THE NEED FOR GOOD OLD FASHIONED AI AND LAW

Trevor Bench-Capon[1]

Abstract: «Good Old Fashioned AI and Law» typically involved a partnership between legal expertise and computer knowledge. Nowadays, however, there is a new fashion in AI. Algorithms that learn from large datasets have moved to the centre stage in many domains. In this paper I set out a number of features of the legal domain which mean that we should use the new AI with caution. For systems which support legal decision making and predict case outcomes, there is not yet a real alternative to Good Old Fashioned AI and Law. Machines may be able to teach themselves chess, but not yet law.

Table of contents

---

[1]   Department of Computer Science, The University of Liverpool, Liverpool, UK.

A. INTRODUCTION

I first met Erich Schweighofer in the early nineties through the DEXA series of conferences. His work then was mainly concerned with exploring the use of AI techniques to support legal tasks[2]. His collaboration in this work was the classic AI and Law partnership of a legal expert, Erich, and a computer specialist, Werner Winiwerter. Pairings of a legal specialist with a computer specialist was a fruitful source of AI and Law research at the time. Examples of such partnerships include Don Berman and Carole Hafer[3], John Zeleznikow and Dan Hunter[4] and Henry Prakken and Giovanni Sartor[5].

---

[2] Three example papers are: 1) SCHWEIGHOFER, E., & WINIWARTER, W. (1993). Legal expert system KONTERM-automatic representation of document structure and contents. In *International Conference on Database and Expert Systems Applications* (pp. 486–497). Springer, Berlin, Heidelberg. 2) MERKL, D., SCHWEIGHOFER, E., & WINIWARTER, W. (1994). CONCAT-Connotation analysis of thesauri based on the interpretation of context meaning. In *International Conference on Database and Expert Systems Applications* (pp. 329–338). Springer, Berlin, Heidelberg. 3) WINIWARTER, W., SCHWEIGHOFER, E., & MERKL, D. (1995). Knowledge acquisition in concept and document spaces by using self-organizing neural networks. In *International Joint Conference on Artificial Intelligence* (pp. 75–86). Springer, Berlin, Heidelberg.

[3] Don and Carole produced several important papers in the 1990s including 1) BERMAN, D. H., & HAFNER, C. D. (1989). The potential of artificial intelligence to help solve the crisis in our legal system. *Communications of the ACM*, *32*(8), 928–938. 2) BERMAN, D. H., & HAFNER, C. D. (1991). Incorporating procedural context into a model of case-based legal reasoning. In *Proceedings of the 3rd international conference on Artificial intelligence and law* (pp. 12–20). ACM. 3) BERMAN, D. H., & HAFNER, C. D. (1993). Representing teleological structure in case-based legal reasoning: the missing link. In *Proceedings of the 4th international conference on Artificial intelligence and law* (pp. 50–59). ACM and 4) BERMAN, D. H., & HAFNER, C. D. (1995). Understanding precedents in a temporal context of evolving legal doctrine. In *Proceedings of the 5th international conference on Artificial intelligence and law* (pp. 42–51). ACM.

[4] ZELEZNIKOW, J., VOSSOS, G., & HUNTER, D. (1993). The IKBALS project: Multi-modal reasoning in legal knowledge based systems. *Artificial Intelligence and Law*, *2*(3), 169–203 and ZELEZNIKOW, J., & HUNTER, D. (1995). Reasoning paradigms in legal decision support systems. *Artificial Intelligence Review*, *9*(6), 361–385.

[5] Among the important works produced by this pair are 1) PRAKKEN, H., & SARTOR, G. (1995). On the relation between legal language and legal argument: assumptions, applicability and dynamic priorities. In *Proceedings of the 5th international conference on Artificial intelligence and law* (pp. 1–10). ACM. 2) PRAKKEN, H., & SARTOR, G. (1996). A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning. *Artificial Intelligence and Law*, *4*(3–4), 331–368 and 3) PRAKKEN, H., & SARTOR, G. (1998). Modelling Reasoning with Precedents in a Formal Dialogue Game. *Artificial Intelligence and Law*, *6*, 231–287.

