

Legal Cases: Argumentation *versus* ML

Trevor BENCH-CAPON

Department of Computer Science, University of Liverpool, UK

In 1965 Herb Simon, one of the founding fathers of AI, wrote “Machines will be capable, within twenty years, of doing any work a man can do” [29]. While this appears as one of the top 10 bad tech predictions [19], futurologists have said much the same thing ever since, although switching to gender neutral language, and sometimes increasing the timescale to as much a fifty years. Given the confidence of the early days of AI (the proposal for the 1956 Dartmouth workshop spoke of machine simulation of *every* feature of intelligence) it was natural to ask *Can/should computers replace judges?* [16]. The question was always asked more by laymen than actual practitioners of AI and Law. It was noted in [16] that Thorne McCarty, a founding father of AI and Law, had reservations about the ability of a program to “capture many of the significant [aspects] of legal reasoning”. By the mid-eighties the consensus was very much that “Computers cannot replace the lawyer, but computers can aid the legal profession in several useful ways.” [21]. This was the basis on which AI and Law proceeded for the next two or three decades, but recently the idea of computers replacing lawyers seems to have undergone something of a revival (at least in the popular mind).

The British Broadcasting Corporation (BBC) has broadcast a number of radio programmes on the possibility of AI replacing people in white collar jobs, including various legal jobs. An *Analysis* programme [5] discussed replacement of legal jobs with AI, and their flagship programme on law, *Law in Action*, devoted a programme to *AI and the Law* [6]. This revival has little to do with “Good Old Fashioned AI and Law”(GOFAIL) (adapting Haugeland’s phrase [20], to refer to symbolic AI and Law) of the sort described in [9]. Rather people are thinking of AI as algorithms applied to large volumes of data or a program like IBM’s Watson [15] targetted at legal questions. The greater availability of large data sets, the growth of computer power and improvements in algorithms for knowledge discovery and machine learning have all led to increased expectations in such programs. In AI and Law, E-Discovery (the topic of the DESI workshops held in conjunction with ICAIL since 2007 [17]) relies on such approaches.

A prime example of this new AI is [2]. This caused a great deal of press interest: both the *Guardian* [23] and the BBC news website [30] ran articles. [23] summarises the program: “Software that is able to weigh up legal evidence and moral questions of right and wrong has been devised by computer scientists at University College London, and used to accurately predict the result in hundreds of real life cases. The AI judge has reached the same verdicts as judges at the European court of human rights in almost four in five cases involving torture, degrading treatment and privacy.”

Given these improvements, should we not abandon GOFAIL for a variant on the new AI? It has always been argued by AI and Law practitioners that the decision is of secondary importance, and what matters is the explanation, the *argument*. I wish to argue that there are differences between some central legal problems and problems suitable for machine learning and knowledge discovery, which mean that we still need arguments.

- Machine learning is retrospective (intended to classify past decisions), whereas case law is prospective (intended to influence future decisions). When deciding a case we are not *discovering* something common to the previous cases, we are *creating* a rule to decide a particular case, and intended to constrain *future* cases. This rule should be consistent with previous cases [9], but the new rule may introduce new features distinguishing these previous cases. A new case may lead to a new theory and reinterpretation of the existing theory [24], [26], [14].
- Machine learning does not need to explain itself. Some ML systems do produce rules (KDD, Association Rule Mining, Decision Trees) but others (Neural Nets, Algorithmic “Black Boxes”) do not. In Law the explanation/rule is what matters. Even where black box techniques such as neural networks are used in AI and Law, an effort is made to extract meaningful rules from the trained network, [12] and [7]. Note, however, that [7] showed that highly accurate performance was no guarantee that a complete set of accurate rules had been discovered.
- Success is not judged statistically. A ML system is considered successful if it correctly classifies enough future cases. But a new judge-made rule classifies correctly since the decision itself determines the classification of future cases: there is no independent “fact of the matter”. A rule is successful if endorsed by a relevant consensus, and survives any appeal process challenging it. Law requires a very high success rate. Whereas 80% is reported as successful for most prediction programs, wrongly deciding 20% of cases would not be acceptable. Improvement may be possible, but around 80% has remained the best for a long time [4], perhaps suggesting a ceiling, whereas GOFAIL generally reports 90+%.
- In machine learning we typically have a large dataset (e.g. of precedent cases) which we are trying to classify, whereas legal decisions are primarily concerned with a single case. ML relies on large data sets but case law systems can work with a few landmark cases. HYPO [3], the best known legal CBR system, used fewer than 30: reasonable theories have been developed for the wild animals domain with only half a dozen [8]. ML derives its authority from numbers whereas case law derives its authority from the status of the court or of the judge or the cogency of the argument. ML needs many ordinary cases. Case law is driven by hard cases.
- Past data may not be homogeneous. Attitudes and case law change over time and later cases are preferred to older cases, so that it can be difficult to determine the (degree of) relevance of a decision. Although case law is often considered without a notion of sequence, in fact sequence is important, and domain theories evolve over time as landmark cases appear (e.g. [28], [22]).
- Past data may contain noise (including prejudices, conscious and unconscious [25], [18]). Although D’Amato says [16] “ By simply not programming into the computer any such character traits, we can be assured that the computer will be impartial as to them”, this is not true of many learning systems which simply assume the future will resemble the past, bias and all [13].
- Case Law requires an argument to ground the new rule. This can be clearly seen in Supreme Court oral hearings [1]. Often these arguments reflect ideas of purpose or value: e.g.[11], [10], rarely considered in ML.

Although I have here concentrated on legal applications, the same considerations also apply in other areas of social science. Microsoft’s chatbot Tay [27] provides a dire warning to those who take AI learning at face value.

References

- [1] L. Al-Abdulkarim, K. Atkinson, and T. Bench-Capon. From Oral hearing to Opinion in the U.S. Supreme Court. In *Proceedings of JURIX 2013*, pages 1–10, 2013.
- [2] N. Aletras, D. Tsarapatsanis, D. PreoŃuc-Pietro, and V. Lampos. Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science*, 2: DOI: 10.7717/peerj-cs.93, 2016.
- [3] K. Ashley. *Modeling legal arguments: Reasoning with cases and hypotheticals*. MIT press, 1991.
- [4] K. Ashley and S. Brüningshaus. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, 17(2):125–165, 2009.
- [5] BBC. Analysis: When robots steal our jobs, 2015. <https://www.bbc.co.uk/programmes/b0540h85>.
- [6] BBC. Law in action: AI and the Law, 2016. <https://www.bbc.co.uk/programmes/b07dlxmj>.
- [7] T. Bench-Capon. Neural networks and open texture. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*, pages 292–297. ACM, 1993.
- [8] T. Bench-Capon. Representing Popov v Hayashi with dimensions and factors. *Artificial Intelligence and Law*, 20(1):15–35, 2012.
- [9] T. Bench-Capon. HYPO’S legacy: introduction to the virtual special issue. *Artificial Intelligence and Law*, 25(2):205–250, 2017.
- [10] T. Bench-Capon and G. Sartor. A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*, 150(1-2):97–143, 2003.
- [11] D. H. Berman and C. D. Hafner. Representing teleological structure in case-based legal reasoning: the missing link. In *Proceedings of the 4th ICAIL*, pages 50–59. ACM, 1993.
- [12] L. Bochereau, D. Bourcier, and P. Bourguine. Extracting legal knowledge by means of a multilayer neural network application to municipal jurisprudence. In *Proceedings of the 3rd International Conference on Artificial Intelligence and Law*, pages 288–296. ACM, 1991.
- [13] S. Buranyi. Rise of the racist robots: how AI is learning all our worst impulses, 2017. <https://www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses>.
- [14] A. Chorley and T. Bench-Capon. An empirical investigation of reasoning with legal cases through theory construction and application. *Artificial Intelligence and Law*, 13(3-4):323–371, 2005.
- [15] B. Cooper. Judges in Jeopardy! Could IBM’s Watson beat courts at their own game? *Yale Law Journal*, 121:87–100, 2011.
- [16] A. D’Amato. Can/should computers replace judges? *Georgia Law Review*, 11:1277, 1976.
- [17] DESI. DESI VII workshop on using advanced data analysis in e-discovery and related disciplines, 2017.
- [18] M. Dunn, L. Sagun, H. Şirin, and D. Chen. Early predictability of asylum court decisions. In *Proceedings of the 16th International Conference on Artificial Intelligence and Law*, pages 233–236. ACM, 2017.
- [19] G. Goble. Top 10 bad tech predictions, 2012. <https://www.digitaltrends.com/features/top-10-bad-tech-predictions/5/>.
- [20] J. Haugeland. *Artificial Intelligence: The very idea*. MIT press, 1989.
- [21] M. M. Hazen and T. L. Hazen. Simulation of legal analysis and instruction on the computer. *Indiana Law Journal*, 59:195, 1983.
- [22] J. Henderson and T. Bench-Capon. Dynamic arguments in a case law domain. In *Proceedings of the 8th International Conference on Artificial Intelligence and Law*, pages 60–69. ACM, 2001.
- [23] C. Johnson. Artificial Intelligence ‘judge’ developed by UCL computer scientists, 2016. <https://www.theguardian.com/technology/2016/oct/24/artificial-intelligence-judge-university-college-london-computer-scientists>.
- [24] E. H. Levi. An introduction to legal reasoning. *U. of Chicago Law Review*, 15(3):501–574, 1948.
- [25] F. E. Marouf. Implicit bias and immigration courts. *New England Law Review*, 45:417, 2010.
- [26] L. T. McCarty. An implementation of Eisner v. Macomber. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 276–286. ACM, 1995.
- [27] R. Price. Microsoft is deleting its AI chatbot’s incredibly racist tweets, 2016. <http://uk.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3>.
- [28] E. L. Rissland and M. T. Friedman. Detecting change in legal concepts. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, pages 127–136. ACM, 1995.
- [29] H. Simon. *The shape of automation for men and management*. Harper and Row, 1965.
- [30] J. Wakefield. AI predicts outcome of human rights cases, 2016. <https://www.bbc.co.uk/news/technology-37727387>.