# THE TRUTH IS OUT THERE: HOW LAWYERS CAN (MAYBE) PREDICT JUDGMENTS WITH MACHINE LEARNING

NICHOLAS CARDACI*

## I INTRODUCTION AND CONTEXT

Fox Mulder was right in saying 'the truth is out there.'[1] Much like the various antagonists on *The X-Files,* the world's truths can be confronting. If we look, we may find out that an institution is not operating as expected (i.e., according to its professed ideology.)[2] Indeed, in the imperfect real world, it is always possible for institutions to profess an ideology while engaging in contrary practices. This is as true today as it was historically. Take, for instance, Machiavelli's much-beloved Florentine Republic; which the Medici family covertly what was meant to be a free republic.[3] More recently, Western liberal democracies have experienced multi-generational 'power families' exercising dynastic political power,[4] despite liberal democracies being intended and indeed designed to *avoid* dynastic political power.[5] Additionally, Europe's civil law (i.e., 'continental system') systems have

---

* Lawyer and researcher. LLB, B.Comm (Finance). www.nicholascardaci.online.
[1] *The X Files* (Directed by Chris Carter, 20th Century Fox Television, 10 September 1993).
[2] See Alan Ryan, *On Machiavelli: The Search for Glory* (Liveright, 1st ed, 2014) 44; Ludger Helms, 'Leadership Succession in Politics: The Democracy/Autocracy Divide Revisited' (2020) 22(2) *The British Journal of Politics and International Relations* 328, 339; Daron Acemoglu and James A Robinson, *Why Nations Fail: The Origins of Power, Prosperity and Poverty* (Crown Publishers, 1st ed, 2012) 329–32. See especially Virginia Bell, *Report of the Inquiry into the Appointment of the Former Prime Minister to Administer Multiple Departments* (Commonwealth of Australia, 2022) 1–6, 89–90, 95 <https://www.ministriesinquiry.gov.au/publications/report-inquiry>.
[3] At this time, Florence was constitutionally a popular republic and was underpinned by an ideology that valued citizens' liberty. Nevertheless, due to Medici rule, 'Florentine ideology was at odds with Florentine practice.' See Ryan (n 2) 44.
[4] According to Ryan (2014),'[Florentine] practice suggested that as long as the Medici did not claim to rule by hereditary right and were good managers, they would be accepted as the rulers of Florence. This is not very different from modern liberal democracies, professional politicians beget professional politicians, or acquire them as sons- and daughters-in-law' according to Ibid. See also Helms (n 2) 339.
[5] Helms (n 2) 329–30, 339–40.

arguably had an unspoken doctrine of precedent applied by their contemporary courts, which goes against the ideas that these systems are based on.[6]

Given the possibility of institutional ideology and practise diverging, it is not unreasonable for lawyers and academia to look for such divergences – as it can have practical significance for legal practice and scholarship.[7] Fortunately, empirical legal research ('ELR')[8] is one method for identifying if and where any such divergences occur.[9] Thus, the truth is out there; *in the data*.

For instance, one such line of divergence-hunting inquiry has been to determine whether the rule of law is strongly operating in a given country, by quantitatively analysing court rulings using traditional statistics and machine learning.[10] This is a worthy inquiry as the rule of law is foundational to the proper functioning of liberal democracies such as Australia.[11] Such inquiries are especially popular among legal realists in attempts to validate their claims about the true nature of the legal system.

---

[6] John Henry Merryman, *The Civil Law Tradition* (Stanford University Press, 2nd ed, 1985) 83.
[7] 'Practical significance is concerned with whether an effect is big enough to affect practical action' according to DR Cox, 'Statistical Significance Tests' (1982) 14 *British Journal of Clinical Pharmacology* 325, 327.
[8] ELR is a broad field of study that 'encompass[es] both quantitative and qualitative studies of legal phenomena.' See Peter Cane and Herbert M Kritzer (eds), *The Oxford Handbook of Empirical Legal Research* (Oxford University Press, 1st ed, 2012) 1026.
[9] See e.g., S Danziger, J Levav and L Avnaim-Pesso, 'Extraneous Factors in Judicial Decisions' (2011) 108(17) *Proceedings of the National Academy of Sciences* 6889, 6889; Christina L Boyd, Lee Epstein and Andrew D Martin, 'Untangling the Causal Effects of Sex on Judging' (2010) 54(2) *American Journal of Political Science* 389; David Benjamin Oppenheimer, 'Verdicts Matter: An Empirical Study of California Employment Discrimination and Wrongful Discharge Jury Verdicts Reveals Low Success Rates for Women and Minorities' 37 57.
[10] Danziger, Levav and Avnaim-Pesso (n 9) 6889; Boyd, Epstein and Martin (n 9); Oppenheimer (n 9). The rule of law requires *inter alia*, the law applying equally to all litigants, and, fair and impartial resolution of court cases by the judiciary. See Anton Gradisek et al, *The World Justice Project: Rule of Law Index 2021.* (The World Justice Project, 2021) 14, 18–9 <https://worldjusticeproject.org/sites/default/files/documents/WJP-INDEX-21.pdf>; *Flowers v New South Wales (No 5)* [2021] NSWSC 887, [99]; Acemoglu and Robinson (n 2) 305–18; *A Matter of Judgment: Judicial Decision-Making and Judgment Writing* (Judicial Commission of New South Wales, 2003) 29 (Mason J) <https://www.judcom.nsw.gov.au/wp-content/uploads/2016/07/education-monograph-2.pdf>. Indeed, in furtherance of the rule of law, the *ICCPR* (to which Australia is a party) mandates that people be treated equally before impartial tribunals and courts. See *International Covenant on Civil and Political Rights (1966) 999 UNTS 171* art 14(1); Westlaw AU, *The Laws of Australia* (at 18 September 2015) '[21.6.610] Obligations under International Covenant on Civil and Political Rights'.
[11] '[T]he rule of law is fundamental in advancing democracy' according to Massimo Tommasoli, 'Rule of Law and Democracy: Addressing the Gap Between Policies and Practices' (2012) 49(4)

Despite the relevance of such divergence-hunting inquiries for lawyers, they often lack the quantitative data abilities required to conduct divergence-hunting inquiries, or other forms of ELR for that matter.[12] This isn't a huge surprise though as these skills are not necessarily part of an Australian legal education. This is because data skills do not appear in the mandatory Priestly 11 units taught in law school, or in the Graduate Diploma in Legal Practice's curriculum.[13]

This skill gap in the legal profession is a barrier to entry for lawyers undertaking ELR generally and divergence-hunting inquiries more specifically. This is to the detriment of lawyers as assessing the probable case outcomes is a part of legal practice and of keen interest to clients.[14] The author seeks to contribute, with this article, to breaking down this barrier.

The author aims to achieve this by providing readers with the theoretical basis *and* step-by-step guidance for using machine learning for a particular form of ELR; *attempting to*

---

*UN Chronicle* <https://www.un.org/en/chronicle/article/rule-law-and-democracy-addressing-gap-between-policies-and-practices>. See especially *Flowers v New South Wales (No 5)* (n 10) [104], [109]. See also Acemoglu and Robinson (n 2) 305–18, 333.

[12] Mark A Cohen, 'Law's Looming Skills Crisis', *Forbes* <https://www.forbes.com/sites/markcohen1/2019/05/21/laws-looming-skills-crisis/>; Michael Legg, *New Skills for New Lawyers: Responding to Technology and Practice Developments [2018] UNSWLRS 51* 4. The legal profession and legal academia traditionally use qualitative analyses of law characterised by manual collection, reading, summarisation and contextualisation of case law. See Masha Medvedeva, Michel Vols and Martijn Wieling, 'Using Machine Learning to Predict Decisions of the European Court of Human Rights' (2020) 28(2) *Artificial Intelligence and Law* 237, 239. This is an unfortunate omission as machine learning and other data analytic tools are becoming increasingly popular in legal practice. See Michael Legg and Felicity Bell, 'Artificial Intelligence and the Legal Profession: Becoming The AI-Enhanced Lawyer' (2019) 38(2) *University of Tasmania Law Review* 34, 44–53; Anthony E Davis, 'The Future of Law Firms (and Lawyers) in the Age of Artificial Intelligence' (2020) 27(1) *The Professional Lawyer* 3, 4–5; 'Lawyer vs Artificial Intelligence: A Legal Revolution', *LexisNexis* (2018) <https://www.lexisnexis.com.au/en/insights-and-analysis/practice-intelligence/2018/Lawyer-vs-AI-A-legal-revolution>; Catherine Nunez, 'Artificial Intelligence and Legal Ethics: Whether AI Lawyers Can Make Ethical Decisions' 20 16, 189–90; Monika Zalnieriute and Felicity Bell, 'Technology and the Judicial Role [2019] UNSWLRS 90' 1–8.

[13] Law Admissions Consultative Committee, 'Model Admission Rules 2015' schs 1, 2 <https://www.lawcouncil.asn.au/files/web-pdf/LACC%20docs/212390818_9_LACC%20Model%20Admission%20Rules%202015.pdf>.

[14] Daniel Martin Katz, Michael J Bommarito and Josh Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States', ed Luís A Nunes Amaral (2017) 12(4) *PLOS ONE* e0174698, 8; John A Humbach, 'Property as Prophesy: Legal Realism and the Indeterminacy of Ownership' (2017) 49 *Case Western Reserve Journal of International Law* 211, 213–4.

*predict judgments using machine learning*.[15] Along the way, the article gives a mock (i.e., purely for demonstration) analysis. This mock analysis helps illustrate how the described theory and guidance is applied *practically*.

Fortunately, this theory and guidance will be relevant to other types of ELR – so readers who are uninterested in predicting judgments can still obtain utility from this article for other ELR pursuits.

This article proceeds as follows.

- Part II explores the legal realist underpinnings of ELR which attempts to predict judgments using machine learning.
- Part III explains what machine learning is, and briefly outlines how it has been used in prior ELR for predicting judgments.
- Part IV walks the reader through a process that can be used in attempts to predict judgments using machine learning, while illustrating it with a mock analysis.
- The article then gives its concluding remarks, and briefly alerts readers to some ethical concerns involved with attempting to predict judgments using machine learning.

## II LEGAL REALISM

The practise of using machine learning to attempt to predict judgments is fundamentally underpinned by the philosophy of legal realism;[16] which posits that, to the detriment of the rule of law,[17] judicial decisions are affected by the 'judicial personalities' of the

[15] Masha Medvedeva et al, 'Automatic Judgement Forecasting for Pending Applications of the European Court of Human Rights' in *Proceedings of the Fifth Workshop on Automated Semantic Analysis of Information in Legal Text* (2021) 1 <http://ceur-ws.org/Vol-2888/paper2.pdf>; Thomas J. Miles and Cass R. Sunstein, 'The New Legal Realism' (2008) 75(2) *The University of Chicago Law Review* 831, 835–6.

[16] Thomas J. Miles and Cass R. Sunstein (n 15) 831–832, 835–6, 843; S Nayak-Young, 'Delimiting the Proper Bounds of the "New Legal Realism"' (2014) 12(4) *International Journal of Constitutional Law* 1008, 1008–1009. A new school called New Legal Realism ('NLR') is emerging as a result of rapid advances in computing power, online access to case law, and data collection. Unlike 'old' legal realism, NLR relies on large-scale quantitative analysis of facts and outcomes rather than personal anecdotes or impressions. As such, NLR hypotheses are tested using a scientific method. See Medvedeva et al (n 15) 1; Thomas J. Miles and Cass R. Sunstein (n 15) 835–6; Nayak-Young 1008–10; Frank B Cross, 'The New Legal Realism and Statutory Interpretation' (2013) 1(1) *The Theory and Practice of Legislation* 129, 136.

[17] Thomas J. Miles and Cass R. Sunstein (n 15) 836, 841, 844; John Hasnas, 'The Myth of the Rule of Law' [1995] *Wisconsin Law Review* 199.

presiding judges.[18] Further, judicial personality is allegedly comprised of a judge's beliefs (e.g., personal attitudes, political beliefs) and characteristics (e.g., demographic.)[19] Nonetheless, legal realists are realistic enough to recognise that judicial personality is not the only factor affecting judicial decision-making.[20] Instead, the effects of judicial personality are constrained by institutional realities and (ironically) the *personal* characteristics of judicial officers.[21] Institutional realities that constrain judicial personality include the presence of unambiguous law in some cases,[22] the doctrine of precedent,[23] the requirement for courts to give reasons for their decisions,[24] and the unattractive possibility of legislative or appellate abrogation of a judge's decision.[25]

---

[18] Cross (n 16) 130–1; Nayak-Young (n 16) 1008–9; Thomas J. Miles and Cass R. Sunstein (n 15) 831, 832, 835. Legal realism is the antithesis legal formalism, which holds that the law is applied without regard for judicial personality. See Brian Z Tamanaha, 'Understanding Legal Realism' (2009) 87 *Texas Law Review* 731, 731–2; Nayak-Young (n 16) 1009 fn 2, 1030; Anthony Mason, 'Future Directions in Australian Law' (1987) 13 *Monash University Law Review* 149, 156; Nikolaos Aletras et al, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 *PeerJ Computer Science* 93, 11.

[19] Thomas J. Miles and Cass R. Sunstein (n 15) 831, 832, 835; Robert French, 'Judicial Activism – The Boundaries of the Judicial Role' (2009 at the Law Asia Conference, Ho Chi Minhh City, Vietnam) 7; Nayak-Young (n 16) 1008–9; Tamanaha (n 18) 731–2, 768. Perhaps controversially, some researchers inquire into whether racial and/or sexual characteristics are part of this personality. See Boyd, Epstein and Martin (n 9); Jonathan P Kastellec, 'Racial Diversity and Judicial Influence on Appellate Courts' (2013) 57(1) *American Journal of Political Science* 167.

[20] Thomas J. Miles and Cass R. Sunstein (n 15) 835–6; Tamanaha (n 18) 731–2, 767–70; Nayak-Young (n 16) 1017; Albie Sachs, 'The Myth of Judicial Neutrality: The Male Monopoly Cases' (1975) 23(1) *The Sociological Review* 104, 112.

[21] See generally Nayak-Young (n 16) 1017; Tamanaha (n 18) 732–3, 767–70; Cross (n 16) 133–7.

[22] Thomas J. Miles and Cass R. Sunstein (n 15) 836, 844; Tamanaha (n 18) 732–3; Cross (n 16) 135.

[23] Sachs (n 20) 111–2; Cross (n 16) 133–5, 137; Tamanaha (n 18) 766–8. See also *Palmer v McGowan (No 5)* [2022] FCA 893, [71]-[72], [212]-[213]; David Rolph et al, *Law of Torts* (LexisNexis Australia, 2021) [11.28]; James Allsop, 'The Future of the Independent Bar in Australia' (at the Australian Bar Association and NSW Bar Association Biennial Conference, Federal Court of Australia, 17 November 2018) <https://www.fedcourt.gov.au/digital-law-library/judges-speeches/chief-justice-allsop/allsop-cj-20181117>.

[24] Tamanaha (n 18) 767. See especially *A Matter of Judgment: Judicial Decision-Making and Judgment Writing* (n 10) 44–5 (Kirby J). See also TF Bathurst, 'Who Judges the Judges, and How Should They Be Judged?', *Handbook for Judicial Officers* (2021) at 'Accountability' <https://www.judcom.nsw.gov.au/publications/benchbks/judicial_officers/who_judges_the_judges.html>; *A Judicial Officer v the Judicial Conduct Commissioner* [2022] SASCA 42, [53], [54]. Cf. Jerome Frank, 'Are Judges Human? Part One: The Effect on Legal Thinking of the Assumption That Judges Behave like Human Beings' (1931) 80(1) *University of Pennsylvania Law Review and American Law Register* 17, 37–8 ('Are Judges Human?').

[25] Thomas J. Miles and Cass R. Sunstein (n 15) 835–6; Sachs (n 20) 111–2; Tamanaha (n 18) 767–8; Cross (n 16) 138–9; Bathurst (n 24) at 'Accountability'; *Williams v The Minister, Aboriginal Land Rights Act 1983* [1999] NSWSC 843, [97]-[98]. Ashley B. Antler, 'The Role of

Similarly, personal characteristics that constrain judicial personality include judges' genuine desire to act 'judicially' (and thus in accordance with their oaths),[26] and judges' respect for democracy.[27]

## A *'HARD CASES'*

Legal realists disagree amongst each other on just how much a judge's personality matters, but they do agree that it matters.[28] Further, the legal realist consensus is that in 'hard cases,' judicial personality becomes more important to the outcome of the case.[29] 'Hard cases' are cases in which the law is silent or ambiguous in regard to the legal issues in the case.[30] 'Hard cases', which populate law reports,[31] are unavoidable given the inexorable emergence of unique factual circumstances to which the law must be applied.[32]

---

Litigation in Combating Obesity among Poor Urban Minority Youth: A Critical Analysis of Pelman v. McDonald's Corp' (2009) 15 *Cardozo Journal of Law & Gender* 275, 295–6.

[26] See generally Tamanaha (n 18) 766–7; Nayak-Young (n 16) 1012–1013; Cross (n 16) 134–5; Thomas J. Miles and Cass R. Sunstein (n 15) 835–6; *A Matter of Judgment: Judicial Decision-Making and Judgment Writing* (n 10) 36-38 (Mason J). See especially S Roach Anleu and K Mack, 'Impartiality and Emotion in Judicial Work' in *Handbook for Judicial Officers* (Judicial Commission of New South Wales, 2021) 'The Judicial Role' <https://www.judcom.nsw.gov.au/publications/benchbks/judicial_officers/impartiality_and_emotion_in_judicial_work.html>; James Thomas, *Judicial Ethics in Australia* (LexisNexis Butterworths, 3rd ed, 2009) 35–6.

[27] Tamanaha (n 18) 767; *5 Boroughs NY Pty Ltd v Victoria* [2021] VSC 785, [52]; *Elliott v Minister administering Fisheries Management Act 1994* [2018] Supreme Court of New South Wales 117, [151]-[156]; *Re Town Planning Appeal Tribunal; Ex Parte Environmental Protection Authority* [2003] Supreme Court of Western Australia - Court of Appeal 248, [101]; Allsop (n 23); James Allsop, 'Being a Judge: Judicial Technique, Independence and Labels' (Speech at the Samuel Griffith Society Conference, Federal Court of Australia, 30 April 2022) <https://www.fedcourt.gov.au/digital-law-library/judges-speeches/chief-justice-allsop/allsop-cj-20220430>; Anthony Mason (n 18) 156; *Pacific Gas & Elec Co v State Energy Resources Conservation and Development Comm'n, 461 US 190 (1983)* 222–3; *Faris v Risdea Holdings Limited* [2005] NZLLA 430, [46]. See also Nayak-Young (n 16) 1017.

[28] Tamanaha (n 18) 731–3, 767–9; Thomas J. Miles and Cass R. Sunstein (n 15) 835–6. See also Humbach (n 14) 216; Cross (n 16) 133–6.

[29] Nayak-Young (n 16) 1009 fn 2, 1016; Humbach (n 14) 211, 213–4; Frederick Schauer, 'Legal Realism Untamed' (2012) 38 *Virginia Public Law and Legal Theory Research Paper* 3–6 <http://www.ssrn.com/abstract=2064837>; Tamanaha (n 18) 732–3; Thomas J. Miles and Cass R. Sunstein (n 15) 836, 844.

[30] Thomas J. Miles and Cass R. Sunstein (n 15) 835–6; Nayak-Young (n 16) 1008, 1015–6; Humbach (n 14) 211, 213–6; Cross (n 16) 135–6, 143, 145; Schauer (n 29) 3–6. Cf. *Harriton v Stephens* (2006) 226 CLR 52, [110], [143]-[144] (Kirby J).

[31] Sachs (n 20) 116. See also Thomas J. Miles and Cass R. Sunstein (n 15) 841.

[32] Cross (n 16) 130, 145. See also *Palmer v McGowan (No 6)* [2022] FCA 927, [6]; *Hurt v The Queen* [2022] ACTCA 49, [134].

Not all cases are 'hard cases' though. Indeed, many cases involve settled law and potentially simple facts as well.[33] According to legal realists, in these 'easy cases' judicial personality is less important and so such cases are more readily decided without (or at least with less influence from) judicial personality.[34] These 'easy cases' are arguably more common in lower courts than appellate courts, due to the role of appellate courts.[35]

### B *Australia's Legal (Un)Realism*

Legal realism is controversial, and there are good reasons to believe it does not accurately describe Australia.

First, legal realism seems to be implicitly rejected by Australia's judiciary given its public insistence that it must exercise judicial power independently, impartially, and without ideological bias.[36] This is consistent with Australia's judicial oaths – which the judiciary take very by all accounts .[37]

Second, Australia's judiciary is explicit that its judicial power exists to determine the law's content and apply it to each case in accordance with established judicial techniques and *stare decisis*, rather than as a means for setting public policy.[38]

---

33 Cross (n 16) 135; Tamanaha (n 18) 732–3; Thomas J. Miles and Cass R. Sunstein (n 15) 836, 844. See also Humbach (n 14) 216; *Guneser v Aitken Partners* [2019] VSC 649, [71].

34 Schauer (n 29) 7, 12–15; Cross (n 16) 133–4, 135–7; Nayak-Young (n 16) 1009 fn 2.

35 Cross (n 16) 134, 136–7; *A Matter of Judgment: Judicial Decision-Making and Judgment Writing* (n 10) 44, 46 (Kirby J). See also *Judiciary Act 1903* (Cth) s 35A; Brendan Sweeney, Mark Bender and Nadine Courmadias, *Marketing and the Law* (LexisNexis Australia, 5th ed, 2015) 24; *Valve Corporation v Australian Competition and Consumer Commission* [2018] HCASL 99.

36 See especially *A Matter of Judgment: Judicial Decision-Making and Judgment Writing* (n 10) 36–7, 39 (Mason J), 44–5 (Kirby J). See also Anleu and Mack (n 26) 'The judicial oath'; *Harriton v Stephens* (n 30) [110],[205]; Allsop (n 27); *Re Town Planning Appeal Tribunal; Ex Parte Environmental Protection Authority* (n 27) [99]-[102]; *A Judicial Officer v the Judicial Conduct Commissioner* (n 24) [57]-[58]; *R v CK* [2022] QChC 18, [22]-[25].

37 See generally Sachs (n 20) 108. See especially Anleu and Mack (n 26); Allsop (n 27); 'Criticism of the Courts and Judges: Informed Criticism and Otherwise' (21 May 2018) 38–9, 41 <https://www.hearsay.org.au/criticism-of-the-courts-and-judges/>; *A Matter of Judgment: Judicial Decision-Making and Judgment Writing* (n 10) 39 (Mason J).

38 See *Palmer v McGowan (No 5)* (n 23) [71]-[72], [212]-[232]; Sachs (n 20) 111–2; Anthony Mason (n 18) 155–6; Rolph et al (n 23) [11.28]; 'Criticism of the Courts and Judges: Informed Criticism and Otherwise' (n 37) 41. See especially *Harriton v Stephens* (n 30) [110], [143]-[144] (Kirby J). See also Brendan Sweeney, Mark Bender and Nadine Courmadias (n 35) 13, 20, 24; Frank (n 24) 19.

Third (and relatedly), the judiciary is explicitly and consistently disinterested in determining cases on the basis of the judiciary's policy preferences, and instead generally leaves matters of policy to Parliament or the Executive.[39]

Further reasons for doubting legal realism's accuracy in Australia are that (a) empirical evidence shows that the rule of law is strong in Australia,[40] and (b) common sense suggests that common law judiciaries have not, for centuries, wholly 'lied to themselves' (and the public) that they discharge their duties properly.[41]

Given the foregoing, legal realism is seemingly inaccurate in the Australian context. Nonetheless, in the author's view it is still worthwhile using ELR methods to be able to objectively test for the presence of the rule of law in Australia given that even Australian institutions *can* diverge from expected practise.[42] As such, the author moves on to explore machine learning, and its potential application in attempts to predict judgments.

## III WHAT IS MACHINE LEARNING?

Machine learning is a sophisticated way to pull insights from case law, which is a near limitless and increasingly accessible product of courts.[43] Machine learning, in short, involves automated 'learning', by an algorithm, of relationships (i.e., 'patterns' or

---

[39] See especially Thomas (n 26) 39–40. See also *Commonwealth Director of Public Prosecutions v Evans* [2022] FCAFC 182, [36]; *CAL No 14 Pty Ltd v Motor Accidents Insurance Board* (2009) 239 CLR 390, [54]; *Williams v The Minister, Aboriginal Land Rights Act 1983* (n 25) [95]-[98]; *5 Boroughs NY Pty Ltd v Victoria* (n 27) [52](d); *Athavle v State of New South Wales* [2021] FCA 1075, [96], [113]-[115]; *Palmer v McGowan (No 5)* [2022] FCA 893, [353]-[357]; *Dennis v Parramatta City Council* (1981) 43 LGRA 71, 74–5. Similarly, the House of Lords' Viscount Dilhourne opined that the House's judicial function is to 'declare what the law is, not what we think it ought to be' in *Cassell & Co Ltd v Broome (No 1)* [1972] AC 1027. See also Sachs (n 20) 108–9.

[40] The World Justice Project ranks Australia highly at 13th place (out of 128 countries) for rule of law. See Gradisek et al (n 10) 45. Similarly, Australia's strong rule of law is noted in Heritage Foundation, 'Australia', *2022 Index of Economic Freedom* <https://www.heritage.org/index/country/australia>.

[41] Nayak-Young (n 16) 1012–3. Cf. Sachs (n 20) 115.

[42] Virginia Bell (n 2) 1–6, 89–90, 95.

[43] Medvedeva et al (n 15) 1; Thomas J. Miles and Cass R. Sunstein (n 15) 835–6.

'associations') between data points within large datasets.[44] Armed with this learning, an algorithm can then attempt to make predictions about unseen data (i.e., future cases.)[45]

It is worth emphasising that use of the term 'learning' does not imply that machine learning algorithms are mimicking the cognitive process of human learning.[46] Instead, the algorithms detect relationships by using statistical, probabilistic or optimisation formulas.[47]

There are various types of machine learning,[48] but in the author's view the simplest to understand and apply is supervised machine learning – which is the type that the author focuses on for this article.

### A *Supervised Machine Learning*

Supervised machine learning involves algorithms being 'trained' – which involves 'learning' the relationships/patterns within categorised (also called 'labelled') datasets so that, after training, the now-trained algorithm can potentially predict the categories of new/unseen data that is not discernibly categorised (i.e., the new dataset *actually* has no categories, or, the new data's category is made unavailable to the trained algorithm.)[49]

Such technology can be used to predict unseen cases' outcomes. To do this, a supervised algorithm is trained on a dataset containing information from/about many court cases (e.g., presiding judge, cited authorities) together with the actual verdicts (e.g., appeal dismissed).[50] By training the algorithm on this dataset (in the so-called 'training phase'), the algorithm learns the relationships between the case information and each class of

---

[44] Shahadat Uddin et al, 'Comparing Different Supervised Machine Learning Algorithms for Disease Prediction' (2019) 19(1) *BMC medical informatics and decision making* 281, 1; Harry Surden, 'Machine Learning and Law' 89(1) *Washington Law Review* 87, 89.
[45] Uddin et al (n 44) 1; Surden (n 44) 89.
[46] Surden (n 44) 89.
[47] Uddin et al (n 44) 1.
[48] See generally Uddin et al (n 44); Sunan Cui et al, 'Introduction to Machine and Deep Learning for Medical Physicists' (2020) 47(5) *Medical Physics* e127; Jonathan Beach, 'Causation: The Interface Between the Scientific and Legal Methods' (2022) 49(1) *University of Western Australia Law Review* 11, 152.
[49] Surden (n 44) 89; Beach (n 48) 152. An algorithm is a procedure for solving a mathematical problem in a finite number of steps that frequently involves repetition of an operation: *Merriam-Webster* (online at 14 August 2022) 'Algorithm'.
[50] Medvedeva, Vols and Wieling (n 12) 242.

verdict.[51] The trained algorithm can then be used to attempt to predict the verdicts in new cases.[52] This is done by providing the trained algorithm with case information concerning a new case, and instructing the trained algorithm to predict the verdict based on that information *based on the relationships learned during the training phase*.[53] These predictions may or may not be accurate. As such, the trained algorithm can and indeed must be tested for accuracy in order to assess the predictive power of the trained algorithm.[54] Indeed, as explained by Medvedeva, Vols and Wieling (2020):

> To evaluate the performance of the machine learning program, it is provided with a case without the judgement (in the 'testing phase') for which it has to provide the most likely judgement. To make this judgement (also called: 'classification') the program uses the information it identified to be important during the training phase.[55]

There are numerous supervised machine learning algorithms for classification.[56] For instance, the naïve Bayes algorithm is based on Bayes' theorem of conditional probability.[57] In contrast, the simpler K-Nearest Neighbour ('KNN') algorithm classifies unseen data in the same way as the most similar training data.[58]

## B *Illustrative Examples of Supervised Machine Learning*

Consider the detection of spam emails as an example of supervised machine learning in action. This requires training the algorithm on a set of emails containing spam emails and non-spam emails, which are categorised as such.[59] In the training phase, the algorithm would learn that there is a strong positive relationship between emails mentioning 'Extra

---

[51] Ibid.
[52] Jenni AM Sidey-Gibbons and Chris J Sidey-Gibbons, 'Machine Learning in Medicine: A Practical Introduction' (2019) 19(1) *BMC medical research methodology* 64, 12 ('Machine Learning in Medicine')
[53] Masha Medvedeva, 'Identification, Categorisation and Forecasting of Court Decisions' (University of Groningen) 16–8, 90–100 <https://research.rug.nl/en/publications/80f24952-d911-4c20-962c-61e751efd9f1>; Uddin et al (n 44) 2; Surden (n 44) 89. Potentially, a trained algorithm 'is able to automatically predict the category (i.e., a verdict) associated to a new element (i.e., a case).'[53] See Medvedeva, Vols and Wieling (n 12) 242.
[54] Medvedeva, Vols and Wieling (n 12) 242.
[55] Ibid.
[56] See generally Uddin et al (n 44).
[57] Ibid 4.
[58] Ibid 5.
[59] Surden (n 44) 91.

Cash' and being spam.[60] Consequently, the trained algorithm will predict that future emails containing that phrase are also spam.[61]

Another illuminating example of the machine learning process is provided by Medvedeva, Vols and Wieling (2020):

> To illustrate how supervised machine learning works, let's imagine a non-textual example. Suppose we want to write a program that recognises pictures of cats and dogs. For that we need a database of images of cats and dogs, where each image has a label: either cat or dog. Then we show the system those pictures with labels one by one. If we show enough pictures, eventually the program starts recognising various characteristics of each animal, e.g., cats have long tails, dogs are generally more furry. This process is called training or fitting the model. Once the program learns this information, we can show it a picture without a label and it will guess which class the picture belongs to.[62]

### C *Prior Use of Machine Learning in Empirical Legal Research*

The legal academy has realised the potential of machine learning. Accordingly, many studies have validated the use of machine learning in attempting to better predict the outcomes of future cases.[63]

For instance, Aletras et al (2016) collected judgments from the European Court of Human Rights. These judgments were divided into two categories, (a) rights violation found, and (b) rights violation not found. From all judgments, words and phrases were collected (except from the parts of the judgments which state the verdict.) The resulting dataset was used to train a supervised machine learning algorithm. The algorithm learned which words and phrases were most associated with a rights violation being found. Consequently, when presented with the equivalent words and phrases from newer cases, the trained algorithm could predict the correct verdict with 79% accuracy.[64]

---

[60] Ibid.
[61] Ibid.
[62] Medvedeva, Vols and Wieling (n 12) 242.
[63] Medvedeva et al (n 15) 1; Thomas J. Miles and Cass R. Sunstein (n 15) 835–6.
[64] Aletras et al (n 18).

In addition, Katz, Bommarito and Blackman (2017) used machine learning in an attempt to predict appeal outcomes in the Supreme Court of the United States ('SCOTUS').[65] The authors used a dataset of 28,000 case outcomes (from 1836 to 2015).[66] For each case, the dataset contained information including which lower court the case originated from, whether the court below was unanimous in its judgment, the manner in which SCOTUS took jurisdiction, and why SCOTUS gave leave to appeal.[67] By training a time-evolving random-forest algorithm with this dataset, the trained algorithm could accurately with significant accuracy, by correctly predicting unseen cases' outcomes 70.2% of the time.[68]

While not explored in depth here, other studies have attempted to measure the influence of the time of day[69] and a judicial officer's gender[70] on judicial decision-making.

## IV WALKTHROUGH OF SUPERVISED MACHINE LEARNING

As foreshadowed, this article will now provide guidance on conducting ELR which uses supervised machine learning in an attempt to predict future cases. The guidance progresses step-by-step and is complemented by an illustrative mock analysis.

### A *Understanding the Legal Context of the Data*

A researcher must first understand the law underpinning the relevant cases and be able to explain it clearly.[71] For instance, Creyke (2017) is right to carefully understand, describe and explain the relevant administrative law before describing her empirical survey of case law.[72]

---

[65] Katz, Bommarito and Blackman (n 14).
[66] Ibid.
[67] Ibid 5.
[68] Katz, Bommarito and Blackman (n 14).
[69] Danziger, Levav and Avnaim-Pesso (n 9) 6889.
[70] Boyd, Epstein and Martin (n 9).
[71] See Robin Creyke, 'Judicial Review and Merits Review: Are the Boundaries Being Eroded?' (2017) 45(4) *Federal Law Review* 627 ('Judicial Review and Merits Review'); *Haritos v Commissioner of Taxation* (2015) 233 FCR 315; *Administrative Appeals Tribunal Act 1975* (Cth) s 44(7) ('*AAT Act*').
[72] Creyke (n 71).

As such, for the sake of completeness in the mock analysis, this article provides information on the relevant law and institutions.

### 1 *The Federal Court of Australia*

The Federal Court of Australia is a federal superior court of record with general appellate and original jurisdiction in both law and equity.[73] In the federal judicial hierarchy, the Federal Court sits below the High Court of Australia but above the Federal Circuit and Family Court of Australia ('FCFCOA').[74] The Federal Court is seated across Australia, in all State and Territory capital cities, but also sits elsewhere from time to time.[75].

The Federal Court as a whole has limited docket control; it generally must take cases as they come, without picking and choosing the cases that it finds important to the exclusion of the others.[76] Even for appeals, the Court's docket control is limited.[77] Indeed, appeals from the FCFCOA generally *must* be heard.[78]

---

[73] Federal Court of Australia, 'The Court's Jurisdiction' (22 September 2021) <https://www.fedcourt.gov.au/about/jurisdiction>; *Federal Court of Australia Act 1976* (Cth) ss 5, 19 ('*FCA Act*'); *Judiciary Act 1903* (n 35) s 39B(1A)(c). See also *Palmer v McGowan (No 6)* (n 32) [6]. In the hierarchy of Commonwealth courts, it is inferior only to the High Court of Australia. See Federal Court of Australia, *The Court's Jurisdiction*; *FCA Act* s 33; *Judiciary Act 1903* (n 35) s 40(1); Federal Court of Australia, 'Appeals from Courts' (4 November 2012) <https://www.fedcourt.gov.au/law-and-practice/guides/appeals/from-courts>.

[74] *Storry v Weir* [2022] FCA 1484, [1]-[13], [27]-[28]; *FCA Act* (n 73) ss 24(1)(d), 25(1AA), 33.

[75] Federal Court of Australia, *The Court's Jurisdiction* (n 73). The Court's registries provide operational support to the judges, provide registry services to legal practitioners and members of the public, receive court and related documents, assist with the arrangement of court sittings, and facilitate the enforcement of orders made by the Federal Court. See Federal Court of Australia, 'Registry Services: What Staff Can and Cannot Do' (April 2013) <https://www.fedcourt.gov.au/services/registry-services>.

[76] See generally *FCA Act* (n 73) ss 20, 25. See also Ashley B. Antler (n 25) 295–6. Cf. David Fontana, 'Docket Control and the Success of Constitutional Courts' in *Comparative Constitutional Law* (Edward Elgar Publishing, 2011) 624–6 <http://ssrn.com/abstract=2256946>. Unlike the Federal Court, the High Court of Australia has broad discretion to refuse to hear appeals. This discretion is broad enough that the High Court can refuse to hear appeals that it deems to be publicly unimportant or have low prospects of success. See *Judiciary Act 1903* (n 35) s 35A; Brendan Sweeney, Mark Bender and Nadine Courmadias (n 35) 24; *Valve Corporation v Australian Competition and Consumer Commission* (n 35). See also Cross (n 16) 135–6.

[76] Federal Court of Australia, 'Allocations of Judicial Matters under the NCF' (21 April 2018) <http://www.fedcourt.gov.au/about/national-court-framework/allocations>.

[77] *FCA Act* (n 73) ss 19, 20, 25.

[78] Ibid. Cf. appeals concerning an interlocutory ruling of the FCFCOA: *Parmar v Minister for Immigration and Border Protection* [2018] FCA 502, [10]-[11], [25]; *FCA Act* (n 73) s 25(1A).

Once cases make it onto the Federal Court's docket (by making the requisite filings at a Court registry), the Court's cases are managed by 'individual docket.'[79] This means that at any given time, a case is managed by a single Justice – usually the same Justice from start to finish.[80] Cases are assigned to each Justice indiscriminately (i.e., via assignment in rotation at the time of filing.)[81]

An implication of this docket control system is that it precludes a registry from selecting a distinct subset of cases. The indiscriminate allocation of cases to Justices, in combination with the Court's general inability to choose its docket, makes such selectivity impossible. If such selectivity could occur, it would complicate any empirical analysis of the Federal Court's judgments. Fortunately, that difficulty does not arise.

### 2 *Merits Review of Protection Visa Matters*

Protection visas are available to people who arrive in Australia without a visa but are owed asylum under Australia's international obligations.[82] Applications for protection visas are assessed and determined by a delegate of the Minister administering the *Migration Act 1958* (Cth) ('*Migration Act*').[83]

If a person applies for a protection visa but their application is denied, or their protection visa is cancelled, the person can apply to have that administrative decision 'merits-

[79] Federal Court of Australia, *Allocations of Judicial Matters under the NCF* (n 76).
[80] Ibid.
[81] Federal Court of Australia, 'Glossary of Legal Terms' (21 April 2018) <http://www.fedcourt.gov.au/digital-law-library/glossary-of-legal-terms>. See also Federal Court of Australia, *Allocations of Judicial Matters under the NCF* (n 76).
[82] *Migration Act 1958* (Cth) ss 35A, 36, 36A ('*Migration Act*'); Department of Home Affairs, 'Australia's Protection Obligations', *Immigration and Citizenship* (21 August 2020) <https://immi.homeaffairs.gov.au/what-we-do/refugee-and-humanitarian-program/about-the-program/seek-protection-in-australia/australia-protection-obligations>; Australian Human Rights Commission, '1. What Are Temporary Protection Visas?' <https://humanrights.gov.au/our-work/1-what-are-temporary-protection-visas>. The relevant provisions of the *Migration Act* incorporate article 1(A) of the *1951 Refugees Convention*, to which Australia is a party. See *Convention Relating to the Status of Refugees, Opened for Signature 28 July 1951, 189 UNTS 137 (Entered into Force 22 April 1954)*; Australian Law Reform Commission, *Family Violence and Commonwealth Laws: Improving Legal Frameworks: Final Report* (Final Report No 117, 2011) [22.5]-[22.6] <https://www.alrc.gov.au/wp-content/uploads/2019/08/whole_alrc_117.pdf>.
[83] *AXY17 v Minister for Immigration* [2017] FCCA 2006, [4].

reviewed' by the non-judicial[84] Administrative Appeals Tribunal ('AAT').[85] Upon review, the AAT decides whether the delegate made the most preferable decision permitted by law.[86] If the AAT is satisfied that a better decision could have been made, then the AAT can provide a remedy to the applicant.[87]

### 3 Judicial Review of Protection Visa Matters in the Federal Court

Jurisdiction to conduct judicial review in the Commonwealth jurisdiction is variably vested in the High Court of Australia, Federal Court and FCFCOA.[88] As such, the AAT's decisions concerning protection visas can be *judicially* reviewed in the Federal Court of Australia.[89]

The orthodox understanding of judicial review is that it involves courts reviewing the *legality* of administrative decisions and providing a remedy where such decisions are

---

[84] Importantly, the AAT is administrative, not judicial, in nature. Consequently, the AAT lacks jurisdiction to determine questions of law. See Judith Bannister, Gabrielle Appleby and Anna Olijnyk, *Government Accountability: Australian Administrative Law* (Cambridge University Press, 1st ed, 2015) 308; *Kirk v Industrial Court of New South Wales* (2010) 239 CLR 531, 572.
[85] Bannister, Appleby and Olijnyk (n 84) 326; *AAT Act* (n 71) ss 25, 43. See also *Tribunals Amalgamation Act 2015* (Cth). An administrative decision is an executive exercise of statutory power that serves to confer or affect legal rights or obligations. See *Eastman v Australian Capital Territory* [2014] ACTSC 105, [35], [40]; *Skiba v Commonwealth Ombudsman* [2022] FedCFamC2G 216, [22]-[26]. However, the author notes that can sometimes be ambiguous whether a decision is administrative or not. See *Director-General of Social Services v Hales* (1983) 47 ALR 281, 305–6.
[86] *Drake v Minister for Immigration and Ethnic Affairs* (1979) 24 ALR 577, 591; *Comcare v Wuth* [2017] FCA 433, [80]; *Hutchinson v Comcare* [2018] FCA 505, [73].
[87] *AAT Act* (n 71) s 43. A list of grounds of appeal to the Federal Court and FCFCOA can be found in the *Administrative Appeals (Judicial Review) Act 1975* (Cth) ('*ADJR Act*').
[88] See generally Bannister, Appleby and Olijnyk (n 84) 329. See especially *AAT Act* (n 71) ss 44, 44AAA; *Migration Act* (n 82) s 476A; *ADJR Act* (n 87) ss 4, 8; *Federal Court Rules 2011* (Cth) rr 31.22, 33.12(2) ('*FCR*'); *Constitution* ss 73, 75(v). Judicial review decisions of the FCFCOA can be further appealed to the Federal Court. See *FCA Act* (n 73) s 24(1)(d); *Storry v Weir* (n 74) [1]-[13]; *CVT19 v Minister for Immigration, Citizenship and Multicultural Affairs* [2022] Federal Court of Australia 1482, [5]. The High Court is supreme in its powers of judicial review as it has a constitutionally entrenched 'minimum provision of judicial review' that cannot be ousted by Parliament. See *Constitution* ss 73, 75(v); Lisa Burton Crawford, 'The Entrenched Minimum Provision of Judicial Review and the Limits of Law' 45 *Federal Law Review* 569; *Plaintiff S157/2002 v Commonwealth* (2003) 211 CLR 476, 511–3; *Storry v Weir* (n 74) [27]-[28].
[89] *AAT Act* (n 71) s 44; *FCA Act* (n 73) s 19. See also Marilyn Bromberg and Nicholas Cardaci, 'Playing with Fire: Why Australian Legislators Must Legalise E-Cigarettes' (2021) 24(2) *Quinnipiac Health Law Journal* 125, 150–3. Note, in such cases the defendant is the federal Minister administering the *Migration Act.*

unlawful.[90] This scope of review is necessarily narrow, and so the *merits* of an AAT decision are not to be dealt with by the Federal Court in judicial review.[91]

Grounds for judicial review are set out, relevantly, in the *ADJR Act* – which incorporates grounds of appeal from the common law.[92] Common grounds of appeal include, in relation to the AAT:

a) procedural unfairness;
b) failing to consider relevant material;
c) asking itself the wrong question(s);
d) applying the wrong law or principle of law;
e) considering irrelevant considerations;
f) making factual findings without evidence; and
g) making an unreasonable decision.[93]

### B *Formulating Hypotheses*

Once the law undergirding the dataset's cases is understood, the next step is to formulate two hypotheses that are capable of being *falsified* (i.e., disproven/disconfirmed.)[94] The first is the null hypothesis, which posits that there *is no* relationship between two

---

[90] Robert French, 'United States Influence on the Australian Legal System' (2018) 43(1) *University of Western Australia Law Review* 11, 17–8; *Citta Hobart Pty Ltd v Cawthorn* [2022] HCA 16, [17]-[22]; *Cau v Victorian Building Authority* [2022] FCA 45, [57]. Such remedies include substituting the decision with a new decision, or remitting the decision back to the original decision-maker with directions. See *AAT Act* (n 71) s 43(1).

[91] Creyke (n 71) 630. This scope of judicial review has traditionally been conceived as necessarily narrow by being limited to 'enforcing the law which determines the limits and governs the exercise of the repository's power', rather than determining if there was a more preferable decision on the merits. See *SDCV v Director-General of Security* [2022] HCA 32, [156]; Creyke (n 71) 630. Cf. *AAT Act* (n 71) s 44(7).

[92] *ADJR Act* (n 87) ss 5, 6; *Murphy v Trustees of Catholic Aged Care Sydney* [2019] NSWCATAP 37, [14].

[93] *ADJR Act* (n 87) ss 5, 6; *Murphy v Trustees of Catholic Aged Care Sydney* (n 92) [14].

[94] Beach (n 48) 120–2. See also See MJ Bayarri et al, 'Rejection Odds and Rejection Ratios: A Proposal for Statistical Practice in Testing Hypotheses' (2016) 72 *Journal of Mathematical Psychology* 90, 92; Jacob Cohen, 'Things I Have Learned (So Far)' [1990] *American Psychologist* 9, 1308.

variables.[95] The second is an alternative hypothesis which posits that there *is* a relationship between two variables.[96]

Formulating these falsifiable hypotheses is crucial to conforming to the scientific method. The generally accepted scientific method was the brainchild of Karl Popper.[97] This Popperian method involves formulating hypotheses and then testing them with experimentation that attempts to disprove (i.e., falsify) the hypotheses.[98] Judicially, this Popperian scientific method has seen support from the Supreme Court of the United States in *Daubert v Merrel Dow*.[99]

With consideration paid to the foregoing, for the mock analysis the *question* is whether the State/Territory in which a Federal Court appeal concerning protection visas, is heard, is *significantly* predictive of the outcome. As such:

- the *null* hypothesis is that a trained supervised machine-learning classification algorithm, trained on a dataset of cases containing information on the State/Territory in which a case is heard, *will not be significantly accurate* at predicting case outcomes. Proving the null hypothesis would support the idea that the rule of law is strong in the Federal Court; and
- the *alternative hypothesis* is simply that the supervised machine-learning classification algorithm *will be significantly accurate* at predicting case outcomes. Proving the null hypothesis supports the idea that legal realism is strong in the Federal Court, to the detriment of the rule of law.

The author emphasises that the alternative hypothesis is only supported by the mock analysis' results if the trained algorithm is accurate to a *significant* degree.[100] Significance testing is a general requirement in science, and there are different ways of determining

---

95 Rebecca Bevans, 'A Step-by-Step Guide to Hypothesis Testing', *Scribbr* (29 October 2021) <https://www.scribbr.com/statistics/hypothesis-testing/>. See also Cox (n 7) 325.
96 Bevans (n 95). To illustrate, an alternate hypothesis is that Court Z *does* decide similar cases differently from Court X, while the corresponding null hypothesis is that Court Z does *not*. Theoretically, quantitatively analysing the judgments from these courts could falsify either of these hypotheses – so they are valid hypotheses.
97 Beach (n 48) 120–2.
98 Paul van Helden, 'Data-Driven Hypotheses' (2013) 14(2) *EMBO reports* 104, 104; Cohen (n 94) 1308; Beach (n 48) 120–2.
99 Beach (n 48) 120–2.
100 See Cox (n 7) 325; Beach (n 48) 146–7. See especially Katz, Bommarito and Blackman (n 14) 9–10.

significance.[101] For present purposes, one simple way is to use a heuristic prediction method ('HPM') to predict that unseen cases (i.e., cases in the testing set) will all have the outcome that was most common in past cases (i.e., in the training set's cases).[102] Practically speaking, the HPM sets a benchmark that the trained algorithm must surpass in order for the alternative hypothesis to be confirmed.[103] As such, for the mock analysis, the trained algorithm will only be considered *accurate* if it is more accurate than the HPM's prediction.[104]

At the outset, the author notes that the null hypothesis is consistent with the rule of law, while the alternative hypothesis is not. As such, this mock analysis can be considered a (mock) divergence-hunting inquiry into the true extent to which the rule of law operates. The author acknowledges their expectation that the null hypothesis for the mock analysis is, prima facie, unlikely to be disconfirmed by the mock analysis.[105]

### C *Data Collection and Preparation*

After the hypotheses have been formulated, the next steps are to:

- formulate inclusion criteria to demarcate what *exactly* is to be included in the dataset;
- collect the data with appropriate methods and in accordance with the inclusion criteria;.
- prepare the data for analysis through quality assurance; and
- understand and describe the data with descriptive statistics.

---

[101] Cox (n 7) 327. See generally Cohen (n 94).

[102] Conor O'Sullivan and Joeran 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) 8–10 <https://arxiv.org/ftp/arxiv/papers/1912/1912.10819.pdf>; Katz, Bommarito and Blackman (n 14) 9–10, 13.

[103] O'Sullivan and Beel (n 102) 8–10; Katz, Bommarito and Blackman (n 14) 9–10, 13.

[104] O'Sullivan and Beel (n 102) 8–10; Katz, Bommarito and Blackman (n 14) 9–10, 13. To illustrate, if a training set has 90% of its cases as appeals being denied, the HPM consequently predicts that *all* testing set cases will be appeals being denied (because appeals being denied was the most common outcome.) If the testing set has appealed being denied in 55% of its cases; therefore the HPM is 55% accurate. This means that the trained algorithm has to be over 55% accurate in order for the alternative hypothesis to be confirmed. See

[105] See earlier in this article at Part II(B Australia's Legal (Un)Realism

These steps are now explained in turn.

### 1 *Inclusion Criteria*

Generally, inclusion criteria dictate what characteristics a case must have in order for it to be included in the dataset.[106] Inclusion criteria are very useful for giving certainty to researchers and their audiences in regard to exactly what data is relevant and thus forms part of the dataset.

Appropriate inclusion criteria vary depending on the exact analysis. Nonetheless, researchers generally should consider using inclusion criteria based on legal area, forum, recency, legal representation and applicable law – as will be shown.

#### *(a) Using Inclusion Criteria to Ensure a Dataset's Homogeneity*

It is a truism that no two cases are identical – especially in the Federal Court.[107] Nonetheless, in the current context, it is crucial to strive to have a dataset that only includes cases that are qualitatively homogeneous (in terms of factors that could affect a case's outcome.)[108] This is to ensure that apples are only being compared to other apples and not oranges.[109]

Indeed, the importance of using a homogenous dataset when analysing the Federal Court's work has been stressed in a public statement of the Court, in response to a statistical analysis of the Court's work (which was commissioned and published by a newspaper):

---

[106] Nicholas Cardaci, 'Costs Orders in Federal Court Migration Litigation: An Empirical Analysis' (2018) 44(1) *University of Western Australia Law Review* 172, 177–8, 184.
[107] See especially 'Federal Court Response to AFR Ranking of Judges', *Australian Financial Review* (online, 25 October 2018) <https://www.afr.com/companies/professional-services/federal-court-response-to-afr-ranking-of-judges-20181024-h171nv>; *Palmer v McGowan (No 6)* (n 32) [6]. See generally Brendan Sweeney, Mark Bender and Nadine Courmadias (n 35) 20 [1.32]; *DPP (Cth) v D'Alessandro* [2010] VSCA 60, [40]; *DPP v Rongonui* [2007] Supreme Court of Victoria - Court of Appeal 274, [45].
[108] See Medvedeva et al (n 15) 4–5; *R v ACN* [2018] EWCA Crim 1507, [11]-[15]. See also Kevin M Clermont and Theodore Eisenberg, 'Xenophilia or Xenophobia in U.S. Courts? Before and After 9/11' (2007) 4(2) *Journal of Empirical Legal Studies* 441, 443 ('Xenophilia or Xenophobia in U.S. Courts?').
[109] *Sully v CBMG North Pty Ltd [2020] FWC 3509* [77]-[78]. See also *Hungry Jacks Pty Ltd v City of Bayswater [2013] WASC 199* [21]-[24].

> [T]hese statistics (and indeed any mere collection of numbers of judgments) say (and says) nothing about the varied character, difficulty and nature of the work of the Court.
>
> Simple metrics of numbers of judges, of judgments, and of arithmetically-derived time and page production are meaningless. This is particularly so in a Court of complex and widely-varied jurisdiction (at both first instance and appellate level) involving often complex matters concerning commercial law; intellectual property (patents, trade marks and copyright); taxation; native title; industrial and employment law; shipping, Admiralty and maritime law; Constitutional and administrative law; and other miscellaneous federal jurisdictions such as defamation; as well as a high volume of (often, though not always) less complex appellate work, in particular migration appellate work.
>
> The Court's work is not homogeneous and of a repetitive character. In this wide variety of work, including jurisdictions of significant speciality, the Court delivers judgments each year in a number that varies but is from 1,600 to 2,000 per annum.[110]

The methodological cruciality of a case database being homogenous is also reflected in jurisprudence which stresses the importance of using homogenous cases when conducting statistical analysis of prior cases' sentencing or penalties.[111]

### *(b) Inclusion Criteria for the Mock Analysis*

Given the foregoing considerations, the inclusion criteria ('Criteria') used for assembling the mock analysis' raw dataset (i.e., filtering out the unwanted Federal Court cases out) are as follows:

1) the case concerns a Federal Court's judicial review, between 2012 to 2018, of an administrative decision, merits review decision, or judicial review decision

---

[110] 'Federal Court Response to AFR Ranking of Judges' (n 107).
[111] *Scherini v Cleveland Freightlines Pty Ltd* [2018] WASC 5, [161]-[169]; *C E Oates & Sons Pty Ltd t/a Narrogin Retravision v Balla* [2015] WASC 144, [109]-[112]; *Hurt v The Queen* (n 32) [134]-[136]; *Dragon Pacific Group Pty Ltd v City of Cockburn [2019] WASC 449* [24]-[27]; *Radianct Holdings (Australia) Pty Ltd v City of Gosnells [2022] WASC 217* [21]-[24]; *Hungry Jacks Pty Ltd v City of Bayswater [2013] WASC 199* (n 109) [21]-[24]. See also National Judicial College of Australia, 'Consistency in Federal Sentencing' (5 February 2015) 3.2 <https://csd.njca.com.au/principles-practice/consistency-in-federal-sentencing/>.

concerning the cancellation or refusal to grant a protection visa (including refusals to grant leave to appeal or extend time to file an appeal);[112] and

2) the case was an appeal from a decision of the FCFCOA, AAT or the Minister administering the *Migration Act*; and
3) the defendant was the Minister administering the *Migration Act*; and
4) the visa applicant was legally represented; and
5) the appeal was decided by a single Justice of the Federal Court; and
6) the judgment contained catchwords.

These inclusion criteria, like any other set of inclusion criteria, need to be operationalised. To do this, a researcher must identify the part of the judgments containing the relevant (to the criteria) information and have tools capable of detecting and extracting that information.[113] So, if the criterion is that X variable has Y value, then you need to (A) identify where X is in the judgments, (B) interpret the information given at X in order to determine if X is indeed Y.[114] For example, if you only want cases where the type of law (X) is migration law (Y), you must know (a) where the judgments indicate what X is, and (b) be able to interpret whether the text provided at that location indicates that X is Y.[115]

The inclusion criteria for the mock analysis were operationalised as follows.

a) The catchwords contained 'protection visa'.
b) The judgment lists a judgment below (i.e., 'appeal from: …') and/or the catchwords indicated that the judgment concerns an appeal.
c) The listed respondent was the relevant federal Minister.
d) The 'Solicitors for the Applicant' section in the judgment lists a representing lawyer, instead of 'Appeared in person' or some variation of that.
e) No judgments with 'FCAFC' in their medium neutral citation.

[112] It is appropriate include cases about granting leave because they still involve the Court considering merits of the underlying application to some degree. See *Tu'uta Katoa v Minister for Immigration, Citizenship, Migrant Services and Multicultural Affairs* [2022] HCA 28, [12]-[20], [53]-[65]; *BZAHM v Minister for Immigration and Border Protection* [2015] FCA 675, [39]-[42]; Medvedeva et al (n 15) 4–5.
[113] Medvedeva, Vols and Wieling (n 12) 245; Theodore Eisenberg and Charlotte Lanvers, 'What Is the Settlement Rate and Why Should We Care?' 37, 127; Masha Medvedeva (n 53) 47, 87–8.
[114] Medvedeva, Vols and Wieling (n 12) 245; Eisenberg and Lanvers (n 113) 127; Masha Medvedeva (n 53) 47, 87–8.
[115] Medvedeva, Vols and Wieling (n 12) 245; Eisenberg and Lanvers (n 113) 127; Masha Medvedeva (n 53) 47, 87–8.

f) The Script searched for and recorded catchwords for each case.

The author now explains the rationale for each of the mock analysis' inclusion criterion. This is good practice because it helps a researcher and their audience be certain that the relevant dataset isn't being constructed arbitrarily.

The following explanations are also intended to assist future researchers in forming their own inclusion criteria for various types of case-related research.

### *(i) Legal Homogeneity*

Ensuring any dataset is *legally* homogeneous is important because the dataset becomes less relevant to a hypothesis to the extent that it includes cases that concern irrelevant legal issues and underlying law.

One element of ensuring legal homogeneity is ensuring all cases relate to the *same legal area*.[116] Criteria 1, 2 and 3 all ensure that the cases in the database all deal with the same subject matter – protection visas.

Legal homogeneity also requires that the applicable law be materially identical over the time in which the dataset's cases were decided.[117] Criterion 1 helps to ensure this as well by limiting the relevant time period. The author is certain that the relevant law has remained materially unchanged throughout the relevant period (2012-2018) after examining the amendment history of the essential statutory provision, which is *Migration Act* s 36.[118]

As a side note, protection visa cases are good cases for ELR tests searching for the potential operation of legal realism. This is because the grant of asylum has been a

[116] Medvedeva et al (n 15) 1–2, 9–10.
[117] A single dataset should not incorporate cases from both before and after a major legal development. See Ibid. To illustrate, if there had been a major rewrite of s 36 in 2015, the pre-rewrite case law may not be applicable to the new provision and thus should not be included in the dataset.
[118] The relevant statutory provision (*Migration Act* s 36) was only amended twice in 2014; and the amendments didn't relevantly change the provision in the author's view. One amending statute made non-substantive changes to the provision, and another added a requirement that the applicant is not assessed by ASIO to be a security risk. See respectively *Migration Amendment Act 2014* (Cth) sch 3 cl 1; *Maritime Powers Legislation Amendment (Resolving the Asylum Legacy Caseload) Act 2014* (Cth) sch 2 cls 6, 7, 8, 9.

contentious moral and political issue in Australia, and cases about such contentious subjects are ones in which judicial personality is more likely to express itself.[119]

### *(ii) Relatively Meritorious Cases*

It is important to exclude unmeritorious (i.e., 'without merit')[120] cases from a dataset, as they are qualitatively different from meritorious cases. This is important because if unmeritorious cases are included in a dataset alongside meritorious cases, this inclusion can skew a subsequent analysis' result by giving a result that suggests meritorious cases have a lower prospect of success than they actually do, or vice versa.

For the purposes of this article, unmeritorious cases are judicial review cases in which the appellant argues grounds that defy common sense[121] and/or cannot reasonably point to an appealable error by the earlier decision maker.[122]

Criterion 4's exclusion of cases brought by self-represented litigants is an imperfect and admittedly crude attempt to filter out unmeritorious cases and conversely include those that are *relatively* meritorious.[123] Despite its imperfection, there are several reasons why Criterion 4 serves as this filter.

First, several laws discourage *lawyers* from representing applicants with unmeritorious cases. Specifically, there are (a) professional ethics rules that discourage lawyers from wasting courts' time by arguing unmeritorious cases,[124] and (b) adverse costs orders that

---

[119119] See Cross (n 16) 133–4, 144; Janet Phillips and Harriet Spinks, 'Boat Arrivals in Australia since 1976', *Parliament of Australia* <https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BN/2011-2012/BoatArrivals>. See especially Thomas (n 26).

[120] *Sawtell v P J Clarke Investments (Qld) Pty Ltd* [2019] FCA 385, [18]; *BUK16 v Minister for Immigration and Border Protection* [2020] FCA 558, [53]-[55].

[121] *AXY17 v Minister for Immigration* (n 83) [42].

[122] *Sawtell v P. J. Clarke Investments (Qld) Pty Ltd* (n 120) [15]-[16].

[123] Though of course it is still *possible* for represented parties to have their lawyers spout inarguable legal theories in court. See *Neitzke v Williams* (1989) 490 US 319, 328; *Ennis v Credit Union Australia* [2016] FCCA 1705, [19]-[21]; *AXY17 v Minister for Immigration* (n 83). Such cases could be inadvertently included in the dataset under Criterion 4. Nonetheless, this is unlikely for the reasons given in this subheading of the article.

[124] *Legal Profession Uniform Law Australian Solicitors' Conduct Rules 2015* (NSW) rr 3, 4, 17; *Legal Profession Uniform Conduct (Barristers) Rules 2015* (NSW) r 23; WA Bar Association, 'Western Australian Barristers' Rules (Amended as at 23 February 2017)' [25] <https://wabar.asn.au/wp-content/uploads/2018/10/Western-Australian-Barristers-Rules-23-

may be awardable against lawyers that represent clients with unmeritorious cases[125] (precisely to *discourage* such cases.) [126] Overall, these legal rules, which the relevant lawyers would *certainly* be aware of, would effectively deter lawyers from taking on unmeritorious cases.[127]

<u>Second</u>, self-represented litigants, who are not paying for lawyers, lack the same economic incentive to refrain from bringing meritless legal actions, as they aren't wasting any money on lawyers to bring such cases.[128] Indeed, the Supreme Court of the United States has noted that 'a litigant whose filing fees and court costs are assumed by the public, unlike a paying litigant, lacks an economic incentive to refrain from filing frivolous, malicious, or repetitive lawsuits.'[129]

<u>Third</u> (and conversely), legally represented litigants do not generally initiate unmeritorious lawsuits because they are aware of the expenses of litigating, namely legal fees and potential adverse costs orders (which lawyers must inform their clients about before litigation.)[130]

---

February-2017.pdf>; *Guneser v Aitken Partners* (n 33) [53]; D.A. Ipp, 'Lawyers' Duties to the Court' (1998) 114 *Law Quarterly Review* 63, 80, 99–100. Cf. Robert Mazza, 'Ethical Issues for Defence Counsel on a Plea of Guilty' (March) 14, 16; D.A. Ipp 85–6.

[125] *Migration Act* (n 82) ss 486E, 486F(1)(c); *BUK16 v Minister for Immigration and Border Protection* (n 120) [58], [65]-[75]; *AXY17 v Minister for Immigration* (n 83) [42]-[56].

[126] 'Parliament intended to discourage persons from encouraging others to make and continue unmeritorious applications in migration cases', subject to public interest considerations, according to *BUK16 v Minister for Immigration and Border Protection* (n 120) [55] (Charlesworth J).

[127] See e.g., *CVT19 v Minister for Immigration, Citizenship and Multicultural Affairs* (n 88) [7]. Cf. *BUK16 v Minister for Immigration and Border Protection* (n 120).

[128] *Neitzke v Williams* (n 123) 324.

[129] Ibid 325. See also Cardaci (n 106) 186.

[130]Masha Medvedeva, Michel Vols and Martijn Wieling, 'Judicial Decisions of the European Court of Human Rights: Looking into the Crystal Ball' in *Proceedings of the Conference on Empirical Legal Studies in Europe* (2018) 8 <https://research.rug.nl/en/publications/judicial-decisions-of-the-european-court-of-human-rights-looking->; Katz, Bommarito and Blackman (n 14) 14; Masha Medvedeva (n 53) 90.

Masha Medvedeva, Michel Vols and Martijn Wieling, 'Judicial Decisions of the European Court of Human Rights: Looking into the Crystal Ball' in *Proceedings of the Conference on Empirical Legal Studies in Europe* (2018) 8 <https://research.rug.nl/en/publications/judicial-decisions-of-the-european-court-of-human-rights-looking->; Katz, Bommarito and Blackman (n 14) 14; Masha Medvedeva (n 53) 90.

Clients would be aware of the potential litigation costs because lawyers in Australia must disclose such costs to their clients. See Cardaci (n 106) 184–5; *Legal Profession Uniform Law 2014* (NSW) Part 4.3 Div 3.

Fourth, there is judicial recognition that 'there has been an unrelenting stream of applications challenging decisions made under the [Migration] Act. Those applications are often brought by litigants in person who, in many cases, have been given assistance by others. History reveals that many of those applications are unmeritorious and are doomed to fail.'[131]

Overall, these realities further support the use of Criterion 4 as a means to ensure homogeneity in the dataset – by helping to ensure that all the dataset's cases are at least relatively meritorious.

With these four foregoing reasons in mind, applying Criterion 4 means the resulting cases (all of which had legally represented applicants) can be reasonably expected to have a level of merit that (a) lawyers would be willing to argue it, and (b) their clients would be willing to pay to have argued in court.

### *(iii) Avoiding Disadvantaged Litigants*

Criterion 4 also serves to exclude cases in which there is a significant disparity between the advocacy abilities of the litigants – which is another factor that could affect the outcome of cases.[132]

Self-represented litigants typically face a significant disadvantage in conducting court proceedings, *especially* when facing off against a represented party such as a Commonwealth Minister.[133] Further, courts can only provide very limited assistance to self-represented litigants (i.e. to reduce unfairness) during proceedings.[134]

In contrast, where both parties are represented, there should be less disparity in legal and advocacy skills – giving both parties a fair go at conducting their case. As such, to ensure homogeneity, only these cases are included as per Criterion 4.

---

[131] *SZFDZ v Minister for Immigration and Multicultural Affairs* [2006] FCA 1366, [26]; *BUK16 v Minister for Immigration and Border Protection* (n 120) [54].
[132] Cane and Kritzer (n 8) 511–3.
[133] *Stewart v Hames* [2021] WADC 93, 224, [65]; *Storry v Weir* (n 74) [1]-[13], [27]-[33].
[134] *Stewart v Hames* (n 133) [65].

*(iv) Eliminating Panel Effects*

Criterion 5 seeks to avoid the confounding influence of any ‘panel effects’, which are theorised effects on judges that cause them to decide cases differently (when co-judging) as compared to when they decide cases alone.[135] Notably, one alleged panel effect is that, when sitting on a panel of judges, there is a greater tendency for each judge to conform with the opinions of their co-judges.[136]

As such, excluding Full Federal Court judgments (as per Criterion 5) means that the theoretically confounding panel effects cannot be present in a subset the dataset’s cases, and thus panel effects cannot affect that subset’s outcomes.

*(v) Convenience for Quality Assurance*

Criterion 6 was for convenience – by making it easier to perform quality assurance on the data. Having the catchwords allowed the author to quickly confirm if each case’s information (collected by the Script) was accurate.

2 *Collecting a Raw Dataset*

To analyse a dataset of prior judgments that fit the inclusion criteria, you obviously need to acquire a dataset of prior judgments. The author emphasises that the dataset must be sufficiently large for it to be used as a dataset for machine learning.[137] Crudely speaking,

[135] Thomas J. Miles and Cass R. Sunstein (n 15) 837–9, 843, 846, 848; Kastellec (n 19) 169–70. Such conformity is suggested (in the literature) as reflecting the general human tendency and desire to conform with the views of others – to such an extent that humans are prone to changing their beliefs in the face of unanimous opposition. See Thomas J. Miles and Cass R. Sunstein (n 15) 839.

[136] Thomas J. Miles and Cass R. Sunstein (n 15) 837–9, 843, 846, 848; Kastellec (n 19) 169–170. Some evidence suggests this panel occurs in U.S. Circuit Court, with its politically ‘liberal’ judges tending to rule less ‘liberally’ when co-judging with ‘conservative’ judges, and vice-versa. See Thomas J. Miles and Cass R. Sunstein (n 15) 837–9, 843, 846, 848; Kastellec (n 19) 169.

[137] Masha Medvedeva, Michel Vols and Martijn Wieling, ‘Judicial Decisions of the European Court of Human Rights: Looking into the Crystal Ball’ in *Proceedings of the Conference on Empirical Legal Studies in Europe* (2018) 8 <https://research.rug.nl/en/publications/judicial-decisions-of-the-european-court-of-human-rights-looking->; Katz, Bommarito and Blackman (n 14) 14; Masha Medvedeva (n 53) 90. See also Cui et al (n 48) e129–e130.

the more data that is available for the training phase, the better the ultimate analysis' results will be.[138] On this point, Cui et al (2020) explain that:

> Applying a small dataset to train a complex algorithm can be problematic, as it may lead to overfitting pitfalls, where the complex algorithm starts to fit noise or errors in the limited-size training set, in other words, the algorithm memorizes the data rather than learns from it. Under this circumstance, generalization of the model is usually not good, that is, the model performs poorly on new, unseen out-of-sample datasets.[139]

Acquiring a dataset can be done most simply by using an existing suitable dataset.[140] However, the desired dataset may not exist already. If so, researchers likely must use computerised techniques, in accordance with the chosen inclusion criteria, to collect the relevant case law from the internet, and then extract the relevant information into a usable format.[141] Sometimes this entails creating and using bespoke scripts or software tools for downloading judgments from internet repositories (if permissible by the database provider) and extracting the judgments' relevant information into a usable format (e.g., a CSV file or plain text files.)[142] Such bespoke tools can be created using the popular Python programming language. In the author's experience, this is a worthy choice as it is a flexible language with powerful third-party libraries such as *Requests* and *BeautifulSoup*. The in-built *Regex* library is also very useful for parsing data pulled from the aforementioned databases.[143]

It should be noted that bespoke tools will rarely be perfect, and may not collect all the judgments,[144] or information therein,[145] that a tool's creator desires. Notwithstanding, if

[138] Medvedeva et al (n 15) 247; Masha Medvedeva (n 53) 90.
[139] Cui et al (n 48) e129–e130.
[140] See e.g., the CASELAW4 database described in Alina Petrova, John Armour and Thomas Lukasiewicz, 'Extracting Outcomes from Appellate Decisions in US State Courts' in Serena Villata, Jakub Harašta and Petr Křemen (eds), *Frontiers in Artificial Intelligence and Applications* (IOS Press, 2020) <http://ebooks.iospress.nl/doi/10.3233/FAIA200857>.
[141] Medvedeva, Vols and Wieling (n 137) 3; Masha Medvedeva (n 53) 10. See e.g., Medvedeva, Vols and Wieling (n 12) 246.
[142] Medvedeva, Vols and Wieling (n 12) 245–6; Cardaci (n 106) 177–9; O'Sullivan and Beel (n 102) 3; Benjamin Strickson and Beatriz De La Iglesia, 'Legal Judgement Prediction for UK Courts' in *Proceedings of the 2020 The 3rd International Conference on Information Science and System* (ACM, 2020) 204, 205 <https://dl.acm.org/doi/10.1145/3388176.3388183>.
[143] Strickson and De La Iglesia (n 142) 3.2.
[144] See e.g., Medvedeva, Vols and Wieling (n 12) 246.
[145] See e.g., row 3 of Figure 2 in this article, where the identity the presiding judge is missing due to the Script not collecting it for that case.

the dataset has *enough* judgments[146] with all the information (e.g., textual features) required for subsequent analysis, that is sufficient for machine learning analysis.[147]

### *(a) Documenting the Data Collection Method*

The methods by which the data was collected must be recorded by researchers in sufficient detail. The record should include *how* and *when* the collection was done and *who* did the collection, and the *number* of items collected.[148] This requirement is analogous to the requirement for medical literature reviews to detail their searches of existing literature.[149]

In the author's view, an instance of good practice in documenting a data collection method is Creyke (2017). That article  helpfully provides the details of that author's case law searches (on AustLII) including the exact search terms, method, search date, and number of results.[150]

### *(b) Collection for the Mock Analysis*

For the mock analysis the initial collection of judgments was done using a Python 3 script ('Script') in February 2022. The Script downloaded all single-judge Federal Court judgments dated from 2012 to 2018 (inclusive). For each of these downloaded judgments. To 'clean' the judgments' text (for ease of future use), it was manipulated using Python's regex library, to make all the text lowercase, and, replace all whitespace with single space characters.[151]

---

[146] Medvedeva, Vols and Wieling (n 12) 246.
[147] Masha Medvedeva (n 53) 90. The authors note that there are also ways of 'imputing' missing data. See '6.4. Imputation of Missing Values', *scikit-learn* <https://scikit-learn/stable/modules/impute.html>.
[148] See e.g., Cardaci (n 106) 177–9.
[149] See e.g., Lauren P Manning, Caroline J Tuck and Jessica R Biesiekierski, 'The Lived Experience of Irritable Bowel Syndrome: A Focus on Dietary Management' (2022) 51(6) *Australian Journal of General Practice* 395, 396.
[150] Creyke (n 71) 631, 632, 638, 647.
[151] Similar to O'Sullivan and Beel (n 102) 5; Strickson and De La Iglesia (n 142) 3.3.

The Script extracted several *features* using regular expressions ('regex') from each judgement's plain text. Features are the qualitative attributes of each case, and are a necessary part of data collection. The features are:

a. judge's name.
b. first part of the judgment's catchwords.
c. A Python re.match value recording whether the judgment contains text indicating that the original decision was affirmed (i.e., appeal refused).
d. A Python re.match value recording whether the judgment contains text indicating that an order was granted in favour of the applicant, (i.e., appeal allowed.)
e. Registry in which the case was heard.
f. Text recording the parties' legal representation in the matter.
g. The catchwords for the judgment.

2. These features were then passed to the raw dataset's data structure. The data structure for this mock analysis was a list of dictionaries with the case name as a key and the regex-extracted features as the values. This data structure is shown below and the author suggests that this data structure is a good example for readers to employ in their own research.

*Figure 1.Data Structure of Example Dataset*

case_dictlist = list()[152]

case_dict = {}[153]

[152] This is a list containing all the *case_dict* dictionaries.
[153] This is a dictionary created for each case in the dataset. The key is a string containing the name of the case.

> case_dict[str(title)] = [judge_string,[154] str(case_matter),[155] str(affirm_search),[156] str(allow_search),[157] registry,[158] rep,[159] catchwords[160]]

After all the prior judgments' features were collected, they were saved in a CSV file, and opened in Excel (after being converted to an .xlsx file.) In Excel, the inclusion criteria were semi-manually applied by the author to isolate the cases that fit those inclusion criteria. The cases that fit the criteria formed the resulting *raw dataset* of 145 unreported judgments from the Federal Court. A short extract from the raw dataset, as displayed in Excel, is displayed below to assist the reader in visualising the data.

[154] A string containing the judges' names.
[155] The first part of the judgment's *catchwords*.
[156] A Python re.match value recording whether the judgment contains text indicating that the decision on appeal was affirmed (i.e., appeal refused). Records 'None' where no such text is found.
[157] A Python re.match value recording whether the judgment contains text indicating that an order was granted in favour of the applicant, e.g., matter remitted, appeal allowed. Mutually exclusive with str(*affirm_search*). Records 'None' where no such text is found.
[158] Registry in which the case was heard.
[159] Text recording the legal representation in the matter.
[160] The catchwords for the judgment.

Figure 2. Extract of Original Data

| Case Title | Judge | Case Matter | Plaintiff fail | Plaintiff success | Registry | Appearance / Representation | All Catchwords | Case below |
|---|---|---|---|---|---|---|---|---|
| **SZQZK v Minister for Immigration and Citizenship [2012] FCA 1229** | robertson j | migration | <re.Match object; span=(2809, 2821), match='be dismissed'> | None | sydney | solicitor for the appellant: allens | migration – appeal from federal magistrates court – judicial review of recommendation of independent merits reviewer to refuse protection visa – procedural fairness – whether failure to put adverse conclusions to appellant – whether conclusions not obviously open on known material – whether material credible, relevant and significant – discretion to grant relief – whether separate or independent finding to sustain reviewer's determination | szqzk v minister for immigration & anor [2012] fmca 490 |
| **ANK15 v Minister for Immigration and Border Protection [2017] FCA 1493** | dowsett j | migration | <re.Match object; span=(1154, 1170), match='appeal dismissed'> | None | victoria | solicitor for the appellant asylum seeker resource centre | migration – appeal from the federal circuit court – decision to refuse a protection visa – where the applicant did not attend a hearing before the administrative appeals tribunal due to alleged illness – where the tribunal made a decision to refuse his application in his absence – appeal dismissed | ank15 v minister for immigration & anor [2017] fcca 1269 |
| **CJT15 v Minister for Immigration and Border Protection [2018] FCA 618** | | migration | <re.Match object; span=(1649, 1665), match='appeal dismissed'> | None | victoria | solicitor for the appellant vrachnas and co | migration – refusal of a protection visa application pursuant to s 36 of the migration act 1958 (cth) and cl 866.221 of schedule 2 of the migration regulations 1994 (cth) — whether the primary judge erred in refusing to grant an adjournment of the hearing — whether the tribunal erred in failing to consider whether the appellant may suffer persecution, serious harm or significant harm in detention by reason of general, random or ubiquitous violence by the authorities — whether the tribunal fell into jurisdictional error because there was no logically probative evidence to suggest that the appellant's family would be willing and able to act as guarantors for the appellant's release on bail — leave to appeal granted in part — appeal dismissed | application for leave to appeal: cjt15 v minister for immigration and border protection & anor [2017] fcca 1039 |
| **ASD16 v Minister for Immigration and Border Protection [2018] FCA 1165** | tracey j | migration | <re.Match object; span=(4151, 4163), match='be dismissed'> | None | victoria | solicitor for the appellant: ambi | migration – appeal from a judgment of the federal circuit court ("the fcc") – where the fcc had dismissed an application for judicial review of a decision of the refugee review tribunal – where the tribunal had affirmed a decision of a delegate of the minister for immigration and border protection to refuse to grant the appellant a protection visa – whether the fcc had erred in failing to find that the tribunal had committed a jurisdictional error – whether the tribunal had erred by failing to consider whether, upon his return to sri lanka, a member of the appellant's family would be able to act as guarantor so that he would be released on bail – whether the tribunal had failed to consider the risk to the appellant of being followed up and suffering torture in his home area as part of the general pattern of such abuse – whether the tribunal had misapplied the tests for whether there was a "real chance" of persecution or a "real risk" of significant harm | asd16 v minister for immigration and border protection [2016] fcca 3091 |

### 3 *Quality Assurance*

As the adage goes, 'garbage in, garbage out.'[161] This means that low-quality data will yield low-quality (i.e., unrealistic or inaccurate) predictions if a machine learning algorithm is trained on that data. Given this, the raw dataset must be subjected to quality assurance ('QA'). Using a high-quality dataset is essential for any machine learning algorithm to accurately predict reality. Data quality requires *inter alia* that the dataset be (a) 'accurate' in reflecting the source material, and (b) 'reliable' by consistently containing all features, representing things (e.g., case names) consistently.[162]

Given the need for a high-quality dataset, researchers should undertake QA of their raw dataset. This involves manual and/or semi-automatic procedures to check the dataset for inaccurate or unreliable data. Generally, data that is missing, incomprehensible, or incorrect should be located and either corrected or removed as part of QA.[163] It is equally crucial that researchers record and report (in any written findings) how QA was performed.[164]

#### *(a) QA for the Mock Analysis*

For the mock analysis, the author has undertaken a QA of the raw dataset. The following QA checks were undertaken to ensure the accuracy and reliability of the dataset that would ultimately be used for training a supervised machine learning algorithm:

1) Converting regex output strings to Boolean values.
2) Ensuring there are no missing values for *catchwords*, *registry* and *rep*.
3) Ensuring that only appeals of protection visas were included by manually reviewing the catchwords <u>of all cases</u> in the raw dataset for any unusual cases that contained 'protection visa' as well as either 'judicial review' or 'appeal', but weren't actually appeals of protection visa decisions.
4) Ensuring the verdicts for each case are correct by manually reviewing the judgment where the values in each case's *affirm_search* and *allow_search* values are both empty (i.e., the Script could not determine the result) and updating these values according to the (a) original judgment, or (b) *catchwords* where possible.
5) the author manually added and populated a column titled *State* to the raw dataset. This column's values to indicate which jurisdiction the registry that heard each case was located in (e.g., cases in the Sydney registry were given values of 'NSW' in the *State* column.)

---

161 'Garbage in, Garbage Out', *TheFreeDictionary.com* <https://idioms.thefreedictionary.com/garbage+in%2c+garbage+out>.
162 Cui et al (n 48) e144.
163 Strickson and De La Iglesia (n 142) 3.2.
164 See e.g., Ibid.

The result of these QA checks was the removal of three cases as a correction to the dataset. This is because because QA check 3 revealed these cases were irrelevant (i.e., unrelated to the requisite subject matter.)[165]

### 4 *Describing a Finalised Dataset*

Undertaking QA checks and implementing the required corrections results in a *finalised* dataset. As such, for the mock analysis, the QA checks resulted in a finalised dataset of 142 cases (after the removal of the three aforementioned irrelevant cases).

Generally, before a researcher begins experimenting on the finalised dataset, it must first be descriptively understood. This can be done in the first instance by providing descriptive statistics about the finalised dataset,[166] including but not limited to:

a) number of cases;
b) date range of cases;
c) verdict(s) issued;
d) type(s) of cases;
e) location of cases; and
f) presiding judge(s).[167]

Describing a dataset provides further context (and potentially insights) to the researchers and ultimate readers of the findings. The prior literature and the author's experience reveal that tables and graphs can be very useful in understanding a case law dataset.[168] As such it is recommended that researchers use visualisation tools for the benefit of themselves and the ultimate readers of the research. As such, the following two tables provide a visual breakdown of the finalised dataset of 142 cases.

---

[165] *MZYWC v Minister for Immigration and Citizenship* [2012] FCA 1457; *DKB18 v Minister for Home Affairs* [2018] FCA 1465; *Steyn v Minister for Immigration and Border Protection* [2017] FCA 1131.
[166] Medvedeva, Vols and Wieling (n 137) 11, 90–1; Frans Leeuw and Hans Schmeet, *Empirical Legal Research* (Edward Elgar Publishing, 2016) 166 <http://www.elgaronline.com/view/9781782549390.xml>.
[167] Cardaci (n 106) 6–9; Medvedeva, Vols and Wieling (n 12) 246–7.
[168] See generally Cohen (n 94) 1305. See e.g., Cardaci (n 106) 180; Medvedeva, Vols and Wieling (n 12) 246–7; Danziger, Levav and Avnaim-Pesso (n 9) 6890.

*Figure 3. Finalised Dataset Cases Grouped by State*

| State | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Grand Total of Cases |
|---|---|---|---|---|---|---|---|---|
| **NSW** | 6 | 5 | 10 | 11 | 12 | 6 | 25 | 75 |
| **NT** | 2 | | | | | | | 2 |
| **QLD** | 6 | 1 | | 5 | 1 | 1 | 2 | 16 |
| **SA** | | 1 | 1 | 1 | 3 | 2 | 2 | 10 |
| **VIC** | 2 | | 3 | 2 | 6 | 5 | 15 | 33 |
| **WA** | | 1 | | 1 | | | 4 | 6 |
| **Grand Total** | **16** | **8** | **14** | **20** | **22** | **14** | **48** | **142** |

*Figure 4. Finalised Dataset Cases Grouped by Plaintiff Result*

| Plaintiff Result | State | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Grand Total of Cases |
|---|---|---|---|---|---|---|---|---|---|
| Failed | **NSW** | 6 | 4 | 9 | 11 | 11 | 6 | 19 | 66 |
| | **NT** | | | | | | | | |
| | **QLD** | 6 | 1 | | 3 | 1 | 1 | 2 | 14 |
| | **SA** | | | 1 | | 1 | 1 | 2 | 5 |
| | **VIC** | 2 | | 2 | 2 | 4 | 3 | 10 | 23 |
| | **WA** | | 1 | | 1 | | | 4 | 6 |
| **Failed Total** | **All States** | **14** | **6** | **12** | **17** | **17** | **11** | **37** | **114** |
| Succeeded | **NSW** | | 1 | 1 | | 1 | | 6 | 9 |
| | **NT** | 2 | | | | | | | 2 |
| | **QLD** | | | | 2 | | | | 2 |
| | **SA** | | 1 | | 1 | 2 | 1 | | 5 |
| | **VIC** | | | 1 | | 2 | 2 | 5 | 10 |
| | **WA** | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Succeeded Total** | **All States** | **2** | **2** | **2** | **3** | **5** | **3** | **11** | **28** |
| **Grand Total** | **All States** | **16** | **8** | **14** | **20** | **22** | **14** | **48** | **142** |

Interestingly, these visualisations make apparent some curious aspects of the 142 cases in the finalised dataset:

a) the vast majority of the finalised dataset's cases resulted in an unsuccessful appeal (114 out of 142; 80.3%);
b) the finalised dataset is dominated by New South Wales judgments (75 total), followed distantly by Victorian judgments (33 total);
c) the number of judgments each year varied significantly, but 2018 was a particularly case-heavy year; and
d) the (annual) rate of appeals being successful rose significantly in 2018 as a result of NSW and Victorian judgments.

Having now finalised, understood and described the dataset, it is time to hand the dataset over to a supervised machine learning algorithm.

### D *Supervised Machine Learning: Training, Prediction and Assessing Accuracy*

For any machine learning experiment, a piece of software must be chosen. For the mock analysis, the author used the sci-kit learn library (version 0.1.24.2) for Python 3.7.4. This software will be essential in the following steps:

- pre-processing the finalised dataset;
- selecting a classification algorithm;
- selecting features and categories; and
- splitting the finalised dataset into training and testing sets.

#### 1 *Dataset Pre-processing*

The finalised dataset must have its irrelevant columns stripped out. To this end, the finalised dataset had unnecessary columns removed, so that all that remained were the *Plaintiff Fail* and *State* values.

Then, the finalised dataset must be pre-processed into a format that a machine learning algorithm can use. Accordingly, the remaining values were rendered in numerical form (as required for sci-kit learn.) To this end:

- Values in *Plaintiff Fail* had TRUE (i.e., appeal dismissed) converted to 2, and FALSE (i.e., appeal allowed) converted to 1.

- Values in *State* value were converted to a corresponding number.[169]

The finalised dataset, now rendered numerically and with irrelevant columns removed, was saved into a CSV file.

### 2 *Choice of Classification Algorithm*

The type of classifier must then be chosen. The author chose scikit-learn's KNN classifier algorithm (clf = neighbors.KNeighborsClassifier()) using scikit-learn's default hyperparameters.[170] However, there are numerous other classifiers that could be used instead, such as the 'naïve Bayes' or 'support vector machine' algorithms.[171]

### 3 *Selecting Features and Categories*

A researcher must select the categories into which the KNN algorithm will categorise the results of unseen cases. For instance, if researchers want the algorithm to predict whether an unseen case is a successful or failed appeal (as we do for the mock analysis), the two categories would be:

1. Successful Cases (i.e., appeal allowed, *Plaintiff Fail = 1*); and
2. Failed Cases (i.e., appeal dismissed, *Plaintiff Fail = 2*).[172]

A researcher must also select the features of the data that the KNN algorithm will consider. As mentioned and listed earlier, features are qualitative attributes for each case (e.g., judge's name).[173]

To avoid 'data leakage', the only features that can be used are those that do not contain details of the cases' verdicts.[174] Depending on what sort of analysis is being conducted, it could be relevant to

---

[169] NSW = 1, NT =2, Qld = 3, SA = 4, Vic = 5, WA = 6. This conversion was done manually to ensure accuracy but it can be done in an automated or semi-automated fashion, as was tested by the author. There were no cases from the ACT or Tasmania in the finalised dataset, hence those jurisdictions' omission. The author takes this chance to note that the Northern Territory, is not an actual State of Australia. See *Northern Territory (Self-Government) Act 1978* (Cth) s 5. Nonetheless, for convenience, the Northern Territory is referred to as a State in this article, which is functionally accurate given that there is no distinction between how the Federal Court operates in Territories vis a vis States. This reflects how the Northern Territory functions as a State for most intents and purposes since it attained full self-government. These comments are equally true for the ACT. See *Australian Capital Territory (Self-Government) Act 1988* (Cth) ss 7, 28; Bromberg and Cardaci (n 89) 136.
[170] 'Sklearn.Neighbors.KNeighborsClassifier', *scikit-learn* <https://scikit-learn/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>.
[171] See generally Cui et al (n 48); Uddin et al (n 44).
[172] To implement these classes in the script, the class of each case is determined according to whether the 'Plaintiff success' feature/column has a TRUE or FALSE value.
[173] See earlier in this article at Part IV(C)(2)(b) Collection for the Mock Analysis
[174] O'Sullivan and Beel (n 102) 5; Masha Medvedeva (n 53) 50–2.

consider features such as the presiding judge, the number of days elapsed between the hearing and judgment dates, or the number of authorities cited in the judgment.

### 4 *Split Finalised Dataset into Training and Testing Sets*

At this point, the finalised dataset needs to be split into a *training set* and a *testing set*. This is crucial.[175]

The *training set* is a subset of the finalised data upon which the KNN algorithm will be trained. In contrast, a *testing set* is a subset of the final dataset that is not included in the testing set. As this implies, any given case will only appear in *either* the training or testing set, not in both. Indeed, the two sets *must* be mutually exclusive in their content. For both sets, the features and categories must remain formatted/arranged in the same way.[176]

There are different methods splitting the data into training and testing sets. For instance, Medvedeva, Vols and Wieling (2020) split the data into training and testing sets with 77%-23% split.[177] Alternatively, Medvedeva et al (2021), used a chronological split; meaning *older cases* were placed into the *training set*, and newer cases were put into the testing set.[178]

For the mock analysis, the author adopted this chronological-split method because (a) of its simplicity and realism,[179] and (b) it is conceptually harmonious with the goal of predicting *future* judgments with past judgments.[180] In total, this results in a training set of 94 cases (~66% of the finalised dataset.)[181] Consequently, the training set (94 cases from 2012 to 2017) contains cases with the following outcomes:

- 77 judgments (80.91%), were 'appeal denied' (Plaintiff Fail = 2); and
- 17 judgments (18.82%).were 'appeal allowed' (Plaintiff Fail = 1)

Conversely, the testing set (48 cases in 2018, which is ~34% of the finalised dataset) contained cases with the following outcomes:

- 37 judgments (77.08%); were 'appeal denied' (*Plaintiff Fail* = 2); and
- 11 judgments (22.92%).were 'appeal allowed' (*Plaintiff Fail* = 1)

---

175 O'Sullivan and Beel (n 102) 3.
176 Jenni AM Sidey-Gibbons and Chris J Sidey-Gibbons, 'Machine Learning in Medicine: A Practical Introduction' (2019) 19(1) *BMC medical research methodology* 64, 12 ('Machine Learning in Medicine').
177 Medvedeva et al (n 15) 5.
178 Medvedeva et al (n 15); Masha Medvedeva (n 53) 124.
179 Masha Medvedeva (n 53) 124; O'Sullivan and Beel (n 102) 3–4.
180 Masha Medvedeva (n 53) 124; O'Sullivan and Beel (n 102) 3–4. See also Sidey-Gibbons and Sidey-Gibbons (n 176) 12.
181 Practically, this was implemented by using the following sci-kit learn function: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=94, shuffle = False

### E *Training and Testing – Execution, Results and Discussion*

At long last, now that the finalised dataset has been split, it would be time for a researcher to train their chosen algorithm. Then, the researchers would test a trained algorithm on the testing set to see how well it performs *as measured by* the selected test metric (i.e., a metric for measuring performance.)[182]

Accordingly, for the mock analysis, the KNN algorithm was trained on a training set consisting of 94 cases dated from 2012 to 2017, by using sci-kit learn's KNN classifier training function given earlier in the article.[183]

#### 1 Testing

As mentioned earlier,[184] the mock analysis' trained KNN algorithm attempts to predict the verdict of each testing case in the set by predicting that each case shares the same verdict as 'most similar' cases in the training set – with the most similar cases simply being those heard in the same *State* (i.e., State/Territory of the relevant registry.)[185]

The trained algorithm can return the resulting test metric (i.e., accuracy) in a readable format once it has finished testing.[186] Generally, researchers should report (a) the resulting test metric (i.e., accuracy), and (b) whether this result is *significant* (or not) according to the earlier-determined significance requirement (i.e., by surpassing an accuracy benchmark set by the HPM.)[187]

For the mock analysis, the test metric is *accuracy*. Accuracy is measured by comparing the accuracy of the predictions that the trained KNN algorithm has made on the testing set against the true outcomes for the test's cases (which were concealed to the trained algorithm).[188] Put differently, accuracy is the percentage of testing set cases which have their verdict (i.e., appeal denied or allowed) correctly predicted by the trained KNN algorithm.[189]

Given that (a) most cases in the training set resulted in appeals being denied, and (b) <u>77.08%</u> of the testing set's cases resulted in appeals being denied; the

---

[182] Medvedeva et al (n 15) 6; Masha Medvedeva (n 53) 15.
[183] See earlier in this article at Part IV(D)(2 Choice *of Classification Algorithm*. See 'Sklearn.Neighbors.KNeighborsClassifier' (n 170).
[184] See earlier in this article at Part III WHAT IS MACHINE LEARNING?
[185] See Cui et al (n 48) e128. See also Surden (n 44) 89–91; '1.6. Nearest Neighbors', *scikit-learn* <https://scikit-learn/stable/modules/neighbors.html>.
[186] Sidey-Gibbons and Sidey-Gibbons (n 176) 10. E.g., via a Python print() function.)
[187] O'Sullivan and Beel (n 102) 8–10; Cardaci (n 106) 180–3. See earlier in this article at Part IV(B Formulating Hypotheses
[188] Sidey-Gibbons and Sidey-Gibbons (n 176) 10.
[189] See Medvedeva et al (n 15) 6.

- the HPM predicts that all testing set cases will fail; and
- the accuracy benchmark set by the HPM is 77.08%.[190]

Consequently, unless the trained KNN algorithm is more than 77.08% accurate, the null hypothesis will be supported (consistent with the rule of law) and the alternative hypothesis (consistent with legal realism) will not be supported; and vice versa.

## 2 Presenting and Discussing Results

The author, for the mock analysis, finally tested the trained KNN algorithm on the testing set using sci-kit learn's sklearn.neighbors.KNeighborsClassifier.score function.[191] The trained KNN algorithm's accuracy was exactly 87.5%. Due to this accuracy figure being higher than the accuracy benchmark of the HPM (77.08%), the *alternative* hypothesis is supported.[192]

At first glance this result would suggest, surprisingly, that the State in which the relevant matters are being heard is predictive of their outcome. However, before getting too excited or shocked, the author considers this surprising result to be caused by the significant limitations of the mock analysis. The limitation of having relatively few cases in the dataset due to the stringent inclusion criteria and the relative smallness of cases in Australia as compared to other larger jurisdictions such as the United States and European Union.[193] The smallness of the dataset is enough reason for the author to (a) reject that this result shows a breakdown of the rule of law in the Federal Court, and (b) not consider the result to be significant.

Finally, it is crucial for researchers to be conservative in trying to *explain* seemingly anomalous results, especially when dealing with results concerning judicial decision-making.[194] Indeed, Beach J wisely writes that '[f]or any set of empirical observations, there is always more than one theory that can explain them or is empirically adequate.'[195] As such, the author suggests generally refraining from attempting to give a definite explanation for the result.

[190] O'Sullivan and Beel (n 102) 8–10.
[191] 'Sklearn.Neighbors.KNeighborsClassifier' (n 170).
[192] Cox (n 7) 325; Beach (n 48) 146–7.
[193] Medvedeva et al (n 15) 3–4; Medvedeva, Vols and Wieling (n 12) 246–7; Strickson and De La Iglesia (n 142) 3.2; Masha Medvedeva (n 53) 9, 90. See especially Katz, Bommarito and Blackman (n 14) 14.
[194] 'A machine learning algorithm requires a substantial amount of data to be trained with. For this reason, we excluded articles with too few cases. We included only articles having at least 100 cases' according to Masha Medvedeva (n 53) 91.
[195] Beach (n 48) 120–1.

## V CONCLUDING REMARKS AND ETHICAL CONSIDERATIONS

As has been shown, the use of machine learning techniques for judgment-prediction opens up many opportunities for discovering hidden truths in the legal system, and the ability to use it is within the reach of lawyers. This article, through its exposition of theory and step-by-step guidance, is intended to help lawyers access these opportunities.

Before concluding, the author stresses that lawyers must *ethically* use the investigative power afforded by machine learning lawyers.[196] Indeed, as the wise Ben Parker said in *Spider-Man*, 'with great power comes great responsibility,'[197] The ethical use of artificial intelligence in the legal profession is a topic of ongoing conversation in the profession and academia,[198] but it is uncontroversial that lawyers' use of machine learning technology is governed by their legal professional responsibilities.[199] Consequently, any lawyer using machine learning techniques in legal practise must, *inter alia* (a) sufficiently understand the technology,[200] and (b) always exercise *independent professional judgement* by never blindly accepting answers given by computers.[201]

All in all, the author hopes this article better equips and inspires at least some readers to hunt for the truth, just as Mulder and Scully did. Though admittedly, a lawyer's use of machine learning probably wouldn't make for good television.

---

[196] See generally Francesco Contini, 'Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings' (2020) 2(1) *Law, Technology and Humans* 4, 5; Law Society of Western Australia, *The Future of the Legal Profession* (12 December 2017) 7–8 <https://lawsocietywa.asn.au/wp-content/uploads/2015/10/2017DEC12-Law-Society-Future-of-the-Legal-Profession.pdf>.

[197] Quoted from Ben Parker (also known as 'Uncle Ben') in *Spider-Man* (Directed by Sam Raimi, Columbia Pictures, 3 May 2002).

[198] See Legg and Bell (n 12); Law Society of Western Australia (n 196) 7–8; Michael Legg and Felicity Bell, 'Artificial Intelligence and Solicitors' Ethical Duties', *Law Society Journal* (1 February 2022) <https://lsj.com.au/articles/artificial-intelligence-and-solicitors-ethical-duties/>. Incompetent use of the technology could include using a biased or insufficiently large dataset See Beach (n 48) 132, 142.

[199] Law Society of Western Australia (n 196) 8. This is a result a lawyer's duty to be competent. See Legg and Bell (n 198).

[200] This is a consequence of lawyers' duty to be competent. See generally *Legal Profession Uniform Law Australian Solicitors' Conduct Rules 2015* (n 124) r 4. See especially Law Society of Western Australia (n 196) 8; Legg and Bell (n 198).

[201] See generally *Legal Profession Uniform Law Australian Solicitors' Conduct Rules 2015* (n 124) r 4. See especially Law Society of Western Australia (n 196) 8; Legg and Bell (n 12) 54–5; Legg and Bell (n 198); Nunez (n 12) 194–5, 204. 'We do not think that any of the models described in this chapter can or should be used for *making decisions* in court, especially those where human rights are at stake' according to Masha Medvedeva (n 53) 137.