The idea of close cooperation between a domain expert and a computer specialist was central to *Good Old Fashioned AI and Law*[6] (GOFAIL), which was largely working in the expert systems tradition[7] in which knowledge engineers elicited knowledge from domain experts and represented it in executable form.

Today, however, the need for expertise is no longer recognised to the same extent. Increasingly «Artificial Intelligence» has come to mean the application of learning algorithms to large amounts of data[8]. Thus, in Chess, whereas Deep Blue, the first computer to beat a world champion[9] relied on an evaluation function provided by chess experts, Alphazero, the currently most powerful chess program, developed by Deepmind, was trained solely via self-play[10], and placed no reliance at all on the body of knowledge that has been developed in Chess over centuries. Even more remarkably, it was possible to learn the much more complicated game of Go in the same way.

Given the success of such techniques it is unsurprising that there should be a desire to apply them in law. Large amounts of data in the form of decided cases are available, and so it seems that it should be possible to train an algorithm to predict the outcome of future cases. An example of this work in the academic realm is a program designed to identify examples of human rights violations found by the European Court of Human Rights.[11]

---

[6] The term is meant to evoke «Good Old Fashioned AI» (GOFAI), coined by John Haugeland in HAUGELAND, J. (1989). *Artificial intelligence: The very idea*. MIT press.

[7] Inspirational here was the MYCIN project: SHORTLIFFE, E. H., & BUCHANAN, B. G. (Eds.). (1985). *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*. Addison-Wesley Publishing Company.

[8] For example see https://www.varsity.co.uk/science/13595 (last accessed 20th September 2019), an article in Cambridge University's leading student newspaper.

[9] Hsu, F. H. (2004). Behind Deep Blue: Building the computer that defeated the world chess champion. Princeton University Press.

[10] SILVER, D., HUBERT, T., SCHRITTWIESER, J., ANTONOGLOU, I., LAI, M., GUEZ, A., ... & LILLICRAP, T. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*.

[11] ALETRAS, N., TSARAPATSANIS, D., PREOȚIUC-PIETRO, D., & LAMPOS, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science*, *2*, e93. A more popular account of the work is at https://www.theguardian.com/technology/2016/oct/24/artificial-intelligence-judge-university-college-london-computer-scientists (last accessed 20th September 2019).

Given this revolution in what is possible using learning algorithms on big data sets, should we not abandon GOFAIL in favour of the new AI? Perhaps not: 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. The only form of explanation offered by Aletras et al was «the 20 most frequent words, listed in order of their SVM [Support Vector Machine] weight». These do not, however, look immediately promising: the list for topic 23 of article 6 predicting violation, for example, is:

> «*court, applicant, article, judgment, case, law, proceeding, application, government, convention, time, article convention, January, human, lodged, domestic, February, September, relevant, represented*»

One doubts whether this would be the list expected by a legal expert, or that pointing to the presence of these words would be found a persuasive argument by the European Court of Human Rights.

I wish to argue that there are a number of differences between legal problems, especially those regarding making, or predicting, decisions on cases, and problems suitable for machine learning and knowledge discovery, which mean that we still need arguments based on legal knowledge, and the involvement of legal experts to produce and assess them.

B.   PROBLEMS WITH MACHINE LEARNING

In this section I will point to several differences between case law and domains in which machine learning is successful, to explain why GOFAIL still has a place in AI and Law.

1.   Machine learning is retrospective

Machine learning is retrospective (trained on 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 which is intended to constrain *future* cases. This rule should be consistent with previous cases[12], but the new case may lead to a new theory and a re-

---

12   For an overview of case based reasoning in AI and Law, see BENCH-CAPON, T. J. M. (2017). Hypo's legacy: introduction to the virtual special issue. *Artificial Intelligence and Law, 25*(2), 205–

interpretation of the existing cases[13]. A new rule may broaden or narrow existing features, introduce new features distinguishing previous cases, which may lack the new feature or not have considered the feature, or represent a shift in the common values of society[14] which the court is supposed to express[15]. Thus a rule which can be induced from the previous cases, and which a machine learning system would be expected to derive, may not be the rule which is actually applied to the current case (and to future cases). One of the great strengths of law is that its interpretation is dynamic, able to meet new situations and to adapt as society changes. But how it will develop cannot be predicted from a consideration of the past. A legal expert might be able to conjecture some trends[16], but that art is beyond the capacity of current machine learning algorithms.

2.   Success in law is not statistical

A Machine Learning system is considered successful if it correctly classifies enough future cases. The program in Aletras *et al* was deemed a success on the grounds that it got 79% of its predictions right. But percentages cannot be applied to the decisions of human judges. A new judge-made rule always classifies the case it which it is made correctly since the judge is empowered to say what is so. There is no independent «fact of the matter». Whether the rule is successful is shown by its future treatment: it

---

250. For formal accounts of precedential constraint see HORTY, J. F., & BENCH-CAPON, T. J.M. (2012). A factor-based definition of precedential constraint. *Artificial Intelligence and Law*, *20*(2), 181–214 and RIGONI, A. (2015). An improved factor based approach to precedential constraint. *Artificial Intelligence and Law*, *23*(2), 133–160.

[13] See 1) LEVI, E. H. (1948). An introduction to legal reasoning. *The University of Chicago Law Review*, *15*(3), 501–574. 2) MCCARTY, L. T. (1995). An implementation of Eisner v. Macomber. In *Proceedings of the 5th International Conference on AI and Law (* 276–286) and 3) CHORLEY, A., & BENCH-CAPON, T. (2005). An empirical investigation of reasoning with legal cases through theory construction and application. *Artificial Intelligence and Law*, *13*(3–4), 323–371.

[14] In the words of Justice Marshall in the case of *Furman v Georgia,* 408 U.S. 238 (1972), «*stare decisis* must bow to changing values».

[15] LEVI, op. cit.

[16] For a discussion of this see BERMAN, D. H., & HAFNER, C. D. (1995) op. cit. and RISSLAND, E. L., & XU, X. (2011). Catching gray cygnets: an initial exploration. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law* (pp. 151–160). ACM.

succeeds if it is endorsed by a relevant consensus, and it survives any appeal process challenging it. This will require not just a rule, but also a convincing justification for that rule, in terms that will be understood by, and able to persuade, legal professionals. Law also aspires to a very high success rate. Whereas around 80% is reported as successful for most prediction programs, using a program knowing that 20% of legal cases would be wrongly decided would not be acceptable[17]. This is certainly true in decision making and prediction systems, although a less demanding standard may be applied to particular legal tasks. For example, for E-discovery[18], which acts as a first pass filter, a lower precision may be acceptable. But for systems intended to support the application of law, there should be something closer to the 90%+ often achieved using GOFAIL techniques[19]. Improvements in the performance of the machine learning systems may be possible, but the 80% ceiling has not been raised for a long time, whereas the GOFAIL approaches consistently perform better.

3.   Explanations

Machine learning often does not provide explanations. Some ML systems do produce explanations, but others, like the European Human Rights prediction system of Aletras *et al*, do not. In Law the explanation is what matters. This has always been recognised in the AI and Law community. Even where black box techniques such as neural networks were used in AI and Law, an effort was made to extract meaningful rules

---

[17]   This applies to high level court decisions. As we will see later, some routine welfare benefit decisions do if fact experience a high error rate. This, however, was seen as a problem that must be addressed, not an acceptable feature of the system: knowingly introducing a system with a high error rate would not be acceptable.

[18]   CONRAD, J. G. (2010). E-Discovery revisited: the need for artificial intelligence beyond rror rate would not be acceptable.information retrieval. *Artificial Intelligence and Law*, *18*(4), 321–345.

[19]   Examples of GOFAIL approaches achieving this degree of success are: 1) BRUNINGHAUS, S., & ASHLEY, K. D. (2003). Predicting outcomes of case based legal arguments. In *Proceedings of the 9th international conference on Artificial intelligence and law* (pp. 233–242). ACM. 2) CHORLEY, A., & BENCH-CAPON, T. (2005). AGATHA: Using heuristic search to automate the construction of case law theories. *Artificial Intelligence and Law*, *13*(1), 9–51 and 3) AL-ABDULKARIM, L., ATKINSON, K., & BENCH-CAPON, T. (2016). A methodology for designing systems to reason with legal cases using abstract dialectical frameworks. *Artificial Intelligence and Law*, *24*(1), 1–49.

from the trained network[20]. Note, however, that Bench-Capon (1993) showed that highly accurate performance (let alone 80%) was no guarantee that a complete set of accurate rules had been discovered. Rules have also been produced by inductive logic programming[21] and using data mining techniques[22]. However, the rules they produce need to be assessed for credibility: a successful rule will have a convincing justification. A legal explanation is normally not simply in terms of a rule, but why that rule is the rule that should be followed. These reasons can readily be obtained from legal experts, or from written decisions, but a machine learning system will tend to quote the numbers of cases supporting it, which provides no more than a probabilistic justification, which will always leave room for doubt as to whether it applies in a particular case.

4. Size of dataset

In machine learning we typically need a large dataset (e.g. of precedent cases) which we are trying to classify, whereas legal decisions are primarily concerned with a single case and a handful of relevant precedents. Machine Learning requires large data sets because its rules have only a statistical justification, but case law systems can work with a few landmark cases, on which intellectually satisfying arguments can be based. HYPO[23] used fewer than 30: CATO[24] used 148, IBP[25] used 186, and reasonable theories

---

[20] For example, BOCHEREAU, L., BOURCIER, D., & BOURGINE, P. (1991). 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* (pp. 288–296). ACM and BENCH-CAPON, T. (1993). Neural networks and open texture. In *Proceedings of the 4th international conference on Artificial intelligence and law* (pp. 292–297). ACM.

[21] MOŽINA, M., ŽABKAR, J., BENCH-CAPON, T., & BRATKO, I. (2005). Argument based machine learning applied to law. *Artificial Intelligence and Law*, *13*(1), 53–73.

[22] WARDEH, M., BENCH-CAPON, T., & COENEN, F. (2009). Padua: a protocol for argumentation dialogue using association rules. *Artificial Intelligence and Law*, *17*(3), 183–215.

[23] ASHLEY, K. D. (1991). Modeling legal arguments: Reasoning with cases and hypotheticals. MIT press.

[24] ALEVEN, V. (2003). Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. *Artificial Intelligence*, *150*(1–2), 183–237.

[25] ASHLEY, K. D., & BRÜNINGHAUS, S. (2009). Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law*, *17*(2), 125–165.

have been developed for the wild animals domain with only half a dozen[26]. Machine Learning and Data Mining derive their authority from their being based on large quantities of data whereas case law derives its authority from the status of the court or of the status of the judge or the cogency of the argument. Machine Learning needs many ordinary cases, so that the core of the concepts can be established. Case law is driven by hard cases, which push (and so define) the boundaries of the concepts. It is at these boundaries that legal disputes arise: easy cases in the core of the concept do not really require support to settle them.

5. Not all data are equal

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 in systems without a notion of sequence[27], in fact sequence is important, and domain theories evolve over time as landmark cases appear[28]. This means that some of the cases may be applying superseded doctrine or were heard in a social situation that no longer applies. Some decisions may even have been explicitly overturned. Thus deciding what cases to use for training the algorithm needs a great deal of care so as not to include decisions that may mislead. This itself requires a good deal of expertise and an understanding of the domain which GOFAIL would have encoded directly. Moreover, the resulting dataset may, once potentially misleading cases have been weeded out, be too small for the algorithms to perform optimally.

---

[26] BENCH-CAPON, T. J.M. (2012). Representing Popov v Hayashi with dimensions and factors. *Artificial Intelligence and Law*, *20*(1), 15–35.

[27] BENCH-CAPON, T. J. M. (2017) op. cit.

[28] See for example 1) RISSLAND, E. L., & FRIEDMAN, M. T. (1995). Detecting change in legal concepts. In *Proceedings of the 5th international conference on Artificial intelligence and law* (pp. 127–136). ACM. 2) HENDERSON, J., & BENCH-CAPON, T. (2001). Dynamic arguments in a case law domain. In *Proceedings of the 8th international conference on Artificial intelligence and law* (pp. 60–69). ACM and 3) HENDERSON, J., & BENCH-CAPON, T. (2019). Describing the Development of Case Law. In *Proceedings of the 17th international conference on Artificial intelligence and law* (pp. 32–41). ACM.

6.   Error and Bias

It is not unusual for a substantial number of past decisions to be simply wrong. This is particularly true of routine welfare benefit decisions. Groothuis and Svensson[29] drew attention to this in connection with the Netherlands General Assistance Act, and reported experiments which suggested that an error rate of more than 20% was typical. The problem is not confined to the Netherlands: The US National Bureau of Economic Research on US Disability Insurance reports «*inconsistencies that suggest a potentially high rate of errors.*»[30] In the UK, an official UK Publication produced by the Committee of Public Accounts reported «*Too few decisions are right first time, with a error rate of 50% for Disability Living Allowance. There are also regional differences in decision making practices that may lead to payments to people who are not eligible for benefits.*»[31] Thus large quantities of past data may contain a significant amount of erroneous material, and it is important that any algorithm be robust in the face of such noise. This was an explicit point in both Mozina *et al* (2003) and Warda *et al (2009)*[32], but does not often receive attention in more recent work such as Aletras et al (2016)[33]. The problem is less acute for GOFAIL, which, rather than working with cases decided, often under time pressure, by low level adjudicators, is working with high level, often landmark, cases, which are heard by the very best judges with enough time to be careful, and which are often subject to a great deal of scrutiny and comment. But if we understand that our data contains erroneous decisions it is essential that the errors are not perpetuated in an AI system. Ironically it is just this sort of low level, high volume, routine cases where errors are most likely, that are most likely to be seen as suitable

---

[29] GROOTHUIS, M. M., & SVENSSON, J. S. (2000). Expert system support and juridical quality. In *Proceedings of JURIX* 2000 IOS Press, Amsterdam (pp1–10).

[30] See http://www.nber.org/aginghealth/winter04/w10219.html (last accessed 20th September 2019).

[31] *Getting it right: Improving Decision-Making and Appeals in Social Security Benefits.* Committee of Public Accounts. London: TSO, 2004 (House of Commons papers, session 2003/04; HC406).

[32] MOZINA et al (2003) op. cit. WARDEH, M., COENEN, F., & BENCH-CAPON, T. (2009). Arguing from experience to classifying noisy data. In *International Conference on Data Warehousing and Knowledge Discovery* (pp. 354–365). Springer, Berlin, Heidelberg.

[33] Aletras et al (2016) op.cit.

for decision using AI. Certainly these cases were seen as the most ripe for the application of GOFAIL techniques[34]

In addition to simple errors, it has been shown that past decisions also contain bias and prejudices, consious and unconscious[35]. Training systems on data containing bias and prejudices will present problems. A striking example from a UK newspaper is

> «*A stunning report claimed that a computer program used by a US court for risk assessment was biased against black prisoners. The program, Correctional Offender Management Profiling for Alternative Sanctions (Compas), was much more prone to mistakenly label black defendants as likely to reoffend wrongly flagging them at almost twice the rate as white people (45% to 24%), according to the investigative journalism organisation ProPublica.*»[36]

The possibility of bias presents something of a dilemma for Machine Learning systems. If the aim is to predict decisions and these decisions are subject to bias, the bias will need to be reflected in the program in order to achieve a satisfactory success rate. On the other hand, it is acknowledged that the bias should not exist, and so should not be reflected in the program. We do not want the system to learn prejudices, but to eliminate them[37]. Note that GOFAIL does not suffer from this problem: its explicit rules can be examined, and biased rules rejected.

---

[34]   One of the envisaged advantages seen in replacing low level routine decisions with AI was that it would enable a significant reduction in the error rate. This was certainly the case for the UK Retirement Pension Forecast Advisor (SPRINGEL-SINCLAIR, S. (1988). The DHSS Retirement Pension Forecast and Advice System: An update. *KBS in government*, Blenheim Online, 89–106) which successfully addressed this problem, greatly reducing the error rate.

[35]   Examples can be found in MAROUF, F. E. (2010). Implicit Bias and Immigration Courts. *New England Law Review*, *45*, 417 and CHEN, D. L. (2019). Judicial analytics and the great transformation of American Law. *Artificial Intelligence and Law*, *27*(1), 15–42.

[36]   S. BURANYI. Rise of the racist robots: how AI is learning all our worst impulses. *The Guardian* August 8th 2017.

[37]   The aim of the systems mentioned in footnote 33 was to investigate whether there was bias in the past decisions, in the hope that revealing bias would be a step towards eliminating it.

7.   Rationales and Reasons

Case law is not simply a matter of identifying and applying an appropriate rule. On the contrary much of a legal dispute is taken up with arguments as to what rule should be applied. This is particularly clear in the US Supreme Court where much of the Oral Hearing stage is taken up with attempting to formulate an appropriate rule[38]. Also there is often no consensus: we often have majority and minority opinions arguing for different rules. These arguments are often not about the facts of the case, but rather reflect ideas of purpose or value[39]. Sometime these arguments require consideration of trade-offs between and balancing of values[40]. It is hard to see these kinds of arguments emerging from contemporary machine learning algorithms.

C.   ROLES FOR MACHINE LEARNING

The last section discussed some reasons why Machine Learning techniques might not be as successful as it currently hoped in applications intended to support legal decision making, the traditional province of GOFAIL.

There are, however, other legal tasks to which Machine Learning may well prove the most suitable technique. One is e-discovery[41] where many thousands of documents

---

[38]   The Oral Hearing Stage is discussed in RISSLAND, E.L. (1999) Dimension-Based Analysis of Hypotheticals from Supreme Court Oral Argument in Proceedings of the 2nd International Conference on AI and Law (pp111–120) and AL-ABDULKARIM, L., ATKINSON, K., & BENCH-CAPON, T. J.M. (2013). From Oral Hearing to Opinion in the US Supreme Court. In *Proceedings of Jurix 2013* (pp. 1–10).

[39]   See for example 1) BERMAN, D. H., & HAFNER, C. D. (1993). Representing teleological structure in case-based legal reasoning: the missing link. In *Proceedings of the 4th international conference on Artificial intelligence and law* (pp. 50–59). ACM, 2) BENCH-CAPON, T., & SARTOR, G. (2003). A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*, *150*(1–2), 97–143 and 3) BENCH-CAPON, T.J.M, (2011). Relating Values in a Series of Supreme Court Decisions. In *Proceedings of Jurix 2011* (pp. 13–22).

[40]   For example BENCH-CAPON, T., & PRAKKEN, H. (2010). Using argument schemes for hypothetical reasoning in law. *Artificial Intelligence and Law*, *18*(2), 153–174 and BENCH-CAPON, T. J.M., & ATKINSON, K. (2017). Dimensions and Values for Legal CBR. In *Proceedings of Jurix 2017* (pp. 27–32).

[41]   CONRAD op. cit.

must be examined to find those which are relevant. In this task false positives are less of a problem, and so the high success rate demanded by legal decisions can be relaxed. Moreover, the documents selected can be assessed according to a gold standard, and so problems of differing arguments do not arise and the performance of the system can be objectively evaluated. Change is also less of an issue: the topic may remain constant even if the law changes.

Other clustering and information retrieval tasks may also be susceptible to Machine Learning techniques. It is to Erich's credit that while others[42] were attempting to apply techniques such as neural networks to legal decision making, he was looking at other legal tasks such as automated indexing and document clustering[43].

D.   CONCLUDING REMARKS

In March 2015 the British Broadcasting Corporation presented a programme entitled *When Robots Steal Our Jobs*[44]. The theme of the programme was that the new AI which used algorithms to learn from large amounts of data meant that AI would start to replace white collar workers, just as a previous generation of robots had replaced workers in manufacturing. The general promise of, and the hype surrounding, the new AI made practioners of GOFAIL such as myself (and perhaps Erich) wonder whether ours might be one of the jobs under threat. After all, if Alphazero could teach itself Chess and Go, why should there not be an AI program capable of teaching itself law? And then GOFAIL would be as obsolete a skill as making flint arrowheads.

In this paper I have advanced some reasons why Erich and I should not worry. Although the new AI may have some legal applications, the heart of AI and Law, support for legal decisions, is likely to remain resistant to its progress. The nature of legal data, subject to disputed interpretation, change and reinterpretation, is so very different

---

[42]   For example, ZELEZNIKOW, J., & STRANIERI, A. (1995). The split-up system: integrating neural networks and rule-based reasoning in the legal domain. In *Proceedings of the 5th international conference on Artificial intelligence and Law* (pp. 185–194). ACM. