, as seminal studies have shown (see e.g. Ramji-Nogales, et al 2009; Rehaag 2012) much of what we know about the operation of refugee law in practice is generally a product of impressionistic assessment.

Computational research enhances the efficiency of reading, but also brings certain challenges. In particular, there is a growing consensus that AI cannot replace the legal doctrinal method. Computational methods ‘are no substitute for careful thought and close reading and require extensive
and problem-specific validation’ (Grimmer and Stewart 2013: 267) and ‘cannot (yet) differentiate between legally significant and legally insignificant text differences’ (Alschner 2021). Data modelling is inherently dependent on the making of data, where the downstream effects result in technical debts over time (data cascades) if not taken seriously (Sambasivan et al 2021). As such, it can introduce systematic biases into the data. Furthermore, computational methods extract specific text and their associated legal meaning arises from ‘multiple and complex contents’ (Baden et al 2022). Simple quantification of observations do not in themselves make them more (or less) true (Passi and Jackson 2017; Møller et al 2021). Hence, asking qualitative ‘small’ questions to quantitative ‘large’ datasets is equally important for data sense making.

By nature of the legal discipline, this will often require a keen eye on the practical element. However, the point also applies at the more theoretical level. As computational legal research replicates a specific (legal) hermeneutic, the underlying theory of refugee law must be subject to critical interrogation. While computational analysis has played a major role uncovering and bringing attention to systematic discrimination and prejudice in human decision-making (Kleinberg et al 2020) and may possibly help to counteract cognitive biases in asylum judging (Ghezelbash, Dorostkar and Walsh, forthcoming), there is an equal risk that computational methods end up replicating existing biases not only in the data, but also based on scholarly preconceptions. While there is an increasing awareness of this issue in computer science, it will always be difficult to reduce the influence of such tendencies. Proposals for good practice include e.g. enabling open datasets and fostering broader machine learning and big data literacy to address current data inequalities (Sambasivan et al 2021; Singh 2021).

As Colin Harvey (1999) has pointed out, it is important for refugee law scholars that are engaged in empirical forms of research to lay bare the normative commitments that underpin their research. Surmounting this epistemic challenge is ultimately a question of reflexivity that will be given special treatment later in this paper. At this stage, it should simply be underscored that computational legal research is not a radical reformulation of the legal method but may help it to see itself better. With the increasing datafication of our world the question is increasingly not whether, but how insights of computation best can be adapted to both national and international bodies of refugee law.
Understanding Outcome Disparity in Refugee Status Determination

Refugee status determination (‘RSD’) is a quintessential focus area of refugee law. A large part of its doctrinal analysis focuses on systematizing the definition of who is a ‘refugee’ and who can claim complementary protections under international law. This literature further presumes that an asylum seeker will be afforded RSD within the terms of the law. However, not only is individual RSD far from universally practiced, RSD is also largely unregulated by international law.¹ As a result, states exercise a high degree of discretion over RSD by invoking its absence from the 1951 Refugee Convention as a matter left to national law (Gorlick 2003). RSD is moreover constrained by epistemic uncertainties, which allow for divergent interpretations of legal rules and enable ‘local yardsticks’ on the process for defining a refugee (Liodden 2020). If RSD is the ‘gateway to protection’ (Kagan 2006) in many, if not all, parts of the world, it remains remarkably inconsistent between, sometimes also within, national jurisdictions.

This issue has proven difficult to challenge through doctrinal legal research. Whilst the UNHCR (2001) has long suggested that RSD must be ‘fair and effective,’ and refugee scholars have pointed out (e.g. Legomsky 2000) that unfair procedures unjustifiably risk refoulement, doctrinal analysis remains limited by what it can take from formal sources of law. As such, refugee law scholarship has mostly focused on problematic circumstances of RSD (e.g. Dastyari and Ghezelbash 2020) or specific national procedures (e.g. Edwards 2006), but even across states with similar legal traditions (see e.g. Kneebone 2008). Prescriptions for best practices have proven difficult to implement, as evident with the EU asylum acquis, which despite common rules has not dispelled significant outcome variation among the Member States (Trauner 2016; Toshkov and de Haan 2013).

What then makes RSD as the great enigma of refugee law scholarship so recalcitrant? Scholars suggest inconsistency may arise through differences in institutional design, procedures and legal interpretations across jurisdictions (Simeon 2010), but also within specific national asylum systems (Vedsted-Hansen 2005). Others point to differences in standards of evidence (Byrne 2007), or problematic use of fact-finding (Gibb and Good 2013) and medical evidence reports (Wallace and

¹ The principal exception being soft law instruments, such as the UNHCR Handbook Procedures and Criteria for Determining Refugee Status under the 1951 Convention and the 1967 Protocol relating to the Status of Refugees.
Wylie 2013). Some suggest that variation can arise from a lack of financial resources or the potential for political interference to judges (Rehaag 2012). Moreover, as Legomsky (2000) notes, some inconsistency in RSD is inevitable because it reflects ‘degrees of specialization, complexity, dynamism, emotional or ideological content, and spectrum of choice’. Yet, even within countries with similar legal traditions, RSD looks very different and is significantly constrained by a variety of socio-institutional factors (Hamlin 2014.) As such, the issues arising in RSD are not easily resolved within the confines of doctrinal law.

In response, interdisciplinary research suggests that law is just but one factor explaining outcome variations. Socio-legal studies propose that rules and standards may be applied in an ad-hoc fashion (Anker 1992) whilst others have argued that the process of refugee credibility assessments is innately intersubjective (Noll 2005). Some research suggests that asylum procedures are prone to forms of explicit and implicit bias (Marouf 2010) and subject to e.g. gendered dynamics (Millbank 2003) or performative scripts (Beard and Noll 2009). Relatedly, public health research has found that cultural recognition, or the lack thereof, between applicants and decision-maker impacts outcomes (Montgomery and Foldspang 2005). Medical studies point to the impact of mental health issues among applicants (Fängström and Sarkadi et al 2019), as well as stress amongst decision-makers (Lustig, Karnik et al 2008). Political science research has alternatively suggested that RSD is impacted by highly extraneous factors, such as past application and acceptance rates, unemployment, prevailing political majorities, judges’ party affiliation and major political events (Neumayer 2005; Rottman et al 2009).

Interdisciplinary research has been instrumental in terms of expanding the analytical gaze in terms of the elements impacting RSD. Yet, existing research usually focus on smaller case studies or specific national contexts, which makes it difficult to explain outcome variation between states. As Noll (2005) points out, research in this area is moreover fragmented along disciplinary divides, which can limit the generalisability of findings and thus further theorising. In sum, as Gorlick (2005) notes, ‘perhaps there is no standard answer for differential recognition rates among like-minded states’.

Nonetheless, recent large-scale empirical research into RSD have offered new insights into the assumptions that pervade these processes within or across national legal systems. Firstly, a number of researchers have used discourse methods to examine corp of legal decisions to identify the presence of ‘non-legal factors’ that can pervade RSD. For instance, Smith-Kahn (2020) has used corpus linguistics to expose the the ideologies underlying asylum discourses in RSD, whilst Herlihy, Gleeson and Turner (2010) have similarly unveiled assumptions about human behaviour that pervade
credibility assessments. A second strand of research has used observational studies to explore causal relations between behaviours and legal processes, as in Rehaag and Cameron’s (2020) simulated study of RSD that used student participants to examine whether an applicant’s appearance might affect the RSD process.

Computational legal methods may further enable scholars to examine these causative elements of RSD, especially where datasets reach sizes where it is difficult, if not impossible, to work with manual coding. Three recent studies exemplify this purchase. Firstly, Chen and Eagel (2017) have used machine learning to analyse 492,903 asylum decisions of 441 United States judges and develop a predictive model of decision making with an accuracy of 82%, which lead the authors to suggest that RSD seems mostly driven by trend features and individual judge characteristics. Ghezelbash, Dorostkar and Walsh (forthcoming) have used computational methods to analyse 6,700 applications asylum decisions of the Australian judges and found variations that may arise from the burden of judges and the role of legal representation. Similarly, a recent study by Rehaag (2019) analysed 30,000 Canadian asylum judgments and found that the outcome of judicial review of asylum cases largely depended on the judge assigned to the panel. These findings echo similar studies undertaken in France by the Supra Legem Project which found significant disparities associated with the background of individual judges and similarly developed a predictive algorithm (see Langford and Madsen 2019). Overall, these studies suggest that the most determinative variables in decision outcomes are the country of nationality and the judge assigned to the case.

*Insights from the Danish Refugee Appeals Board data*

We now move from extant research to provide examples from our own research into large case file data sets from the Danish Refugee Appeal Board (*Flygtningenævnet*; hereinafter, ‘RAB’). As background, the asylum process in Denmark is two-tiered. First instance decisions are made by the Immigration Service. Decisions rejected at first instance are automatically appealed to the RAB, which means that a large proportion of asylum decisions go to appeal. The RAB is a quasi-judicial body and has full legal competence to assess questions of fact and law and in practice serves as the final instance, with no possibility to appeal its decisions to the ordinary courts (see Gammeltoft-Hansen and Ford 2022). The current composition of the RAB hears asylum cases by a board consisting of a chairperson, who is also a judge within the ordinary Danish court system, as well as two other members, representing the Danish Immigration Ministry and the Danish Bar Association. Previous periods have seen different compositions and board sizes (ranging from three to seven
members), involving e.g. members nominated by the Danish Refugee Council, an NGO, thereby enabling inquiries into the effect of who and how many partake in RSD.

Our dataset contains 15,535 asylum cases tried by the RAB in the period of 1995-2020. The case files were obtained by way of special agreement with the Danish Refugee Council, and includes a large proportion of case files that are otherwise not publicly available. The case files are in the form of a judicial decision and stipulate not only the legal reasoning underpinning the decision, but also the procedural history, relevant facts and documents, as well as interview transcripts with the applicant. This is a rich dataset, which allows us to ask complex and sophisticated research questions and enables a whole range of novel large-scale analyses, such as identifying recurrent argumentation grounds, tracing the process and steps of asylum adjudication since the applicant enters Denmark until the final decision, understanding why the RAB overturns first instance decisions, and inspecting the presence of systemic biases. However, also due to the confidential nature of this data, the aforementioned data richness comes with additional costs and challenges.

The data comes as a collection of ~30,000 files including Word documents and PDF files, as well as a number of temporary files and operating system specific metadata files that are unreadable or do not contain any case information. The temporary files come from older Word documents and were likely corrupted during subsequent processes of transferring the digital archive to new platforms. As such, the total number of accessible case files available for analysis is reduced to 15,535. Most of the PDF files were obtained by scanning paper documents. This means that most of the PDF files are not directly machine-readable and we had to rely on optical character recognition (OCR) techniques to convert the documents in machine-readable text. Furthermore, no metadata is available. Therefore, after we obtained a collection of machine-readable texts, we extracted all the information of interest from the text of the decisions. While NLP models to extract information from texts exist for major languages (English in particular), the availability of NLP models for the Danish language is still quite limited (Kirkedal et al 2019). To remedy this, we extracted the information of interest through regular expressions (a technique to define and match textual patterns in texts), leveraging the quasi-structured organisation of the texts. The information extracted include nationality, decision date, age, marital status, and sex of the applicant (explicitly written in the decision), sex of the chairing judge (inferred from their name), and decision outcome. Describing the whole data pre-processing and extraction pipeline is beyond the scope of this paper, but is a task that – at least for this type of data and small language jurisdiction – should not be underestimated.
The information extracted firstly allows us to analyse some factors that extant literature has deemed important to explain biases or differences in asylum recognition rates, such as the sex of the adjudicator or the family status of the applicant (cf. Ramji-Nogales et al. 2009). In our data, we find no significant statistical difference for these factors (Figure 1, above). This suggests that previous findings about outcome variations might not be generalisable across all countries and jurisdictions. In our case, the fact that we do not observe difference in recognition rates based on the sex of the chairperson might be due to the fact that decisions are taken by a panel rather than by a single judge. Other factors that appear not to be relevant to explain outcome variation (not shown in the figure) are the sex and the age of the applicant – again somewhat different to findings of research in other jurisdictions (Cf. Ramji-Nogales et al. 2009). Where we do find significant differences is in recognition rates between the applicants’ country of origin, which is also to be expected. As we expand our dataset to include similar case law repositories from other Nordic countries, cross-jurisdictional differences in recognition for specific nationalities of claims may however form an important starting point for more in-depth analysis and correlation with other factors.

While cross-jurisdictional differences in recognition rate based on nationality is often used as an indicator for bias in the existing literature, it is important to stress that it may also be explained by other underlying factors. For instance, networked-based effects in terms of the geographic origin or type of claimants, shifting availability of internal flight alternatives, or differences in each country’s use of background information and country of origin reports. This discussion prompts us to highlight a general difficulty faced by researchers and practitioners when they desire to assess the presence of bias. Observing a statistically significant outcome variation on some factor (such as the country of origin or judge identity) cannot automatically be inferred as bias. Yet, deeper analyses and richer data
that can help researchers to control for alternative explanations are needed and represent a promising venue for future research.

Understanding the development of specific case types
Having access to a broad repository of asylum appeals within a given jurisdiction allows us to study specific types of claims in more detail. Here, we explore outcome variation for two types of claims – sur place cases arising from post-departure religious conversion and claims based on sexual orientation. Both types of cases have garnered increasing attention as a matter of both practice and scholarship, due to the difficulties and highly varied approaches by authorities assessing the credibility of such claims (Berlit et al 2015; Rose and Given-Wilson 2021). Moreover, both classes of cases have only relatively recently emerged as somewhat common-place within the international normative framework (see e.g. Rehaag 2009; da Lomba 2011). We extracted both types of cases through relevant keyword searches.
Figure 2 shows the evolution of these two types of claims in terms of yearly number of cases and recognition rates. Both sexual orientation and religious conversion cases saw significant growth at the RAB since 2010. Part of the explanation for the marked increase in cases may be related to the change in legal practice by the RAB, inspired by similar developments at the European level (Berlit et al 2015), meaning that in some cases asylum applicants from these groups no longer need to demonstrate an individual risk of being profiled by authorities or non-state persecutors. For these two groups, the recognition rate exhibit a visible jump from 2010 as opposed to the rest of the cases, which show a more stable recognition rate over time. However, after this jump, while the number of sexual orientation and conversion cases keeps increasing, the recognition rates remain stable. The explanations for this effect may be multiple. One possibility is an increase in real or perceived
manufactured claims, spurred by the more lenient legal practice. But it may also be interpreted as the judicial system ‘auto-correcting itself’ by adopting a more restrictive standard for credibility assessments in light of the legal practice change and ensuing increase of applications under particular categories. This would mirror findings in other studies. Daniel Chen and others (2016) for instance have shown that the US asylum system ‘negatively auto-correlates’ in light of an increasing suspicion that sequential success does not occur by chance, particularly with respect to claims which adjudicators might suspect of fabrication.

Understanding and measuring outcome disparity among judges

Several studies (Ramji-Nogales et al. 2009; Rehaag 2019; Ghezelbash, Dorostkar and Walsh, forthcoming) have shown that recognition rates vary between judges. The overall idea behind this stream of research is to point out that RSD is dependent on judge identity and that the allocation of cases to judges may constitute a procedural ‘lottery’. Here, we conduct a similar analysis on our data as well as data for other jurisdictions: the rates of the US judges for the period 2016-2021 (obtained from the TRAC website), the rates of the Australian judges (as reported in Ghezelbash and Dorostkar and Walsh, forthcoming) and the overall rates of the Canadian leave judges in the judicial review stage for the periods 2008-2011 and 2013-2016 (as reported in Rehaag 2019). With this analysis, we have three goals. First, providing legal scholars with tools from statistics to assess the variability of recognition rates in relation to judges; we do this by introducing the concept of statistical distribution and a number of measures to quantify the variation. Second, showing that not all “lotteries” are created equal; we do this by comparing the shape of the distributions for the jurisdictions considered. It should be noted that we do not compare the recognition rates of the judges of one country with those of the judges of another country. Similar comparison would be influenced by all the confounding variables discussed in the previous two sections. Third, showing that it is possible to evaluate the effects of intervention strategies; we do this by comparing the distributions of the Canadian data for the two periods because, as reported in Rehaag 2019, between the two periods the Canadian Federal Court implemented an intervention strategy to reduce the disparity in recognition rates among the judges.

The typical way to evaluate outcome disparity among judges is through a rank plot; that is, by plotting in increasing (or decreasing) order each individual recognition rate. Figure 3 shows the rank plot for the judges of each country. The rank plot is useful to understand the range of variation (i.e. minimum and maximum): for instance, in Denmark the range is [0, 0.44] while in USA the range is
However, it is not easy to understand the amount of variation from the rank plot. For instance, it is not easy to estimate the average recognition rate for US or to see that the US data exhibits the largest inequality in recognition rates. Fortunately, the same data can be represented as a distribution of values (density plot), which makes it easier to understand the amount of variation (Figure 3, subplot ‘Comparison’). The density plot shows the recognition rates on the x-axis and the density on the y-axis. The density values are illustrated as a normalized count, meaning that the area under the curve sums up to one.

---

2 We obtained the density plot through a technique called ‘kernel density estimation’ which is available in many statistical packages. For this study we have used the standard ‘density’ function available in the R programming language.
Figure 3: Rank plots and density plot showing the judges' recognition rates for the jurisdictions considered here

From the density plot it is easy to see that the distribution of the recognition rates for the US judges is skewed, with most of the judges (around 75%) having a recognition rate somewhere in between 0 and 0.4; accentuated differently, 25% of the US judges grants 66% of the applications. Compared to the US distribution, all the other jurisdictions exhibit a markedly different distribution. They exhibit a clear peak, which indicates the existence of a typical recognition rate, and their curve is almost symmetrical around this peak, indicating a ‘bell-shaped’ distribution. Recognition rates in these jurisdictions are clustered around an average and the variation follows a specific pattern. In other words, the probability of observing a judge with a higher (lower) recognition rate than the national average decreases as the magnitude of the difference – between the judge recognition rate and the average rate – increases. The tighter the bell, the lower the variation (or dispersion around the
average) is. Returning to the US data, this distribution does not have a well-defined average and the dispersion around the average is asymmetrical and right-skewed. As such, the US data exhibits the highest degree of dispersion.

The dispersion of the distribution tells us something about the consistency of the legal system. If cases are randomly assigned to judges, the variation observed in countries that exhibit a bell-shaped distribution is less likely to result from individual bias among adjudicators. Adjudicators that exhibit overly large or overly low recognition rates can be understood as outliers not descriptive of the overall system – although outliers may also have an important impact on the system. Under this understanding of consistency, the US system appears to be less consistent because of the variation pattern. However, two caveats apply. Firstly, if the allocation of cases to judges is non-random, this type of variation may be explained on other grounds. Secondly, a certain degree of variation is to be expected given the specific circumstances of individual cases, and its absence would be equally concerning from a legal perspective (Legomsky 2007). Nonetheless, too high degree of variation across judges’ recognition rates as well as skewed distributions may both be seen as indicative of individual bias.

Now that we have an intuitive understanding of statistical distributions, typical values, and dispersion or variability, let us turn our attention to the Canadian data. This data may be used to make a comparison before and after an initiative from the Canadian Federal Court to reduce the disparity in recognition rates among individual judges. By comparing the two distributions, we can observe that the peak of the distribution in the period 2013-2016 is higher than the peak in the period 2008-2011, which means that the average recognition rate has increased. Again, one should be careful not to make strong inferences based on this observation alone, since the composition and type of applications may also had shifted between the two periods. However, the Canadian distribution in the period 2013-2016 is less skewed than the distribution for the period 2008-2011. This indicates that the inequality in the judges’ recognition rates has been reduced. While there could be other factors at play here as well, this effect is more likely to be linked to the intervention strategy.

In sum, this part of our analysis suggests that not all asylum ‘lotteries’ are created equal, and how data-based comparisons in terms of variations across individual judges or panels across jurisdictions can be meaningfully made, even if the underlying asylum systems retain significant differences. We believe that distributions convey more information than rank plots, despite being based on the same data. Finally, distribution plots can facilitate inter-temporal comparisons to test the effects of intervention strategies, as in the case of the Canadian Federal Court.
Mathematical tools to measure outcome disparity

In the previous section, we introduced distributions, typical values and variability/dispersion. In this section, we illustrate some mathematical tools to formally measure these concepts. We can quantify the typical value (central tendency) by computing the mean, median, and mode (the peak of a distribution), while the dispersion can be quantified by computing the entropy, or the Gini coefficient. We also introduce the beta distribution (see Balakrishnan and Nevzorov 2003) as a tool to model recognition rates and to compute entropy and mode – all the other measures can be computed directly from the set of recognition rates.

![Recognition rates distributions for each country modelled with a Beta distribution. The plot show that the theoretical models are a good fit for the empirical data.](image)

The Beta distribution is a flexible tool to model rates or probabilities (more generally values between 0 and 1) and it is defined by a two parameters ($\alpha, \beta$). The parameters can be found by fitting the theoretical model to the empirical data through a technique called maximum likelihood.
estimation. Figure 4 shows the empirical cumulative distribution and the theoretical Beta distribution for each country. The cumulative distribution is another way to visually represent the same information contained in rank plots and distribution plots. The cumulative distribution shows, on the y-axis, the proportion of judges that have a recognition rate lower or equal than a given value (on the x-axis). For example, around 80% of the Canadian judges in the period 2008-2011 have a recognition rate lower or equal than 10%.

Similarly, Table 1 (below) shows measures of central tendency and dispersion for the countries under exam. If there is a well-defined typical value, the mean, median, and mode will have similar values, otherwise they will change significantly. We can see that the latter is the case for the US data, confirming our intuition in the previous section about the lack of a typical value. In particular, the distribution peaks around zero (mode). In contrast, the distributions for the other jurisdictions have a clear typical value. The measures of dispersion, consistent with our intuition, show that the US data exhibits the highest degree of variation. The measures of dispersion do not always agree in how they rank the distributions. This happens because they measure the inequality of a distribution in different ways and with different reference models. The Gini coefficient defines inequality on the basis of a highly skewed distribution, the entropy on the basis of the uniform distribution; that is, a distribution for which each value is equally probable. The disagreement between entropy and the Gini coefficient shows that measuring variation is not a trivial task and the measure needs to be chosen carefully. The Gini coefficient is the measure that best captures the intuition about variation that we developed in the previous section. The Gini coefficient for the Canadian data in the two periods confirms our intuition that outcome variation has seen a reduction in the second period.

\[ \text{estimation}^3 \text{. Figure 4 shows the empirical cumulative distribution and the theoretical Beta distribution for each country. The cumulative distribution is another way to visually represent the same information contained in rank plots and distribution plots. The cumulative distribution shows, on the y-axis, the proportion of judges that have a recognition rate lower or equal than a given value (on the x-axis). For example, around 80\% of the Canadian judges in the period 2008-2011 have a recognition rate lower or equal than 10\%.} \]

\[ \text{Similarly, Table 1 (below) shows measures of central tendency and dispersion for the countries under exam. If there is a well-defined typical value, the mean, median, and mode will have similar values, otherwise they will change significantly. We can see that the latter is the case for the US data, confirming our intuition in the previous section about the lack of a typical value. In particular, the distribution peaks around zero (mode). In contrast, the distributions for the other jurisdictions have a clear typical value. The measures of dispersion, consistent with our intuition, show that the US data exhibits the highest degree of variation. The measures of dispersion do not always agree in how they rank the distributions. This happens because they measure the inequality of a distribution in different ways and with different reference models. The Gini coefficient defines inequality on the basis of a highly skewed distribution, the entropy on the basis of the uniform distribution; that is, a distribution for which each value is equally probable. The disagreement between entropy and the Gini coefficient shows that measuring variation is not a trivial task and the measure needs to be chosen carefully. The Gini coefficient is the measure that best captures the intuition about variation that we developed in the previous section. The Gini coefficient for the Canadian data in the two periods confirms our intuition that outcome variation has seen a reduction in the second period.} \]

---

^3 in R the parameters of a beta distribution can be estimated with the function `beta.mle` contained in the package `Rfast.`
Table 1: measures of central tendency (mean, median, and mode) and dispersion (Gini coefficient and entropy) for the countries under exam. In square brackets, we report 95% confidence intervals computed through the bootstrap procedure, by resampling the data 1000 times. The bootstrap confidence intervals were obtained through the function BootCI in the package DescTools, in the R programming language. Mode and entropy have no confidence intervals as they were computed analytically. Values in the table are round to the second decimal digit.

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Gini</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>0.20</td>
<td>0.20</td>
<td>0.18</td>
<td>0.21</td>
<td>-1.15</td>
</tr>
<tr>
<td></td>
<td>[0.18, 0.23]</td>
<td>[0.18, 0.22]</td>
<td></td>
<td>[0.14, 0.27]</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>0.09</td>
<td>0.08</td>
<td>0.04</td>
<td>0.39</td>
<td>-1.47</td>
</tr>
<tr>
<td></td>
<td>[0.07, 0.12]</td>
<td>[0.06, 0.12]</td>
<td></td>
<td>[0.27, 0.47]</td>
<td></td>
</tr>
<tr>
<td>Canada (Pooled)</td>
<td>0.09</td>
<td>0.09</td>
<td>0.08</td>
<td>0.34</td>
<td>-1.60</td>
</tr>
<tr>
<td></td>
<td>[0.08, 0.10]</td>
<td>[0.08, 0.10]</td>
<td></td>
<td>[0.29, 0.39]</td>
<td></td>
</tr>
<tr>
<td>Canada 2008-2011</td>
<td>0.07</td>
<td>0.07</td>
<td>0.04</td>
<td>0.38</td>
<td>-1.77</td>
</tr>
<tr>
<td></td>
<td>[0.05, 0.09]</td>
<td>[0.05, 0.08]</td>
<td></td>
<td>[0.27, 0.48]</td>
<td></td>
</tr>
<tr>
<td>Canada 2013-2016</td>
<td>0.10</td>
<td>0.10</td>
<td>0.08</td>
<td>0.29</td>
<td>-1.59</td>
</tr>
<tr>
<td></td>
<td>[0.09, 0.12]</td>
<td>[0.07, 0.12]</td>
<td></td>
<td>[0.23, 0.33]</td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>0.31</td>
<td>0.24</td>
<td>0</td>
<td>0.43</td>
<td>-0.22</td>
</tr>
<tr>
<td></td>
<td>[0.29, 0.33]</td>
<td>[0.22, 0.26]</td>
<td></td>
<td>[0.41, 0.45]</td>
<td></td>
</tr>
</tbody>
</table>

Opportunities for Further Research

Our analysis thus far has shown the potential for computational methods in helping to lay bare differences in asylum decision-making and their underlying elements. Yet, our own research on this issue is still at an early stage. As such, we will briefly detail some potential future applications for this line of research by drawing on similar research in adjacent fields of law.

In the above example, we used data science analysis to systematize the variables of asylum outcome variation. Further research could scale up complexity by employing forms of machine learning (‘ML’). ML in legal research has mostly been used to create predictive models for judicial decision making by training algorithms to classify or predict unlabelled text through either ‘supervised’ or ‘unsupervised’ methods (Alschner 2021). In refugee law, ML could be employed to identify patterns of consistency of asylum decisions between jurisdictions and the more common factual and legal bases of outcome variation. For instance, Alschner and Skougarevskiy (2016) have used supervised ML to map investment treaties and track convergence and normative consistency across large data sets through the common terms employed in legal provisions. Applying this method to refugee case law may similarly reveal divergences of legal interpretation both within and between
countries. The concept of ‘persecution’ provides a relevant example for this kind of research, given the range of different interpretations and the inherent scope for subjective considerations that arise in the assessment of this crucial element of refugee claims (see e.g. recently Storey 2022).

ML may further be used to autonomously identify crosscutting topics that escape human reading across large datasets of refugee law. For instance, Sadl, Panagis, and Christensen (2016) have used topic modelling to show the shifts in the content of EU law over time on both legal factual bases, whilst other researchers have applied ‘deep learning’ AI methods to track the implementation of EU law through semantic and topic similarity (Nanda et al 2019). An interesting avenue for such research is to probe epistemological issues arising in credibility assessment. For instance, ML may help discover keywords that arise in credibility assessment between jurisdictions and track their inferences as a basis for exploring the heuristics of decision-makers. ML could also be used to identify how asylum decision-makers apply rules to individual facts when assessing an applicant’s story by isolating semantic placement. Further traction can be gained through the use of ‘process mining’ to better understand the asylum decision-making workflow and identify how differences in the handling of individual cases influences their outcomes. Some governments have shown interest in this method as a way to expand fast-track and accelerated procedures (Capgemini 2020). While process mining can in principle help streamline asylum processing, identify procedural delays and promote outcomes that are more balanced (see e.g. Slaats 2020), there is limited empirical research into whether such tools actually work to achieve these aims in this area.

ML analysis may also help practicing immigration lawyers or decision-makers identify relevant case law from other countries. Existing scholarship suggests that national interpretations of refugee law remain fragmented (Lambert and Goodwin-Gill 2010). Similarly, textbooks and soft law guidelines may be geographically biased in terms of the case law they rely on (Bailliet 2015). Mining case law data across jurisdictions can thus be used to establish searchable networks of similar cases based on underlying characteristics (see e.g. Alschner and Charlotin 2018). Another opportunity is to search for coherence in the use of materials that have proven resistant to authoritative treatment by judges (Thomas 2008), such as country guidance, NGO reports, UNHCR recommendations and other soft law documents (Goodwin-Gill 2010). Promoting stronger links and more exchange between national legal understandings and approaches in this manner could potentially foster more coherent applications of international refugee law.

Finally, computational methods may be used to analyse the impact of external socio-political factors and events on asylum decision-making. Previous studies have found broad correlation
between specific events (such as the September 11 terrorist attacks) and overall rates of asylum acceptance (Neumayer 2005; Rottman et al. 2009). More computationally driven approaches have used sentiment analysis to examine the role of a discourse of ‘achievement’ in the case law of a court (Potts and Kjær 2016). NLP has similarly been used to examine how courts grapple with politically sensitive issues (Busch and Pelc 2019). Similar research in refugee law could examine how asylum decision-makers deal with the political context with reference to the language employed in their decisions. A body of work in the digital humanities has used computational methods alongside critical discourse analysis to detect bias against refugees in public texts such as news materials (see e.g. Haider 2019). An opportunity for refugee law research may thus be to triangulate the political context of RSD by finding correlations between grounds for rejection in law, on the one hand, and policy announcements, broader legislative changes on immigration issues, or public opinion data on refugee issues, on the other. Computational methods may thus help promote greater integration between refugee law and refugee studies in helping to address crosscutting research problems.

In sum, this section has sought to exemplify how data-driven approaches can garner new insights on the determinants of asylum decision-making. We have further proposed some as of yet unexplored paths for computational refugee law research to unpack the epistemic and legal uncertainties in RSD, promote more coherence amongst national RSD systems, and assess how external events and the wider socio-political context might engender outcome variation in asylum decision-making. However, it is necessary to bear in mind that several of these applications may simultaneously face empirical, technical, and normative challenges, which must be borne in mind when engaging with computational legal research. In the final section, we address these in turn.

Towards Reflexive Strategies in Computational Refugee Law Research

It is now widely accepted that machine learning has a ‘normative mirroring effect’ – it reproduces ‘not only the beneficial and desired but also the biased, skewed and discriminatory’ (Larsson 2019). Decades of research on the implementation of machine learning suggests that some forms of data analysis hold a disparate impact and are significantly (or even systematically) discriminatory in their effect, and shape distributional outcomes between the powerful and the oppressed (Barocas and Selbst 2016). Machine learning can be impregnated with human assumptions about race (Benjamin 2019), gender (Hoffman 2017) or class (Eubanks 2018), which has led several scholars to argue that AI can
become a threat to democratic frameworks (Hildebrandt 2016). As such, any turn to computational refugee law – even one motivated by a critical approach to existing governmental practices – must equally confront the challenges that arise from engaging in big data analysis, and reflexively question the researchers’ own data practices.

To this end, a computational refugee law can be enriched by taking cues from critical data studies (‘CDS’). A key insight from this body of scholarship is that big data analysis is never merely empirical or neutral, but always also normative in nature (Illadis and Russo 2016). As Gould (1981) reflects, ‘inanimate data can never speak for themselves, and we always bring to bear some conceptual framework, either intuitive and ill-formed, or tightly and formally structured, to the task of investigation, analysis, and interpretation’. More recently, CDS scholars have drawn inspiration from science and technology studies (e.g. Latour 1993) by arguing that data comes constituted within wider assemblages and thus reflects pre-existing value structures (Thylstrup et al 2021; Kitchin 2014). Other scholars underscore that data practices are performative and must therefore entail ethical commitments arising from the socio-cultural contexts in which they are embedded (Johns and Compton 2022). The highly sensitive nature of asylum data thus makes it important to ask how research can work towards wielding data power back to those who are oftentimes termed the ‘data subjects’ – in this case asylum-seekers, refugees and migrants (d’Ignazio and Klein 2020; Rose 2005).

As Ghezelbash, Dorostkar and Walsh (forthcoming) argue based on related work on asylum adjudication in Australia, AI tools may be able to support the fundamental principle of equality of arms typically lacking in legal proceedings concerning asylum. Providing practicing lawyers with a posteriori predictive analysis of asylum decisions could help lawyers identify relevant precedence and tailor their arguments for specific decision-makers by indicating what kind of legal arguments have been historically successful in similar cases. Similarly, it may enable NGOs to identify and document outlier decisions or possible discrimination in regard to specific types of cases. If asylum-seekers and their advocates have access to such tools, they would be better equipped to appeal decisions on grounds of what is known in common law jurisdictions as ‘apprehended bias.’

Vice versa, computational research on asylum adjudication may be subject to push back from authorities and courts, as happened in France, which sought to introduce a general ban on research on individual judge behaviour following predictive analysis of asylum outcomes mentioned above (Langford and Madsen 2019). For governments and courts, computational legal research may be seen as devaluing the normative aspect of law by highlighting problematic aspects of legal decision making, and thereby undermine institutional trust (Chatziathanasiou 2022). Carefully grounding
computational methods within a practice-oriented legal discourse may thus important, if empirical research is to achieve real-life impact. Legal professionals have long argued that the notion that the outcome of a case depends on ‘what the judge had for breakfast’ (Priel 2020), does not reflect their experience. Empirical studies continue to show that law remains a strong influence on the behaviour of judges (Solum 2014), and this must be borne in mind when addressing the intended audience. Moreover, as Chatziathanasiou (2022) has recently argued, social science evidence must be carefully assessed within a legal framework, to the extent that which it addresses the causative factors of legal judgment.

CDS encounters each side of this challenge by arguing for involving groups of practitioners and affected parties directly into the data construction and analytical process to facilitate grounded sense making (see e.g. Neff and Tanweer 2017). A responsible ethics for computational refugee law may thus seek to integrate the perspectives of refugees themselves, legal practitioners and decision makers through participatory methods of research. In our own work, participatory methods have proved particularly valuable, not only as a means to ensure dialogue and a practice-oriented engagement for our research, but also in terms of probing how data and data categories are continuously produced by both authorities and non-state actors within RSD processes, and how this impacts decision-making processes (Møller et al 2019; Møller and Nielsen 2022).

In our work, close dialogue with NGOs and authorities have further been a prerequisite to ensuring data access and representation. The Danish dataset we have drawn from in the case study above was made possible by a special agreement with a non-governmental organisation, the Danish Refugee Council, which has a formal role in the Danish asylum procedure and has maintained an archive of asylum decisions over decades. For other countries, our work is reliant on access to information requests and the willingness of public authorities to open up repositories of case files as well as internal metadata. In each case, such agreements have been subject to stringent conditions, including for ethical use and sharing of the data due to requirements of anonymization and data storage. Working with NGOs as data partners has further been a strategy to ensure a modicum of representation, since it would be impossible to ensure individual consent from individual asylum-seekers as ‘data subjects.’

We furthermore acknowledge that data access in this area is likely to be highly uneven and itself risks reproducing existing biases in refugee research. Some countries make decisions on asylum applications publicly available. Yet, many countries face almost a blanket ban on publication. Furthermore, structural differences in regard to e.g. digitalisation of case files means that
computational research may not be possible in jurisdictions where this is not generally practiced as part of the legal system. Issues concerning data availability and research resources thus mean that computational refugee law research has so far been dominated by studies of asylum systems in the Global North, thereby reproducing broader structural imbalances in refugee law research (Spijkerboer 2021). More generally, computational research may similarly entrench issues of ‘methodological nationalism’ by essentializing units of analysis due to the inherent dependency of researchers on certain data sources (Wimmer and Schiller 2003; Drouhot et al 2022).

A related problem concerns missing data. Case files are often archived on a historical basis, meaning that older files will often be in a different format and vary in terms of the information contained and structure through which it is presented. This can create issues for coding through computational methods as the text of these files is sometimes not in a machine-readable format, or case files lost or corrupted due to archival transfers across storage platforms – something we experienced in our Danish dataset as well. This problem is further compounded for documents that are not in the English language (Baden et al., 2021). Whilst rendering programs are increasingly becoming more sophisticated across languages, digital delays of this kind is particularly likely to effect the empirical scope of early applications of computational methods in this area, and thereby the generalizability and rigor of such work. We highlight these problems to underscore that computational legal research should never purport to be ‘complete’ – like any form of empirical analysis; it is at best a snapshot through a certain prism.

Last, but not least, we take from CDS and other schools of critical theory the need to actively ‘problematize’ intra-disciplinary perspectives and constructions in order to reflexively engage disciplinary biases in the design of computational legal methods (Petersen et al. 2021). A point similarly recognised in wider debates on interdisciplinary engagements between law and the social sciences (Klabbers 2009; Noll 2005: 6). In our own practice, differences in terms of epistemological and normative assumptions similarly emerge as part the ongoing tripartite dialogue between legal, computer science and critical data studies scholars. For example, the term bias has a specific technical term in data science. Bias from this perspective may emerge already in the construction of data and what is counted or left out, which is substantially different from a legal understanding of bias as the predisposition of a judge or adjudicator to decide against or in favor of one of the parties, or in our case, the asylum-seeker. Similarly, taking a reflexive approach to our own process of data analysis has led to important discussions on data categorisations, deficits, exclusion, equity and risks (Johns 2017; D’Ignazio and Klein 2020). For example, in the above analysis a number of binary distinctions
are applied in relation to e.g. sex and marital status, which may themselves neglect the complexity of gender categories in the context of asylum-seekers. In a follow-up study, we thus specifically sought to unpack how gender categories emerge in the applicants’ presentation of asylum motives and subsequent case handling (Kaltenhäuser et al. 2022). Data science tools stabilise and provide consistency to our analysis. However, working from different epistemologies and disciplines, we believe reflexive strategies are essential for ensuring a balance between, on the one hand, technical explorations of the affordances that data driven approaches make possible, and maintaining an awareness of and an commitment to challenge core concepts such as neutrality and objectivity of data, on the other (D’Ignazio 2021).

Conclusion

As a field of practice law has its own truth conditions and mode of existence (Matthews and Veitch 2016). Consequently, computational studies of refugee law that engage in policy must be pursued cautiously. The significant benefit of a computational refugee law scholarship lies in its ability to raise the stakes of the game by enhancing our ability as researchers to analyse legal texts with greater scale and depth, but at the same time does not replace the need for finely grained qualitative research. Computational approaches can reveal undiscovered synergies and patterns in the law, but should always be connected to a strategy of what forms the ultimate purpose of legal research.

A computational refugee law must therefore also be cognisant of the findings of generations of socio-legal empirical research into the processes of refugee law. Research which has shown refugee status determination to be a highly contextual phenomenon to varying degrees shaped socio-institutional setting, external factors and the normative context of asylum judging that creates law from narratives and uncertain evidence. This explains, in part, the backlash that has already been experienced by researchers from judges who are reluctant to reduce difficult decisions to quantification. However, it also further underscores the importance of a reflexive approach that engages these concerns through dialogue, and with awareness of the ethics that are entailed by encroaching datafication – whether led by researchers or authorities themselves.

In this article, we have sought to show what computational approaches can bring to the table for refugee law research, whether as a means to help unpack recalcitrant research problems within the field, or as critical responses to growing digitalisation and application of algorithmic decision-making.
in practice. As an example, our analysis of asylum decision making at the Refugee Appeals Board highlights how data-driven methods can bring greater scale to the analysis of in-country variations and the ability to automatically extract core features about applicants and adjudicators. Through mathematical methods, we have further sought to lay the foundations for cross-jurisdictional comparisons based on previous studies in Australia, Canada and the United States. At the same time, it is important to emphasise that the examples we have provides represents only early steps. More generally, and save for a few pioneering studies, the application of data-driven approaches in refugee law is still in its infancy. But it is an area which be expected to grow in line with broader transdisciplinary frameworks for computational research in law and the social sciences.

Bibliography


Displaced Persons and Migrant Workers Essays in Memory of Joan Fitzpatrick and Arthur Helton. Brill, pp. 65-89.


Author(s): William Hamilton Byrne, Thomas Gammeltoft-Hansen, Sebastiano Piccolo, Naja Holten Møller, Tijs Slaats, Panagiota Katsikouli

Title: Data Driven Futures of International Refugee Law

MOBILE Working Paper, No. 1, 2023

Publication date: 01/01/2023

© Author

MOBILE Working Paper Series

William Hamilton Byrne, Assistant Professor, University of Copenhagen

E-mail: William.hamilton.byrne@jur.ku.dk

MOBILE (Center of Excellence for Global Mobility Law) was established in 2023 based on a grant from the Danish National Research Foundation. The center’s mission is to systematically study the legal infrastructures of human mobility across geographies, social divides, travel patterns and time.

Papers are available at: https://mobilitylaw.ku.dk/working-papers/

MOBILE

Danish National Research Foundation’s Center of Excellence for Global Mobility Law
The Faculty of Law, University of Copenhagen
Karen Blixens Plads 16, 2300 Copenhagen S