



The Computational Turn in International Law

Runar Hilleren Lie | ORCID: 0000-0002-6799-4156

Department of Public and International Law, University of Oslo, Oslo,
Norway

Corresponding author

r.h.lie@jus.uio.no

Malcolm Langford | ORCID: 0000-0002-5975-4320

Department of Public and International Law, University of Oslo, Oslo, Norway
malcolm.langford@jus.uio.no

Received 5 February 2024 | Accepted 5 February 2024 |

Published online 26 March 2024

Abstract

The recent computational turn carries the potential to both enrich and destabilise international law research. Epistemologically, it offers a repertoire of data science methods, including network analysis, text processing and analysis, machine learning, and large language and agent-based models. Ontologically, with its inductive and macroscopic logic, it can inflect and challenge our established understandings of the nature and role of international law. In this article, we review the growing use of computational approaches in international law research, with a focus on their methodological particularity and potential contributions to doctrinal and non-doctrinal research questions. At the same time, we discuss the limitations of the computational turn, especially issues of suitability, quality, and feasibility.

Keywords

legal research – data science – international law – computational legal studies – adjudication – machine learning – network analysis – quantitative text analysis

1 Introduction

The rise of computational methods carries the potential to both enrich and destabilise research in international law. Like previous ‘turns’ – linguistic, critical, historical, and empirical – the growing computational turn is a project that is both epistemological and ontological.¹ Epistemologically, it offers a repertoire of data science methods for addressing doctrinal *and* non-doctrinal questions. Ontologically, the computational turn inflects and challenges established understandings of the nature and role of international law, with its macroscopic data lens and ability to draw attention to ‘larger networks of forces’² and ‘reveal patterns that researchers did not expect nor actively look for.’³

Because international law is a complex adaptive system, it is in many ways overly ripe for such a computational turn. Constituted by multiple elements (a web of 50,000-plus treaties,⁴ judgments, state practice and *opinio juris*, secondary literature, institutions, and actors), which are deeply interconnected and constantly changing, international law is a system that is challenging to understand.⁵ Herein lies the computational promise. It enables easy access to large and scattered volumes of legal and related text and reduces it automatically to numerical data. Moreover, it provides new tools to analyse

- 1 For a deeper discussion of the computational turn in law, see David Law and Malcolm Langford, ‘The Computational Turn in Constitutional Law’, in David Law and Malcolm Langford (eds.), *Research Methods in Constitutional Law: A Handbook* (Edward Elgar, forthcoming). For analysis of other ‘turns’, see L. Eslava, ‘The Teaching of (Another) International Law: Critical Realism and the Question of Agency and Structure’, 54 *The Law Teacher* (2020) pp. 368–384; M. Koskeniemi, *The Politics of International Law* (Hart Publishing, 2011); G. Shaffer and T. Ginsburg, ‘The Empirical Turn in International Legal Scholarship’, 106 *American Journal of International Law* (2012) pp. 1–46; T. Skouteris, ‘The Turn to History in International Law’, 16 *Journal of International Law* (2005) pp. 539–559.
- 2 S. Spyrou, ‘An Ontological Turn for Childhood Studies?’, 33 *Children & Society* (2019) pp. 316–323.
- 3 W. Alschner, *Investment Arbitration and State-Driven Reform: New Treaties, Old Outcomes* (Oxford University Press, New York, 2022) p. 24.
- 4 W. Alschner, ‘The Computational Analysis of International Law’, in R. Deplano and N. Tzagourias (eds.), *Research Methods in International Law* (Edward Elgar Publishing, 2021) p. 203.
- 5 J. F. Morin, J. Pauwelyn and J. Hollway, ‘The Trade Regime as a Complex Adaptive System: Exploration and Exploitation of Environmental Norms in Trade Agreements’ 20 *Journal of International Economic Law* (2017) pp. 365–390; Page, ‘What Sociologists Should Know About Complexity’, 41 *Annual Review of Sociology* (2015) pp. 21–41; A. Roberts and T. St John, ‘Complex Designers and Emergent Design: Reforming the Investment Treaty System’ 116 *American Journal of International Law* (2022), pp. 96–149; J. Ruhl, D. Katz and M. Bommarito, ‘Harnessing Legal Complexity’ *Science* (2017).

and represent this data, such as network analysis, text analysis, machine learning, and large language models. They provide tools which could make better sense of the systems' legal and social properties of self-organisation, stability, and intermittent volatility, and even its leverage points for reform;⁶ as well as support the empirical dimensions of the other turns in international law.

At the same time, it is important to be aware of the potential dark side of the computational turn. These approaches are contingent on large volumes of data and the careful training of models and algorithms, which can reduce their relevance for more bespoke and non-digitalised areas of international law. Likewise, there is a concern that computational approaches may embed different forms of bias and noise that are difficult to detect. There is also a deeper concern that it may privilege an understanding of international law that is predictive and inductive, rather than normative and deductive.

This article seeks to provide an overview of the ongoing computational turn within international law research. To be sure, scholars have surveyed the rise of computational approaches to law in general⁷ or particular fields within international law.⁸ Three studies also analyse international law scholarship as a whole. In 2017, Šadl and Palmer Oslen focused on the growing use of network and quantitative text analysis;⁹ in 2019, Altwicker reviewed a wide range of computational approaches with a particular focus on their implications;¹⁰ and in 2021, Alschner conducted a fairly comprehensive review, with a focus though on how doctrinal questions in international law could be addressed.¹¹ In this

6 D. Meadows, 'Leverage points: Places to Intervene in a System' *The Sustainability Institute* (1999), p. 28.

7 J. Frankenreiter and M. A. Livermore, 'Computational Methods in Legal Analysis' 16 *Annual Review of Law and Social Science* (2020) p. 39–57; N. Varsava, 'Computational Legal Studies, Digital Humanities, and Textual Analysis' in R. Whalen, *Computational Legal Studies* (Edward Elgar Publishing 2020); R. Whalen, 'The Emergence of Computational Legal Studies: An Introduction' in R. Whalen (ed.), *Computational Legal Studies* (Edward Elgar Publishing 2020); R. Whalen, 'Defining Legal Technology and Its Implications' 30 *International Journal of Law and Information Technology* (2022) p. 47.

8 W. Alschner, J. Pauwelyn and S. Puig, 'The Data-Driven Future of International Economic Law' 20 *Journal of International Economic Law* (2017), p. 217; W. Hamilton Byrne et al. 'Data-Driven Futures of International Refugee Law' *Journal of Refugee Studies* (2023); T. R. Nielsen and N. H. Møller, 'Data as a Lens for Understanding What Constitutes Credibility in Asylum Decision-Making' 6:1 *Proceedings of the ACM on Human-Computer Interaction* (2022).

9 U. Šadl and H. Palmer Olsen, 'Can Quantitative Methods Complement Doctrinal Legal Studies? Using Citation Network and Corpus Linguistic Analysis to Understand International Courts', 30 *Leiden Journal of International Law* (2017) pp. 327–349.

10 T. Altwicker, 'International Legal Scholarship and the Challenge of Digitalization', 18 *Chinese Journal of International Law* (2019) pp. 217–246.

11 Alschner, *supra* note 3.

review we build on and develop this work. We include a host of new studies, focus on both doctrinal and non-doctrinal uses, and examine the potential of new large language models (LLMs) and generative artificial intelligence (AI).

In our survey, we ask five questions: How have computational approaches impacted international law research? What are the key methods and how are they applied? For which research questions are they particularly suitable? What are the limitations and how might they be overcome? And what are the likely developments?

Two important caveats must be made though. First, like other reviews, we do not attempt a ‘systematic’ review. Ours can be best described as ‘*mixed methods*’ – combining an ‘overview’ with a ‘scoping’ and ‘state of the art’ review. Thus, we describe the ‘characteristics’ of the literature, identify the ‘nature and extent’ of existing research, and focus more on ‘current matters’.¹² In so doing, we claim neither comprehensiveness nor representativity, but we do review a substantial portion of the computational international law scholarship in peer-reviewed journals and academic books; and some literature in non-peer reviewed computer science repositories for rapid publishing, such as ArchivX.¹³ Together, we cover 118 publications.

Second, our focus is on *scholarship*. The computational turn in international law is multidimensional. It is also present in the development of *legal information systems* (e.g., online treaty repositories and case databases)¹⁴ and *computational analysis in practice* (e.g., data science tools for human rights monitoring, diplomatic negotiation and adjudication, and prediction of violations or behaviour).¹⁵ We confine ourselves, nonetheless, to research,

12 M. J. Grant and A. Booth, 'A Typology of Reviews: An Analysis of 14 Review Types and Associated Methodologies,' 26 *Health Information and Libraries Journal* (2009) pp. 91–108.

13 We note that here there are 515 articles with ‘legal in the title’ and 28 that address
‘international law’ or the European Court of Human Rights.

14 D. Behn et al., PluriCourts Investment Treaty and Arbitration Database (PITAD), 31 January 2019; D. Cubie, 'Sources of IDL: The New IFRC Disaster Law Database', 3 *Yearbook of International Disaster Law Online* (2022) pp. 637–641.; See, e.g., A. Newcombe, *ITAlaw.Com*, available at <https://www.italaw.com> (visited 15 February 2021); United Nations Conference on Trade and Development, *International Investment Agreements Navigator*, 27 April 2021, UNCTAD Investment Policy Hub, available at <https://investmentpolicy.unctad.org/international-investment-agreements> (visited 27 April 2021); *HUDOC – European Court of Human Rights*, available at <https://hudoc.echr.coe.int/#%22documentcollectionid%22:%22GRANDCHAMBER%22,%22CHAMBER%22%7D> (visited 30 January 2024).

15 A. Deeks, 'Introduction to the Symposium: How Will Artificial Intelligence Affect International Law?', 114 *AJIL Unbound* (2020) pp. 138–140; *Building Capacity in Colombia: Truth and Reconciliation*, HRDAG – Human Rights Data Analysis Group, available at <http://hrdag.org/2021/10/28/colombia-truth-reconciliation/> (visited 30 January 2024).

while noting that computational legal research interacts with, draws on, and contributes to information systems and practice tools.¹⁶

The article proceeds as follows. In section 2, we discuss what is meant by computational approaches to legal research. Section 3 analyses some of the key methods, which is followed by an exploration of the potential of LLMs. Following this, section 4 outlines three key challenges (suitability, quality, and feasibility) and how they might be addressed. Finally, in section 5, we assess the extent of the computational turn and present some perspectives on its future.

2 Defining Computational Approaches to Legal Research

Before examining the details of different approaches, it is worth considering what the computational study of law entails. Put simply, it can be understood as the study of a legal system using largely empirical methods aided by computer software to advance knowledge about law, including its meaning, practice, effects, or culture. Yet, given that computational studies of law encompass a broad range of *techniques* for the extraction, processing, and analysis of information for a diverse range of research *goals*,¹⁷ scholars vary considerably in how they label and describe the field.

Some researchers focus on its *epistemological* features. In arguing for the use of the term ‘computational legal studies’, Whalen presents the different components that form the field. This includes text analysis, data acquisition, visualisation, artificial intelligence, methods, and model creation.¹⁸ By doing so, he seeks to indicate the field’s breadth rather than suggesting a single integrated research method or paradigm. In a somewhat similar vein, Katz’ describes computational legal studies as a project with broad interdisciplinary features: ‘a sub-field dedicated to applying tools from computer science, applied physics, informatics, complex systems and applied mathematics to help enrich positive legal theory.’¹⁹ Likewise, Frankenreiter and Livermore focus

16 See, e.g.: W. Alschner, M. Elsig and R. Polanco, ‘Introducing the Electronic Database of Investment Treaties (EDIT): The Genesis of a New Database and Its Use’ 20 *World Trade Review* (2021), p. 73; A. Quemy, ‘European Court of Human Right Open Data Project’ arXiv:1810.03115 (2019); R. Lie, ‘A Computational Approach and Dataset for Analysing International Investment Treaties’ <<https://papers.ssrn.com/abstract=4474552>> accessed 9 June 2023.

17 Frankenreiter and Livermore, *supra* note 7; M. A. Livermore and D. N. Rockmore, *Law as Data: Computation, Text, & the Future of Legal Analysis* (The Santa Fe Institute Press, 2019); Whalen, *supra* note 7.

18 Whalen, *supra* note 7, p. 3.

19 Katz et al., ‘Reproduction of Hierarchy – A Social Network Analysis of the American Law Professoriate’, 61 *Journal of Legal Education* (2011) p. 79.

on how the tools can add to legal knowledge.²⁰ They use though a different terminology – ‘computational methods in legal analysis’ – and focus more on the methodological contribution rather than the specific methods, dividing studies into the categories of causal inference, prediction and classification, interpretation and description, as well as citational analysis.

Other descriptions are more *ontological* in their focus. Alschner and colleagues foreground the underlying logic in many of the new tools.²¹ Rather than emphasising the methods, they frame the field as a data-driven approach to research.²² For this, three factors must be in place: ‘data-first’ (that the research hypotheses are formed from the patterns observed in the data); comprehensiveness (all available data should be used and not just subsets); and induction (patterns should be identified computationally rather than from manually coding and counting).²³ As Alschner argues in his recent book,²⁴ this method involves a discovery process followed by theorisation:

In the method, one first explores variation in the data and only then tries to rationalize the findings from a theoretical perspective. Inductive, data-driven approaches can thereby reveal patterns that researchers did not expect nor actively look for.²⁵

Drawing these different definitions together, we can observe that the computational turn in law is depicted in two ways. The first is epistemological, with its emphasis on methods and its contribution to existing research questions. It tends to attract monikers like ‘computational legal studies’, ‘computational analysis’ or ‘macroscopic’, and represents, in many ways, the evolution and expansion of the empirical turn in legal studies. The other is ontological, with a focus of the new possibilities in understanding law. It is often accompanied by terms like ‘data-driven’, and emphasises the distinctive and often inductive characteristics of computational approaches.

In our view, these two aspects are sometimes, or often, different sides of the same coin. Many of the methods – especially unsupervised text analysis, exploratory network analysis, or predictive analytics – may not only provide

²⁰ Frankenreiter and Livermore, *supra* note 7.

²¹ W. Alschner and D. Charlotin, ‘Data Mining, Text Analytics, and Investor-State Arbitration’ in P. Ortolani et al. (eds.) *International Arbitration and Technology* (Kluwer Law International, 2022).

²² Alschner, Pauwelyn and Puig, *supra* note 8.

²³ *Ibid.*, pp. 221–223.

²⁴ Alschner, *supra* note 6.

²⁵ *Ibid.*, p. 24.

a methodological edge but transform how we see international law and its surrounding community. In any event, both conceptions illustrate helpfully the nature of the contribution of the computational turn and, as we shall see, its limitations.

3 Computational Techniques in International Law

In this section, we discuss some of the major computational methods that have been used to analyse international law and the types of research questions for which they are particularly suitable. To be sure, some computational methods are excluded. For instance, agent modelling, which seeks to simulate the actions and interactions of individuals or actors to understand the behaviour of a system, has been applied in some legal research,²⁶ but remains seminal.

3.1 *Network Analysis*

We begin with network analysis, or graph theorem as it is known in the field of mathematics.²⁷ It constitutes now both a theory and a method.²⁸ Theoretically, if not ontologically, network analysis emphasizes the importance that ties or edges (i.e., relations) between nodes (e.g., individuals, institutions) have for the transmission of resources and inflection of the behaviour of nodes. For example, a court (node) might cite (an edge) another court (a second node) to strengthen the legitimacy or authority of its finding. Methodologically, network analysis provides tools for measuring and visualizing the relationships between individuals and similar individuals.

Within the study of law, network analysis has been shown to be particularly suitable in at least two areas. First, it has been widely used to understand the relationships between different sources of law, whether in adjudication, diplomacy, or advocacy.²⁹ Such uses may be doctrinal in nature (e.g., identifying neglected legal sources, determining the weight of a source, or predicting which sources may be relevant); or non-doctrinal (e.g. examining legitimation

26 See, e.g., K. Troitzsch, 'Legislation, Regulatory Impact Assessment and Simulation' N. Lettieri, S. Faro (eds.), *Law and computational social science* (Napoli, ESI, 2013); A. Schwartz, 'An Agent-Based Model of Judicial Power' *Journal of Legal Metrics* 9 *Journal of Law* (2019) p. 21.

27 N. Biggs, *Graph theory: 1736–1936* (Clarendon Press, 1986).

28 S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, New York, 1994); B. Carolan, 'The social network perspective and educational research introduction' in *Social Network Analysis and Education: Theory, Methods and Applications* (Thousand Oaks, SAGE Publications, 2014) pp. 3–22.

29 Šadl and Olsen, *supra* note 9, pp. 333–4.

practices or the features of influential sources). Second, the methodology has been used to understand the relationships between the actors in a legal system (e.g., the openness of actors to new ideas, their influence on others, or their conflicting roles).³⁰

In both types of investigations, the analysis may be micro (case/individual level), meso (sub-structures of law or actors/institutions) or macro (a system, or systems). Questions posed at the micro and meso level often involve a focus on interconnections while those at the macro level concern overall characteristics of the network: 'how large it is, how dense it is, how does innovation travel the network etc'.³¹

These levels often influence which of the (many) available measures should be used. At the micro level, *centrality* is often deployed to measure a node's influence in the network (e.g. a judicial precedent); while *betweenness* shows the influential quality of certain connections (e.g., the presence of a certain judge, state, or context). At the meso level, *distribution* helps measure the degree of centralization in the influence in the network (e.g., whether the ICJ is a centripetal force in restraining fragmentation in international law);³² while *clustering* uses the links between nodes to create groups with common interconnections, e.g., sub-clusters of contradictory interpretations or distinct epistemic communities. In addition to these measures, there are a range of more complex algorithms and scores such as PageRank (weighting the importance of nodes), HITS (adds a relevance factor when weighing nodes), and Motifs (recurrently determines influence, clusters and distribution).³³

The field of legal network analysis in national and international law has developed to such an extent that it can be divided into four distinct network types: legal citation networks, legal social networks, statutory and regulatory networks, and criminology networks. We discuss the first two, which are the most prevalent in international law.³⁴

30 S. Puig, 'Social Capital in the Arbitration Market', 25:2 *European Journal of International Law* (2014) pp. 387–424.

31 Alschner, *supra* note 3, p. 208.

32 See discussion on fragmentation in *Ibid.*, p. 225.

33 S. Brin and L. Page, 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', 30 *Computer Networks and ISDN Systems* (1998) pp. 107–117; B. Carolan, 'The social network perspective and educational research introduction' in *Social Network Analysis and Education: Theory, Methods and Applications* (Thousand Oaks, SAGE Publications, 2014), pp. 3–22.

34 For examples of the latter two, see, e.g., K. Ashley et al., 'Statutory Network Analysis plus Information Retrieval', in *Second International Workshop on Network Analysis in Law* (2014) pp. 1–7; J. M. McGloin and D. S. Kirk, 'An Overview of Social Network Analysis', in T. C. Pratt (ed.) *Advancing Quantitative Methods in Criminology and Criminal Justice* (Routledge, New York, 2012); R. Whalen, 'Legal Networks: The Promises and Challenges of Legal Network Analysis', 2016 *Michigan State Law Review* (2016) pp. 539–566.

Legal citation analysis has a long pre-computational history. It can be conducted with small-N qualitative methods and large-N quantitative methods.³⁵ Focusing though on data science approaches, an early study used computational network analysis to study the judicial structure of the US Supreme Court from micro, meso, and macro perspectives,³⁶ showing a core of key cases that form the nexus of its two centuries-long jurisprudence. In a similar vein, Leicht and colleagues show the structural divisions, developments, and patterns within the Court's network of citations over time.³⁷ A wide range of authors have used since network analysis of different domestic courts' practice to analyse authority, complexity, hierarchy, influence, citation growth, and reliance on incorrect precedents.³⁸

Similar analysis has been conducted in the realm of international law. In 2012, Lupu and Voeten analysed the European Court of Human Rights' (ECtHR), finding that the court is 'case matter' focused rather than country focused in its use of previous jurisprudence.³⁹ Moreover, the Court cites more frequently when it anticipates scepticism from domestic judges.⁴⁰ Similarly, Alschner and Charlotin find that the ICJ cites increasingly its own judgements, even though not obliged; and, over time, it has created an ever more complex self-citation network.⁴¹ Scholars have used citation network analysis to study other courts and tribunals such as the International Criminal Court (ICC),⁴² the Court of Justice of the European Union (CJEU),⁴³ the World Trade Organisation

35 For an overview, see Smith, 'The Web of Law', 44 *San Diego Law Review* (2007) pp. 309–354.

36 S. J. Chandler, 'The Network Structure of Supreme Court Jurisprudence', 10 *The Mathematics Journal* (2005) pp. 501–526.

37 E. A. Leicht et al., 'Large-Scale Structure of Time Evolving Citation Networks', 59 *The European Physical Journal B* (2007) pp. 75–83.

38 Chandler, *supra* note 36; Katz et al., *supra* note 19; Katz and Stafford, 'Hustle and Flow: A Social Network Analysis of the American Federal Judiciary', 71 *Ohio State Law Journal* (2010) p. 53; Whalen, *supra* note 34.

39 Y. Lupu and E. Voeten, 'Precedent in International Courts: A Network Analysis of Case Citations by the European Court of Human Rights', 42 *British Journal of Political Science* (2012) pp. 413–439.

40 *Ibid.*

41 Alschner and Charlotin, 'The Growing Complexity of the International Court of Justice's Self-Citation Network', 29 *European Journal of International Law* (2018) pp. 83–112.

42 F. Tarissan and R. Nollez-Goldbach, 'The Network of the International Criminal Court Decisions as a Complex System', in A. Sanayei, I. Zelinka and O. E. Rössler (eds.), *ISCS 2013: Interdisciplinary Symposium on Complex Systems* (2014) pp. 255–264.

43 Derlén and Lindholm, 'Goodbye van Gend En Loos, Hello Bosman? Using Network Analysis to Measure the M. Derlén and J. Lindholm, 'Goodbye van Gend En Loos, Hello Bosman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments', 20 *European Law Journal* (2014) pp. 667–687; M. Derlén and J. Lindholm, 'Characteristics of Precedent: The Case Law of the European Court of Justice in Three

(WTO),⁴⁴ investment arbitration tribunals, and the Court of Arbitration for Sport,⁴⁵ while Ridi has examined cross-citational networks between courts and tribunals – covering over 200,000 judgments in the process.⁴⁶

What is notable about legal citation analysis is the diversity of questions it seeks to answer. In some cases, it is driven by doctrinal and legal reform concerns. For example, Alschner's comparison of citation practices within international investment arbitration demonstrated that some tribunals are very attentive as to the underlying treaty in citing a case precedent, while others were highly liberal in their case selection,⁴⁷ raising questions over the use of proper legal method and strengthening claims in the UNCITRAL WG III reform process that tribunals are more concerned with consistency than correctness.⁴⁸ Šadl and Olsen argue that network analysis can increase the validity, reliability, and transparency of doctrinal studies while maintaining

Dimensions', 16 *German Law Journal* (2015) pp. 1073–1098; M. Derlén and J. Lindholm, 'Peek-A-Boo, It's a Case Law System! Comparing the European Court of Justice and the United States Supreme Court from a Network Perspective', 18 *German Law Journal* (2017) pp. 647–686; M. Derlén and J. Lindholm, 'Is It Good Law? Network Analysis and the CJEU's Internal Market Jurisprudence', 20 *Journal of International Economic Law* (2017) pp. 257–277; O. Larsson et al., 'Speaking Law to Power: The Strategic Use of Precedent of the Court of Justice of the European Union', 50 *Comparative Political Studies* (2017) pp. 879–907; H. S. Lund and H. Palmer Olsen, 'The Making of iCourts', in H. P. Olsen and H. S. Lund (eds.), *The Making of iCourts* (2022) pp. 11–40, p. 225–246; Y. Panagis and U. Šadl, 'The Force of EU Case Law: A Multi-Dimensional Study of Case Citations', *Legal Knowledge and Information Systems* (2015) pp. 71–80; Šadl and Olsen, *supra* note 9.

44 K. Pelc, 'The Politics of Precedent in International Law: A Social Network Application', 108 *American Political Science Review* (2014) pp. 547–564.

45 W. Alschner, 'Ensuring Correctness or Promoting Consistency? Tracking Policy Priorities in Investment Arbitration through Large-Scale Citation Analysis', in D. Behn, M. Langford and O. K. Fauchald (eds.), *The Legitimacy of Investment Arbitration: Empirical Perspectives* (Cambridge University Press, 2022) pp. 230–255; J. Lindholm, *The Court of Arbitration for Sport and Its Jurisprudence: An Empirical Inquiry into Lex Sportiva* (Asser Press, Springer, The Hague, 2019).

46 D. Charlotin, 'The Place of Investment Awards and WTO Decisions in International Law: A Citation Analysis', 20 *Journal of International Economic Law* (2017) pp. 279–299; N. Ridi, 'The Shape and Structure of the 'Usable Past': An Empirical Analysis of the Use of Precedent in International Adjudication', 10 *Journal of International Dispute Settlement* (2019) pp. 200–247; N. Ridi, 'Approaches to External Precedent: The Invocation of International Jurisprudence in Investment Arbitration and WTO Dispute Settlement', in D. Behn, M. Langford and S. Gáspár-Szilágyi (eds.), *Adjudicating Trade and Investment Disputes: Convergence or Divergence?* (Cambridge University Press, 2020) 121–148.

47 Alschner, *supra* note 45.

48 See discussion in M. Langford et al., 'UNCITRAL and Investment Arbitration Reform: Matching Concerns and Solutions: An Introduction', 21:2–3 *The Journal of World Investment & Trade* (2020) p. 167.

integrity in ensuring sufficient attention to legal detail. They demonstrate this through a historical network analysis of the step-by-step judicial development of Article 14 of the European Convention of Human Rights, no longer visible in the current citation practice of the ECtHR.⁴⁹ At the same time, others use legal citation analysis to address socio-legal questions. Pelc shows how certain states conduct strategic litigation before the WTO panels to create precedents for future cases.⁵⁰ Notably, by tracking the eventual influence of a citation on the outcome of cases, he was able to calculate the commercial or economic value of citations.

Turning to *social network analysis*, it uses people and organisations for nodes, and their interactions as edges (e.g., tribunals, treaty signatories). An early use can be found in Katz and Stafford's⁵¹ analysis of how law-clerks move throughout the American federal judiciary, creating a proxy for the social influence of the various judges that these clerks served.⁵² It showed that the social structure and authority of the judiciary is dictated by a small group of central judges. Katz and colleagues⁵³ also analysed the network of academic references in hiring in law schools, finding that elite universities have an outsized influence.⁵⁴

In international law, the field of investment law has been the most prominent subject of social network analysis. Its voluminous and highly decentralised nature – with over 3500 treaties and 1000 ad hoc arbitration tribunals – makes it a good candidate. For example, Saban and colleagues created a social network of state signatories to 2460 treaties (until 2005),⁵⁵ showing both growth and decline.⁵⁶ Today, such network analysis of investment treaties is now an underlying part of more sophisticated machine learning methods that analyse the dynamic and linguistic patterns in treaty drafting (see section 3.3).

49 A. Frese and H. Palmer Olsen, 'Citing Case Law: A Comparative Study of Legal Textbooks on European Human Rights Law', 11:1 *European Journal of Legal Studies* (2018) p. 91–131; see also H. Palmer Olsen and A. Küçüksu, 'Finding Hidden Patterns in ECtHR's Case Law: On How Citation Network Analysis Can Improve Our Knowledge of ECtHR's Article 14 Practice', 17 *International Journal of Discrimination and the Law* (2017) pp. 4–22; Šadl and Olsen, *supra* note 9, pp. 344–347.

50 Pelc, *supra* note 44.

51 Katz and Stafford, *supra* note 38.

52 *Ibid.*, p. 484.

53 Katz et al., *supra* note 19.

54 *Ibid.*

55 D. Saban, F. Bonomo and N. E. Stier-Moses, 'Analysis and Models of Bilateral Investment Treaties Using a Social Networks Approach', 389 *Physica A: Statistical Mechanics and Its Applications* (2010) pp. 3661–3673.

56 *Ibid.*, pp. 3671–3672.

Another use is to examine individuals and their behaviour. In 2014, Puig examined the social relations between arbitrators for International Centre for Settlement of Investment Disputes (ICSID) cases.⁵⁷ By counting the number of roles and different types of ties between arbitrators, he measured arbitral power according to their arbitration panel exposure.¹⁵ Tracking these appointments between 1972 and 2014, he identified a core of 25 arbitrators whom he calls ‘power brokers’,¹⁶ mostly the ‘grand old men’ from Europe and North America.¹⁷ According to Puig, the resilience of this group may support Ginsburg’s theory that the imbalance in appointments is a result of strategic action, with insider arbitrators raising the barriers to keep out new entrants.

We expanded on this study in an article with Behn.⁵⁸ We included all investment arbitration cases (1039 at the time) as well as legal counsel, secretaries and expert witnesses – and weighted each actor’s role. We found that the identity of the ‘powerbrokers’ changed partly, and that those that served as both arbitrators and legal counsel moved up the list. This not only raised normative and legal concerns about the risks of ‘double hatting’,⁵⁹ it showed both the potential for strategic action by insiders (playing different roles) and how dual roles boosted an arbitrator’s symbolic capital, making them more attractive for appointment.⁶⁰ Lie also measured the influence of law firms in this network, finding that a central group of law firms held almost comparable influence as arbitrators.⁶¹

Summing up, when studying complex systems such as legal citations or treaty and actors dynamics, network analysis provides a useful tool for structuring and analysing the multiple intricate connections. It can help confirm presumed patterns or discover new patterns – revealing the structures of power and influence.

3.2 *Textual Analysis*

Beyond quantifying and extracting information, computational methods may provide insights into the legal texts themselves, as well as their

⁵⁷ Puig, *supra* note 30.

⁵⁸ M. Langford, D. Behn and R. Lie, ‘The Revolving Door in International Investment Arbitration’, 20 *Journal of International Economic Law* (2017) pp. 301–332.

⁵⁹ *Ibid.*; M. Langford, D. Behn and R. Lie, ‘The Ethics and Empirics of Double Hatting’, 6 *ESIL Reflections* (2017), available at < https://esil-sedi.eu/post_name-118/>.

⁶⁰ M. Langford, ‘The Changing Sociology of the Investment Arbitration Market: The Case of Double Hatting’, in C. Bull, L. Malintoppi and C. Partasides (eds.), *Arbitration’s Age of Enlightenment?* (Wolters Kluwer, 2023).

⁶¹ R. Lie, ‘The Influence of Law Firms in ISDS’, in O. K. Fauchald, D. Behn and M. Langford (eds.), *The Legitimacy of Investment Arbitration: Empirical Perspectives* (Cambridge University Press, 2022).

authors. In traditional quantitative analysis, legal texts are coded with pre-determined variables. This can provide significant insight into the workings of international law,⁶² but can lead to the loss of significant information.⁶³ Legal texts are a hermeneutic construction, with a wealth of information that is embedded in content, expression, and form.⁶⁴ Computational text analysis or natural language processing (NLP) can complement existing methods by providing tools that offer macroscopic insights into a broad range of qualities and attributes in texts.⁶⁵ It combines 'distant reading' with traditional 'close reading'.

Such text analysis in legal research becomes a useful tool at both the data collection and analysis phases of research, and is increasingly used in combination with network analysis and machine learning.⁶⁶ Text is the natural research object.⁶⁷ For data collection, computational text analysis methods

62 See, e.g., E. A. Posner and M. F. P. de Figueiredo, 'Is the International Court of Justice Biased?', 34 *Legal Studies* (2005), p. 599; L. Helfer and E. Voeten, 'International Courts as Agents of Legal Change: Evidence from LGBT Rights in Europe' (2014) 68:1 *International Organization*, pp. 77–110; J. K. Staton and W. H. Moore, 'Judicial Power in Domestic and International Politics', 65 *International Organization* (2011), p. 553; Ø. Stiansen, 'Directing Compliance? Remedial Approach and Compliance with European Court of Human Rights Judgments', 51 *British Journal of Political Science* (2021), p. 899; D. Behn, T. L. Berge and M. Langford, 'Poor States or Poor Governance: Explaining Outcomes in Investment Treaty Arbitration', 38(2) *Northwestern Journal of International Law and Business* (2018).

63 Alschner and Charlotin, *supra* note 21; Livermore and Rockmore, *supra* note 17.

64 M. Langford, D. Behn and R. Lie, 'Computational Stylometry: Predicting the Authorship of Investment Arbitration Awards', in R. Whalen (ed.), *Computational Legal Studies: The Promise and Challenge of Data-Driven Research* (Edward Elgar, 2020); R. Lie, 'The Influencers of International Investment Law: A Computational Study of ISDS Actors' Changing Behavior', 23 *German Law Journal* (2022), pp. 350–375; J. Pauwelyn and K. Pelc, *Who Writes the Rulings of the World Trade Organization? A Critical Assessment of the Role of the Secretariat in WTO Dispute Settlement*, SSRN Scholarly Paper, ID 3458872 (2019), available at <<https://papers.ssrn.com/abstract=3458872>> (visited 12 November 2019); R. Nanda et al., 'Unsupervised and Supervised Text Similarity Systems for Automated Identification of National Implementing Measures of European Directives' 27 *Artificial Intelligence and Law* (2019) pp. 199–225.

65 Alschner and Charlotin, *supra* note 21; K. Bodwin, J. S. Rosenthal and A. H. Yoon, 'A Statistical Approach to Judicial Authorship: A Case Study of Judge Easterbrook', 37 *Advances and Applications in Statistics* (2013) p. 123; W. H. Gomaa and A. A. Fahmy, A Survey of Text Similarity Approaches; Hachey and Grover, 'Extractive Summarisation of Legal Texts', 14 *Artificial Intelligence and Law* (2006) pp. 305–345; *Ibid.*; J. S. T. Howe et al., Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments, 2019; Langford, Behn and Lie, *supra* note 64.

66 Alschner, *supra* note 3, p. 211.

67 S. Graham et al., *Exploring Big Historical Data: The Historian's Macroscope* (2nd edition, World Scientific, 2022); F. Jannidis, H. Kohle and M. alte Rehbein (eds), *Digital Humanities* (JB Metzler 2017).

are especially useful when attempting to extract information or categorise large amounts of largely homogenous legal documents, such as court or panel decisions, legal briefs or treaties. For data analysis, the method allows an efficient analysis of a large volume of data in a short period of time, e.g., seeking to find similarities and differences or reveal discursive and affective traits.

In this sub-section, we discuss the two standard research designs in computational text analysis: (1) ‘unsupervised’ exploration of text and (2) ‘supervised’ interrogation of text to examine whether a particular pattern exists. The underlying measures used may be the same – for example, text similarity, closeness between words, collocation of words, or sentiment type (positive/negative); and both approaches may be engaged to answer a single research question. Nonetheless, the overall ambition with each approach is distinct.

Unsupervised text analysis can be understood as a process of ‘fingerprinting’ a text corpus – sorting, categorising, and comparing. A common approach is clustering, also known as topic modelling, which allows researchers to group similar texts.⁶⁸ It is built on a two-phase process in which the algorithm searches for words, sequences of words, or ‘bags of words’ which commonly appear together and then predicts the likelihood that certain words or groups of words form a common topic. These topics may then be labelled by a researcher on examining the prediction.

In national law, topic modelling has attracted a significant literature.⁶⁹ For example, David Law examined the ideological influences on constitutional drafting, finding that certain ‘hegemonic forces have left a profound impact on constitutional texts’, namely earlier and highly prevalent British, French, and Spanish colonial influences as well as socialism.⁷⁰ In a related paper, he finds two types of human rights ‘dialects’ in constitutions: the ‘universalist’ and ‘positive-rights’.⁷¹ Likewise, Carlsen and colleagues classified and analysed the US Federal Code,⁷² finding strong connections between the structure and

68 I. Titov and R. McDonald, ‘Modeling Online Reviews with Multi-Grain Topic Models’, in *Proceedings of the 17th International Conference on World Wide Web* (2008) pp. 111–120.

69 B. M. Stewart, ‘Computer-Assisted Text Analysis’, in M. Langford and D. Law (eds.), *Research Methods in Constitutional Law* (forthcoming).

70 D. S. Law, ‘Constitutional dialects: the language of transnational legal orders’ (2019) *Constitution-Making and Transnational Legal Order*, pp. 110–55.

71 D. S. Law, ‘The global language of human rights: a computational linguistic analysis’ 12:1 *The Law & Ethics of Human Rights* (2018) pp. 111–50.

72 K. Carlson et al., ‘A Multinetwork and Machine Learning Examination of Structure and Content in the United States Code’, 8 *Frontiers in Physics* (2021), available at <<https://www.frontiersin.org/articles/10.3389/fphy.2020.625241>> (visited 3 November 2023).

content of the law, amplified by a strong cross-referencing network between certain provisions.⁷³

In international law research, unsupervised approaches are still in their infancy. But some examples exist. Examining the ICJ's use of precedent, Alschner and Charlotin use topic modelling to distinguish between different usages of the same case by comparing the text of the citing paragraphs.⁷⁴ They find that: 'The number of issues covered by citations has increased markedly over time, as disputants increasingly look for precedent beyond the subject matter of their case to gain an upper hand in litigation.'⁷⁵ Panagis, Christensen, and Sadl use topic modelling to explore the shifts in the content of the case-law of the ECtHR and CFEU, finding that some themes vary while certain key topics remain consistent.⁷⁶ Such path dependency was also identified by Lie in his computational creation of a family history of clauses in international investment treaties, based on a hybrid of topic modelling and supervised machine learning. It showed that states appear to borrow and reproduce, rather than innovate, when drafting international investment agreements.⁷⁷

Turning to *supervised approaches*, research is guided by an interest in identifying the existence (or not) of a pattern. Like topic modelling, similarity comparisons are used, but in a different manner. The most frequently used practice is the calculation of Jaccard distance.⁷⁸ Alschner⁷⁹ combines this measure to study the historic development of investment treaties. Using a combination of meta-data obtained from the precursor of the EDIT database,

⁷³ *Ibid.*

⁷⁴ Alschner and Charlotin, *supra* note 41.

⁷⁵ *Ibid.* Section B.1.

⁷⁶ Y. Panagis, M. L. Christensen and U. Šadl, 'On Top of Topics: Leveraging Topic Modeling to Study the Dynamic Case-Law of International Courts', *Legal Knowledge and Information Systems* (2016) pp. 161–166.

⁷⁷ R. Lie, 'Treaty Influencers: A Computational Analysis of the Development of International Investment Law', 26 *Journal of International Economic Law* (2023) pp. 500–524.

⁷⁸ Gomaa and Fahmy, *supra* note 65; P. Jaccard, 'The Distribution of the Flora in the Alpine Zone.1', 11 *New Phytologist* (1912) pp. 37–50.

⁷⁹ W. Alschner, 'Locked in Language: Historical Sociology and the Path Dependency of Investment Treaty Design', M. Hirsch and A. Lang (eds), *Research Handbook on the Sociology of International Law* (Edward Elgar, 2018); W. Alschner, 'Sense and Similarity: Automating Legal Text Comparison', R. Whalen (ed.) *Computational Legal Studies* (Cheltenham, Edward Elgar Publishing, 2020) pp. 9–28; W. Alschner, J. Seiermann and R. Skougarevskiy, 'Text-as-Data Analysis of Preferential Trade Agreements: Mapping the PTA Landscape', 5 *UNCTAD Research Paper* (2017); W. Alschner and D. Skougarevskiy, *Consistency and Legal Innovation in the BIT Universe*, SSRN Scholarly Paper, ID 2595288 (2015), available at <<https://papers.ssrn.com/abstract=2595288>> (visited 7 June 2019); W. Alschner and D. Skougarevskiy, 'Mapping the Universe of International Investment Agreements', 19 *Journal of International Economic Law* (2016) pp. 561–588.

together with both manual and automatic calculations of textual similarity, he finds that path dependency offers a better explanation than rational design for how investment treaties evolve.⁸⁰ States have a stronger tendency to reproduce existing treaty structures and legal provision solutions, rather than optimise their treaties according to their own immediate interests.⁸¹

Chaisse et.al. uses text similarity, in the form of computational phrase matching, to find how much influence the participating countries had on the Asia-Pacific Regional Comprehensive Economic Partnership (RCEP).⁸² By comparing the final text of the agreement with the parties' previous treaties, they find that rather than being influenced by the economically powerful parties in the treaty, much of the influence comes from the ASEAN and TPP agreements.⁸³ Potts and Kjær examine instead the influence of the International Criminal Tribunal for Yugoslavia (ICTY) and claims that it had 'irreversibly changed the landscape of international humanitarian law'.⁸⁴ Tracing the 'colocation' of words in other documents, they found 'discursive remnants of the Tribunal's stated "achievements"' but also 'problematic language' that called partly into question the court's contribution.⁸⁵

Others use supervised approaches to understand the changes within a court's jurisprudence. Šadl and Palmer Olsen seek to understand when a court draws innovatively on judicial precedent. They do so by testing use of the word 'effectiveness' in a network of over 10,000 cases and 18,247 paragraphs, finding not only the expected 'classic cases' associated with the Court's dynamic development of EU law, but also very under-recognised cases in tax law which have had system-wide consequences.⁸⁶

Another use of supervised text approaches is the identification of the author of legal texts. While legal texts are issued formally on behalf of an institution – assembly, government, court, lawyer – their actual authors may not be immediately identifiable. The text may have been written collectively or by outside actors. Identifying the writers may reveal the ideology, power, or

80 Alschner, *supra* note 80.

81 *Ibid.*

82 J. Chaisse et al., 'Drafting Investment Law: Patterns of Influence in the Regional Comprehensive Economic Partnership (RCEP)', 25:1 *Journal of International Economic Law* (2022) pp. 110–128.

83 *Ibid.*

84 A. Potts and A. L. Kjær, 'Constructing Achievement in the International Criminal Tribunal for the Former Yugoslavia (ICTY): A Corpus-Based Critical Discourse Analysis', 29 *International Journal for the Semiotics of Law – Revue Internationale de Sémiotique Juridique* (2016) pp. 525–555.

85 *Ibid.*

86 Šadl and Olsen, *supra* note 9, pp. 337–340.

the diversity of the authors; or allow pursuit of applied legal questions such as whether a document has been lawfully made (see e.g., challenges to arbitral decisions on the basis that they were prepared by assistants).⁸⁷

Attempts to identify authorship identification is best known through the method of stylometry. Neurological studies show that preferences for function words develop late in children,⁸⁸ and individual style is captured in features such as word and sentence length, lexical frequencies, punctuation patterns, and rare words. In order to identify judicial authorship, Rosenthal and Yoon model tested the *distribution* of the observed count of 63 function words of US Supreme Court justices over time against an expected distribution.⁸⁹ They found significant variation *across* judgment periods for some judges, suggesting that judicial clerks (who were constantly changing) played a significant role in drafting for some judges. Likewise, Carlson, Livermore and Rockmore (2015) analysed 25,407 opinions of the Supreme Court, finding that there a 'style of the time',⁹⁰ but also that 'writing styles of individual Justices have become less consistent as clerks have taken on a greater role on the Court'.⁹¹

This method has also been applied to international tribunals. Using stylometric comparisons, Pauwelyn and Pelc found that the WTO secretariat has significant influence on the drafting of the Dispute Settlement Body's panel reports.⁹² This study was reproduced by Charlotin,⁹³ and he found also that secretaries and assistants have an influence on the final awards.⁹⁴ Such

87 D. Galagan, *The Challenge of the Yukos Award: An Award Written by Someone Else – a Violation of the Tribunal's Mandate?*, 27 February 2015, Kluwer Arbitration Blog, available at <<https://arbitrationblog.kluwerarbitration.com/2015/02/27/the-challenge-of-the-yukos-award-an-award-written-by-someone-else-a-violation-of-the-tribunals-mandate/>> (visited 3 November 2023); Langford, Behn and Lie, *supra* note 64.

88 J. W. King and M. Kutas, 'A Brain Potential Whose Latency Indexes the Length and Frequency of Words', 10 (1995); L. Menn and L. K. Obler, 'Agrammatic Aphasia: A Cross-Language Narrative', in L. Menn and L. K. Obler (eds.), *Agrammatic Aphasia: A Cross-Language Narrative Sourcebook* (1990) pp. 1369–89.

89 J. S. Rosenthal and A. H. Yoon, 'Judicial Ghostwriting: Authorship on the Supreme Court', 96 *Cornell Law Review* (2010) pp. 1307–1344.

90 K. Carlson, M. A. Livermore and D. Rockmore, 'A Quantitative Analysis of Writing Style on the U.S. Supreme Court', 93 *Washington University Law Review* (2015) pp. 1461–1510.

91 *Ibid.* p. 1461.

92 Pauwelyn and Pelc, *supra* note 64.

93 D. Charlotin, *Who Writes WTO Panel and AB Reports? A Tentative Stylometric Analysis*, 1 May 2018, Medium, available at <<https://medium.com/@damien.charlotin/who-writes-wto-panel-and-ab-reports-a-tentative-stylometric-analysis-565c18f0491d>> (visited 19 October 2018).

94 D. Charlotin, 'Identifying the Voices of Unseen Actors in Investor-State Dispute Settlement', in F. Baetens (ed.), *Legitimacy of Unseen Actors in International Adjudication* (Oxford University Press, 2019) pp. 392–426.

text analysis of authorship has been steadily complemented though machine learning methods, to which we now turn.

3.3 *Machine Learning and Prediction*

The idea of using AI for answering legal questions goes back to the late 1970s. Initially a theoretical exercise,⁹⁵ it was followed in the 1980s by rule-based programming, use of legal databases, and development of tools for statistical analysis.⁹⁶ Yet, while this advanced the project of computer-based analysis of legal issues, it was not until the maturation of machine learning and NLP technologies in the 2010s that the idea of AI began to play a more significant legal role.⁹⁷

Machine learning emulates the way humans learn by allowing computers to learn by example and create recognition patterns that can be reused. Compared to text- and networking analysis, the scope of application with machine learning is broader and better equipped for tackling complexity. Researchers can predict outcomes, answer legal questions, find connections and discover patterns in the legal materials with increasingly impressive results. However, it often requires large amounts of documents and careful annotation for training models.⁹⁸

Predicting court dynamics is a common research agenda in national and international law.⁹⁹ An early successful attempt was conducted by Katz, Bommarito, and Blackman,¹⁰⁰ using hand coded data.¹⁰¹ With a random forest technique, they were able to predict the outcomes of US Supreme Court decisions with close to 80 per cent accuracy, as well as the votes of individual judges, bettering chance, experts, and expert panels.¹⁰² Most prediction studies

95 A. von der L. Gardner, *An Artificial Intelligence Approach to Legal Reasoning* (MIT Press, 1980).

96 K. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (Cambridge University Press, 2017).

97 K. D. Ashley, 'A Brief History of the Changing Roles of Case Prediction in AI and Law', 36 *Law in Context. A Socio-Legal Journal* (2019) pp. 93–112; Frankenreiter and Livermore, *supra* note 7.

98 See eg. D. M. Katz, M. J. Bommarito and J. Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States', 12 *PLOS ONE* (2017); Lie, *supra* note 78; M. Medvedeva, M. Vols and M. Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights', 28 *Artificial Intelligence and Law* (2020) pp. 237–266; Nanda et al., *supra* note 64.

99 For a broad review of predictive works see, M. Medvedeva, M. Wieling and M. Vols, 'Rethinking the Field of Automatic Prediction of Court Decisions', *Artificial Intelligence and Law* (2022), available at <10.1007/s10506-021-09306-3> (visited 23 February 2022).

100 Katz, Bommarito and Blackman, *supra* note 99.

101 Daily, 'The Supreme Court Database Codebook', p. 128.

102 Katz, Bommarito and Blackman, *supra* note 99.

on courts now use machine learning on raw text corpora with NLP. This includes outcomes,¹⁰³ judicial positions¹⁰⁴ and authorship,¹⁰⁵ which lawyers are likely to win, relevant cases, areas of law, and influence of public opinion.¹⁰⁶

For international courts, the use of predictive learning has focused on judicial outcomes, case features such as area of law, and adjudicative authorship. In relation to outcomes, there are now more than ten studies seeking to predict the findings of the ECtHR.¹⁰⁷ The first by Alteras and colleagues received widespread attention.¹⁰⁸ Basing their predictions solely on the text of previous judgments, they could predict, with an 79% accuracy, whether the Court would find a violation or not for particular ECHR articles. However, the approach was criticised by Medvedeva, Vols and Weiling for lacking sufficient inference control.¹⁰⁹ Yet, even by narrowing the training material to the facts

¹⁰³ K. Kowsrihawatt, P. Vateekul and P. Boonkwan, 'Predicting Judicial Decisions of Criminal Cases from Thai Supreme Court Using Bi-Directional GRU with Attention Mechanism', *2018 5th Asian Conference on Defense Technology (ACDT)* (2018) pp. 50–55.; A. Lage-Freitas et al., 'Predicting Brazilian Court Decisions', *arXiv:1905.10348 [Cs]* (2019), available at <<http://arxiv.org/abs/1905.10348>> (visited 21 February 2022); M. B. L. Virtucio et al., 'Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning', in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)* vol. 02 (2018) pp. 130–135; B. Walzl et al., 'Predicting the Outcome of Appeal Decisions in Germany's Tax Law', in *9th IFIP WG 8.5 International Conference, ePart 2017, St* (2017).

¹⁰⁴ Ashley, *supra* note 97.

¹⁰⁵ Bodwin, Rosenthal and Yoon, *supra* note 65; Rosenthal and Yoon, *supra* note 90.

¹⁰⁶ A. von Ungern-Sternberg, 'Demokratische Meinungsbildung Und Künstliche Intelligenz (Democracy, Public Opinion Formation, and Artificial Intelligence)', (2019), available at <<https://papers.ssrn.com/abstract=3400756>> (visited 2 March 2022).

¹⁰⁷ This includes N. Aletras et al., 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective', *2 PeerJ Computer Science* (2016) p. 93; I. Chalkidis, I. Androutsopoulos and N. Aletras, 'Neural Legal Judgment Prediction in English', *arXiv:1906.02059 [Cs]* (2019), available at <<http://arxiv.org/abs/1906.02059>> (visited 21 February 2022); A. Kaur and B. Bozic, *Convolutional Neural Network-Based Automatic Prediction of Judgments of the European Court of Human Rights* (2019); Medvedeva, Vols and Wieling, *supra* note 99; C. O'Sullivan and J. Beel, 'Predicting the Outcome of Judicial Decisions Made by the European Court of Human Rights', in *27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science* (2019); T. Y. S. S. Santosh et al., 'Deconfounding Legal Judgment Prediction for European Court of Human Rights Cases Towards Better Alignment with Experts', *arXiv:2210.13836* (25 October 2022); Leveraging Task Dependency and Contrastive Learning for Case Outcome Classification on European Court of Human Rights Cases, *arXiv:2302.00768* (13 February 2023); T. Y. S. S. Santosh, O. Ichim and M. Grabmair, 'Zero-Shot Transfer of Article-Aware Legal Outcome Classification for European Court of Human Rights Cases', *arXiv:2302.00609*, 13 February 2023.

¹⁰⁸ Aletras et al., *supra* note 108.

¹⁰⁹ Medvedeva, Vols and Wieling, *supra* note 99.

of the case, Medvedeva and colleagues were able to achieve similar results. To address the concern that the facts of the case may be written in a way that provides signals about a case, they also trained a separate model on the initial communication of the complaint. It produced similar results. In the wake of this research, various machine learning methodologies have been applied across multiple international jurisdictions, with fairly accurate results: in the range of 70–90%.¹¹⁰

A second field is predicting the likely area or provisions of law in litigation from the facts of a case, which has many potential practical applications. For example, Sulea and colleagues predict the area of law from the case description of 126,865 judgments of the French Supreme Court over two centuries with 90.9% accuracy.¹¹¹ Chalkidis, Androutsopoulos and Aletras attempted to forecast which articles in the ECHR and its protocols were violated in final judgments¹¹² although the best performing method for prediction of the violated *article* results was 60% (with BERT hierarchical version). The final and most ambitious attempt so far is by Quemy¹¹³ who attempted a multilabel and multiclass prediction of the relevant articles in the ECHR,¹¹⁴ achieving 79.66% accuracy in multiclass prediction of which articles (i.e., plural) are violated.

Beyond judgments, machine learning has been used to predict treaty outcomes. Alschner and Skougarevskiy developed a model that was able to write investment treaties, incorporating a majority of the features found in investment agreements while adjusting the predicted content to the negotiating parties.¹¹⁵ Similarly, we were able to predict the clause-level outcomes of investment treaty negotiations between states with over 90%

¹¹⁰ Chalkidis, Androutsopoulos and Aletras, *supra* note 108; Kaur and Bozic, *supra* note 108; Kowsrihawatt, Vateekul and Boonkwan, *supra* note 104; O'Sullivan and Beel, *supra* note 108; Quemy, *supra* note 16; Virtucio et al., *supra* note 104; Walzl et al., *supra* note 104; H. Zhong et al., 'Iteratively Questioning and Answering for Interpretable Legal Judgment Prediction', 34 *Proceedings of the AAAI Conference on Artificial Intelligence* (2020) pp. 1250–1257.

¹¹¹ O. M. Sulea et al., 'Predicting the Law Area and Decisions of French Supreme Court Cases', *arXiv:1708.01681 [Cs]* (2017), available at <<http://arxiv.org/abs/1708.01681>> (visited 23 February 2022).

¹¹² Chalkidis, Androutsopoulos and Aletras, *supra* note 108.

¹¹³ Quemy, *supra* note 16.

¹¹⁴ Including AdaBoost, Bagging Classifier, Bernoulli Naïve Bayes, Decision Tree, Ensemble Extra Tree, Extra Tree, Gradient Boosting, Linear SVC, Multinomial Naïve Bayes, Neural Net, and Random Forest.

¹¹⁵ W. Alschner and D. Skougarevskiy, *Can Robots Write Treaties? Using Recurrent Neural Networks to Draft International Investment Agreements*, SSRN Scholarly Paper, ID 2984935 (2016), available at <<https://papers.ssrn.com/abstract=2984935>> (visited 14 October 2020).

accuracy. We used a supervised learning model trained on previous negotiation outcomes and textual preferences across a corpus of 3000+ treaties, together with information on the economic and military power of each state and their historical preferences for certain types of provisions.¹¹⁶ Using the same dataset, Lie also used machine learning and network analysis to determine that certain states language have an outsized influence on these treaty negotiations.¹¹⁷

Finally, scholars have sought to complement comparative stylometric text analysis of authorship with machine learning methods. An early domestic attempt concerned the long debate over the identity of the authors behind the Federalist Papers, which had considerable influence on the drafting of the US Constitution.¹¹⁸ Using a neural network trained on the writing styles of known potential authors, Tweedie, Singh and Holmes developed a model that predicted the likely authors of each paper.¹¹⁹ Likewise, Rosenthal and Yoon use machine learning-like methods to test the robustness of their findings on the internal variation in judicial writing styles.¹²⁰ They partitioned the data into a training set and a testing set. Using a linear classifier based on function words, they found that they could distinguish fairly strongly between judges even when their v-scores for function word distribution were similar.

In international law, Pauwelyn and Pelc also used machine learning in their study on the WTO secretariat's role in the drafting of WTO panel reports.¹²¹ With Stylo, an R-package that analyses multiple features of the text, including word selection, frequency and relative placement, they conducted an exclusionary analysis. It showed that the stylometric influence of the secretariat was statistically significant, and may even outweigh the influence of the panellists. Likewise, in a study conducted by Behn and ourselves,¹²² and using a machine learning based approach, we found that there were large inconsistencies between the writing styles for documents issued by a presiding arbitrator in investment law. We suggested that this lack of linguistic consistency may indicate that some arbitrators, 'sub-contract', some of their drafting of awards.¹²³

Finally, some methods used to analyse domestic courts are yet to make their international debut. In a study on oral argument at the US Supreme Court,

116 Lie and Langford, 'Fait Accompli – Predicting the Outcomes of Investment Treaty Negotiations', (2023) presented at the *Conference on Data Science and Law*.

117 Lie, *supra* note 78; Quemy, *supra* note 16.

118 F. J. Tweedie, S. Singh and D. I. Holmes, 'Neural Network Applications in Stylometry: The 'Federalist Papers'', 30 *Computers and the Humanities* (1996) pp. 1–10.

119 *Ibid.*, at 6–9.

120 Rosenthal and Yoon, *supra* note 90.

121 Pauwelyn and Pelc, *supra* note 64.

122 Langford, Behn and Lie, *supra* note 64.

123 *Ibid.*

Dickinson, built predictive models from the transcript texts for outcomes, focusing on the features of number, length and timing of judicial questions and their sentiment.¹²⁴ The analysis was combined to create models for each justice, which was then trained against a set of ten samples with different feature analyses, and used subsequently to predict the voting patterns of justices with an overall accuracy of 73 per cent. Others have shown that certain vocal characteristics of advocates can influence a court outcomes!¹²⁵ This shows the potential for predictive models to capture the many legal and non-legal factors behind judicial behaviour¹²⁶

3.4 *Large Language Models – unleashed Potential*

While the tools and techniques discussed above have augmented and expanded the scope of research on international law, and will continue to do so, the greatest long-term impact on research may come from the emergence of LLMs, or as some of them are more popularly known – ChatGPT, Llama or Bard.¹²⁷

These LLMs are different from traditional language models, which need to be adapted and trained for a specific purpose, LLMs are general tools¹²⁸ that can be used to address a wide range of tasks. The potential increases in capability and accuracy that LLMs bring to the table are significant despite

124 G. M. Dickinson, 'A Computational Analysis of Oral Argument in the Supreme Court', 28 *Cornell Journal of Law and Public Policy* (2019), available at <<https://papers.ssrn.com/abstract=3198401>> (visited 8 November 2023).

125 D. Chen, Y. Halberstam and A. C. L. Yu, 'Perceived Masculinity Predicts U.S. Supreme Court Outcomes', 11:1 *PLoS ONE* (2016).

126 L. Baum, *The Puzzle of Judicial Behavior* (1997), available at <<https://www.fulcrum.org/concern/monographs/p5547s24n>> (visited 31 January 2024).

127 D. M. Katz et al., 'GPT-4 Passes the Bar Exam', 4389233, 15 March 2023; Medvedeva, Vols and Wieling, *supra* note 99; J. Savelka et al., 'Explaining Legal Concepts with Augmented Large Language Models' (GPT-4), arXiv:2306.09525, 22 June 2023; S. Surana, S. Dembla and P. Bihani, 'Identifying Contradictions in the Legal Proceedings Using Natural Language Models', 3 *SN Computer Science* (2022) p. 187.; A. Tamkin et al., 'Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models', arXiv:2102.02503, 4 February 2021; H. Westermann et al., 'Bridging the Gap: Mapping Layperson Narratives to Legal Issues with Language Models' *Proceedings of the Sixth Workshop on Automated Semantic Analysis of Information in Legal Text*, 23 June 2023; F. Yu, L. Quartey and F. Schilder, 'Legal Prompting: Teaching a Language Model to Think Like a Lawyer', arXiv:2212.01326, 8 December 2022.

128 Within AI research, there is a distinction between Artificial General Intelligence (AGI) which refers to human-like behavior and generalized AI, which refers to AI that can master a wide range of tasks, but without reaching any form of what we would consider intelligent. LLMs in their current state are considered to be in the latter category, as they are useful for a multitude of tasks, but has no built in "intelligence".

being only recently available. They have been used to pass the legal bar exam in the United States,¹²⁹ predict judgements¹³⁰, and provide models that can write, annotate and reason over legal texts and problems.¹³¹ In international law, one scholar has tested successfully tested ChatGPTs ability to engage in treaty interpretation, by seeing if it strategically deploy ‘cards’, usually principles of treaty interpretation, to persuade an ‘audience’ that their interpretation is correct.

In our view there are two areas – beyond the typical ask-a-question to the LLM and get-an-answer – that are most promising for international law research.

The first is the democratisation of computational methods. Until now, a broad and deep understanding of programming and sophisticated software tools has been required. Yet, LLMs with suitable frameworks around them able to conduct sophisticated analysis based on instructions in plain language, prompting a focus on the development of simpler ‘prompt engineering’ skills for lawyers.¹³² As an example, we can imagine an empirical study of lawyers in international law. While extracting the names, firms, and associations previously would require complicated software rummaging through each document trying to isolate the names and roles through various complex algorithms, LLMs instead allow us to load in the document, and simply ask the question: “*Within the document provided, please create a table with the names and affiliations of each lawyer mentioned.*” The output would be a nicely formatted table with each lawyer and affiliation. With a few basic lines of code generated by the LLM itself, this could be repeated for every case document

¹²⁹ J. H. Choi et al., ‘ChatGPT Goes to Law School’, 4335905, 23 January 2023.

¹³⁰ C. Jiang and X. Yang, ‘Legal Syllogism Prompting: Teaching Large Language Models for Legal Judgment Prediction’, in *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law* (2023) pp. 417–421.

¹³¹ T. Kojima et al., ‘Large Language Models Are Zero-Shot Reasoners’, 35 *Advances in Neural Information Processing Systems* (2022) pp. 22199–22213; Savelka and Ashley, ‘The Unreasonable Effectiveness of Large Language Models in Zero-Shot Semantic Annotation of Legal Texts’, 6 *Frontiers in Artificial Intelligence* (2023) p. 1279794; S. Yue et al., DISC-LawLLM: Fine-Tuning Large Language Models for Intelligent Legal Services, arXiv:2309.11325, 23 September 2023.

¹³² H. Hassani and E. S. Silva, ‘The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field’, 7 *Big Data and Cognitive Computing* (2023) p. 62; E. Kasneci et al., ‘ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education’, 103 *Learning and Individual Differences* (2023) p. 102274; J. G. Meyer et al., ‘ChatGPT and Large Language Models in Academia: Opportunities and Challenges’, 16 *BioData Mining* (2023) p. 20; R. Pudari and N. A. Ernst, ‘From Copilot to Pilot: Towards AI Supported Software Development’, arXiv:2303.04142, 7 March 2023.

available to the researcher. More sophisticated prompts could further be used to determine presence of legal facts, doctrines and argumentation.

To be sure, these methods require a researcher to be sufficiently skilled in traditional legal or socio-legal approaches – to both create prompts and conduct critical quality assurance. Nonetheless, such LLMs allow for rapid and user-friendly predictive and generative analysis of large volumes of legal text due to the simplicity of the methodological interface.

The second feature of LLMs provide mirrors the first. While LLMs allow simple creation of what previously required sophisticated software, it also allows for rapid development, expansion, and augmentation of previous computational methods. Until recently, computational researchers had to greatly limit the scope of their analyses due to capacity constraints – human and technical. But the use of LLMs allow programmers to address considerably this burden, while obtaining potentially more accurate and reproducible results.¹³³ The experimentation with the potential of leveraging LLMs for greater methodological power and sophistication is only now beginning.

4 Limitations of the Computational Turn

Although computational science provides legal scholars with the ability to answer new and old questions, this form of research is not without its limitations. Like any method, it has its weaknesses. It is possible to point towards three common or core issues obstacles. They concern the breadth of questions that can be answered (suitability); the calibre of datasets and the relevant methods (quality); and the competence to use the methods (feasibility). Within each, there are trade-offs between effectiveness and relevance.

4.1 *Suitability*

Computational legal studies is particularly *suitable* for answering certain types of research questions. It has an advantage in analysing large volumes of information and uncovering less-observable patterns: it is macroscopic.¹³⁴ Moreover, it can accommodate both text and numbers, making it highly dexterous; and some of its methods can be mobilised to predict outcomes and patterns with remarkable accuracy, including generating comprehensible text and dynamic analytics.

However, its macroscopic strength is also its Achilles heel. Computational methods are less adept at providing both interpretive and causal understandings.

¹³³ Meyer et al., *supra* note 133; Pudari and Ernst, *supra* note 133.

¹³⁴ Graham et al., *supra* note 69.

For example, legal network analysis can provide a rich overview of citation patterns and provide indications of the power of certain international judgments and related actors; but the methods are not yet sophisticated enough, like all other methods, to identify conclusively why it was cited and how this citation affected the subsequent legal analysis. Moreover, it often struggles with tasks like understanding if the citing actor agreed with the cited judgment/text. Legal, qualitative analysis, text analysis (sentiment analysis), or LLMs models would be needed for that.

The same applies to machine learning analysis. While it is possible to use feature analysis and reverse engineering methods to understand how these algorithms make their predictions, these provide only a limited insight into highly complex techniques, such as those that use neural networks.¹³⁵ A black box often remains. Moreover, given the reliance on past data, even when recurrently updated, these methods can be hamstrung in identifying how new and different situations might be tackled.¹³⁶

Nonetheless, it is possible to reduce the magnitude of these challenges through different management and mitigation strategies. Suitability can be managed through a greater focus on the appropriateness of the method for a specific research question. Descriptive and exploratory challenges can be often managed through technical fixes; but causal, critical, or creative challenges will often require mixed methods. The latter is especially appropriate when a singular method will result in evidential ambiguity: additional methods are thus needed to confirm a finding or compensate for methodological weaknesses.¹³⁷ For example, Cosette Creamer used computational text analysis on thousands of WTO documents to create a mood indicator of states, which became the central independent variable in a standard regression analysis of possible reflexive behaviour by adjudicators in the WTO Dispute Settlement Body.¹³⁸

¹³⁵ P. Linardatos, V. Papastefanopoulos and S. Kotsiantis, 'Explainable AI: A Review of Machine Learning Interpretability Methods', 23 *Entropy* (2021) p. 18; R. Roscher et al. 'Explainable Machine Learning for Scientific Insights and Discoveries' 8 *IEEE Access* (2020) available at <<https://ieeexplore.ieee.org/abstract/document/9007737>> (visited 23 February 2022).

¹³⁶ H. Surden, 'Machine Learning and Law Essay', 89 *Washington Law Review* (2014) pp. 87–116.f.

¹³⁷ M. L. Small, 'How to Conduct a Mixed Methods Study: Recent Trends in a Rapidly Growing Literature', 37 *Annual Review of Sociology* (2011) pp. 57–86.

¹³⁸ C. Creamer, 'Between the Letter of the Law and the Demands of Politics: Judicial Balancing of Trade Authority within the WTO', (2015) with results published also in Langford, Creamer and Behn, 'Regime Responsiveness in International Economic Disputes', in D. Behn, M. Langford and S. Gáspár-Szilágyi (eds.), *Adjudicating Trade and Investment Disputes: Convergence or Divergence?* (2020) pp. 244–284. For a similar use

4.2 *Quality*

The second challenge is *quality*, which in turn can be separated into quality of data and quality of method. Typically, when we consider the accuracy of computational methods, we measure to what extent an algorithm produces the expected and correct answer, compared to when it does not. While the above-mentioned results are undoubtedly impressive, the models are wrong in 5–40% of cases. In some cases, this result is much better than achieved by human forecasters; at other times it is worse.¹³⁹ As such, the idea of ensuring optimal quality in each step of the method is key to ensuring reliable results.

As to the quality of the data, there are questions of completeness and reliability. If one is to do a prediction-based analysis of, say, the practice of the ECtHR, a broad and representative dataset – a complete dataset – is key to achieving sufficiently trustworthy results. Missing cases, poor case selection, or the presence of legal components within fact summaries with legal components will distort the method: otherwise, it is ‘garbage in, garbage out’.¹⁴⁰ This challenge is not only germane to computational studies – it is a longstanding issue with empirical studies in general. Yet, due to the inductive nature of many computational methods, combined with large volumes of data, flawed data may be more difficult to identify, isolate, and correct. Likewise, reliability issues can emerge in data collection. While recent legal documents are available as ‘digital-first’, older documents may have been scanned and been subject to several layers of processing before reaching the final stages of analysis. If the quality of such processing is low, key features may be incorrect or missing.

Even with complete and reliable datasets, the choice and implementation of methods will further have an impact on the quality of any analysis. Three factors are relevant. The first is the choice of method or algorithm. Algorithms/computational methods are often specific and closely tailored to address one type of research question under a narrowly defined set of conditions. If an incorrect algorithm is chosen, this may have significant seen and unseen effects on accuracy. The second are ‘bugs’. Computational methods are often complex to implement, and even small errors, such as misspellings or incorrectly placed

of computational methods at the data gathering stage, but for doctrinal analysis see B. Joubert, ‘Sanitary and Phytosanitary Measures as Barriers to Trade: A South African Perspective’ (2022) (PhD Dissertation thesis, available at University of Oslo and North-Western University).

139 N. Hilborne, *AI Crunches Lawyers in Case Prediction Challenge*, 31 October 2017, Legal Futures, available at <<https://www.legalfutures.co.uk/latest-news/ai-crunches-lawyers-case-prediction-challenge>> (visited 9 November 2023); Katz, Bommarito and Blackman, *supra* note 99.

140 See Medvedeva, Vols and Weiling’s discussion of this issue.

commas, may have large effects. The third is the probabilistic non-deterministic strategy of most modern machine learning algorithms, in particular neural networks, i.e., given the same data and the same algorithm the results may differ, which creates an inherent level of uncertainty when interpreting results.¹⁴¹ In the case of LLMs, this challenge can be acute: some of them currently have a tendency to hallucinate, creating false data as the generative power of the model overrides its predictive foundation.¹⁴² When using these tools for legal analysis, detecting and filtering out such hallucinations is necessary.¹⁴³

Bringing the issues of data and methodological quality together, it is important that researchers are conscious of these challenges and prudent in addressing them. This includes the necessary safeguards and robustness checks, such as validation techniques.¹⁴⁴

4.3 *Feasibility*

This brings us naturally to the final limitation: feasibility. A key question for many is how to master, technically and ethically, these new methods as well as the broader literatures behind them.¹⁴⁵ In legal education, in most jurisdictions, methodological singularity reigns supreme. Thus, one risks 'dilettantism', seeking to rise above a single method yet failing to master multiple methods,¹⁴⁶ and sustain this expertise over time.¹⁴⁷

In facing these practical challenges, there are four possible responses. The first is to *moderate ambition* according to capacity and context. One can start with small steps, such as learning basic computational skills (e.g. simple computational text analysis) or good prompt engineering for LLM

¹⁴¹ S. Scardapane and D. Wang, 'Randomness in Neural Networks: An Overview', 7 *WIREs Data Mining and Knowledge Discovery* (2017) p. e1200.

¹⁴² H. Alkaissi and S. McFarlane, 'Artificial Hallucinations in ChatGPT: Implications in Scientific Writing', *Cureus* (2023), available at <<https://www.cureus.com/articles/138667-artificial-hallucinations-in-chatgpt-implications-in-scientific-writing>> (visited 26 July 2023); R. Azamfirei, S. R. Kudchadkar and J. Fackler, 'Large Language Models and the Perils of Their Hallucinations', 27 *Critical Care* (2023) p. 120; B. Peng et al., 'Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback', arXiv:2302.12813, 8 March 2023.

¹⁴³ Peng et al., *supra* note 143.

¹⁴⁴ T. Burzykowski et al., 'Validation of Machine Learning Algorithms', 164 *American Journal of Orthodontics and Dentofacial Orthopedics* (2023) pp. 295–297.

¹⁴⁵ D. Beach and J. G. Kaas, 'The Great Divides: Incommensurability, the Impossibility of Mixed-Methodology, and What to Do about It', 22 *International Studies Review* (2020) pp. 214–235.

¹⁴⁶ M. Nissani, 'Ten Cheers for Interdisciplinarity: The Case for Interdisciplinary Knowledge and Research', 34 *The Social Science Journal* (1997) pp. 201–216.

¹⁴⁷ Small, *supra* note 138.

approaches.¹⁴⁸ The second is a *full-blooded commitment* to learn, and undertake the necessary training.¹⁴⁹ To be sure, this approach is challenging, and requires humility, but can be also meaningful, providing new perspectives, cognitive distance from one's existing paradigms, and space for creativity.¹⁵⁰ The third is to *collaborate* with others who have acquired the requisite expertise, a common solution in more complex interdisciplinary research.¹⁵¹ Importantly, this is equally applicable to data scientists who cross into legal territory,¹⁵² they may need to 'need to lawyer up' in researching international law.¹⁵³

The final approach is *systemic*. It is to improve undergraduate and postgraduate education, as has been done in some bachelor, master, and Ph.D programmes.¹⁵⁴ The increasing use of computational methods by international NGOs and law firms will also trigger a demand for changes to legal education.¹⁵⁵

5 Conclusion

Computational methods have arrived in international law research. In the last decade, researchers have turned to techniques such as network analysis, text analysis, and machine learning to tackle long-standing and new questions. Beyond this epistemological contribution, many see ontological advantages, with the potential for the inductive and macroscopic logic of computational approaches to reorient established understandings of international law. However, the challenges of fusing macroscopic methods with the

¹⁴⁸ Alkaissi and McFarlane, *supra* note 143; Azamfirei, Kudchadkar and Fackler, *supra* note 143; Peng et al., *supra* note 143.

¹⁴⁹ For example, a legal Ph.D scholar, based at a South African university in South Africa, recently used some of the time in her doctoral degree to learn the R-statistics package and quantitative content analysis. She did this by joining an online scholarly collective. See, Joubert, *supra* note 139.

¹⁵⁰ T. B. Kashdan and P. J. Silvia, 'Curiosity and Interest: The Benefits of Thriving on Novelty and Challenge', in S. J. Lopez and C. R. Snyder (eds.) *Oxford Handbook of Positive Psychology* (Oxford University Press, 2009) pp. 367–374.

¹⁵¹ For examples, see, Langford, Behn and Lie, *supra* note 58; Medvedeva, Vols and Wieling, *supra* note 99; Pauwelyn and Pelc, *supra* note 64.

¹⁵² A. Huneeus, 'Human Rights between Jurisprudence and Social Science', 28 *Leiden Journal of International Law* (2015) pp. 255–266.

¹⁵³ T. E. George, 'An Empirical Study of Empirical Legal Scholarship: The Top Law Schools', 81:1 *Indiana Law Journal* (2006).

¹⁵⁴ *Ibid.*

¹⁵⁵ See e.g. P. Ball and M. Price, 'Using Statistics to Assess Lethal Violence in Civil and Inter-State War', 6 *Annual Review of Statistics and Its Application* (2019) pp. 63–84.; A. Deeks, *supra* note 15.

often-bespoke character of doctrinal, and even socio-legal, research, the lack of researcher competence and easy-to-use software tools, and the varying quality of datasets¹⁵⁶ means that the computational turn is more *seminal* than transformational.

Nonetheless, and gazing into a crystal ball, we may see a more *transformational* turn in international law research. This is for three reasons. The first is the increased availability and completeness of datasets, already well on its way within international law. Efforts such as the EDIT database for investment treaties,¹⁵⁷ Jus Mundi for general international law,¹⁵⁸ Quemy's ECtHR computational dataset,¹⁵⁹ and a multitude of other database initiatives that contain large text samples in processable form allow researchers to skip the labour-intensive data collection phase. Moreover, the increased tendency of journals to require datasets to be published alongside the articles, makes replication and repurposing possible.¹⁶⁰

The second is the democratisation of computational methods brought forward by LLMs. Tools that were only available to computer scientists are becoming increasingly available to less tech-savvy researchers, with increasing quality as more data is added and models are fine-tuned. Moreover, the repurposing of published code also allow researchers to adapt an ever-growing pool of algorithms to international law research.

The final, and potentially most impactful driver is the development of goal-oriented AI platforms.¹⁶¹ These quasi-autonomous tools use LLMs to direct and instruct other software tools such as machine learning, statistics, NLP, and database interaction to autonomously, or with regular human interaction, pursue a goal stated by a researcher.¹⁶² In many ways these systems operate similarly to a reasonably skilled human assistant. While such systems are in

156 See, e.g., discussion on missingness in N. Strain et al., 'Compliance Politics and International Investment Disputes: A New Dataset' (2024), <available at <https://ssrn.com/abstract=4397922>>.

157 Alschner, Elsig and Polanco, *supra* note 16.

158 Jus Mundi, available at <<https://jusmundi.com>> (visited 7 June 2022).

159 Quemy and Wrembel, 'ECHR-OD: On Building an Integrated Open Repository of Legal Documents for Machine Learning Applications', 106 *Information Systems* (2022) p. 101822.

160 See e.g. Journal of International Economic Law *General Instructions*, Oxford Academic, available at <https://academic.oup.com/jiel/pages/general_instructions> (visited 31 October 2023).

161 M. Fezari, A. Al Dahoud and A. Al-Dahoud, 'From GPT to AutoGPT: A Brief Attention in NLP Processing Using DL' (2023); Firat and Kuleli, 'What If GPT4 Became Autonomous: The Auto-GPT Project and Use Cases', 3 *Journal of Emerging Computer Technologies* pp. 1–6.

162 Fezari, Al Dahoud and Al-Dahoud, *supra* note 162; Firat and Kuleli, *supra* note 162.

their infancy, and are currently too unreliable for research purposes, their accuracy can be significantly increased by natural language interactions with a skilled researcher. This possibility to interact and instruct research systems using natural language rather than programming algorithms could be a significant driver for research into areas of law that have previously not been explored.

Acknowledgements

The authors would like to thank the two peer reviewers for their very helpful comments, as well as acknowledge David Law for the framing of the arrival of computational legal studies as a ‘computational turn’.

This research was supported by the NordASIL project (financed by Nordforsk grant no. 105178) and the COPIID project (financed by the Research Council of Norway grant 326629).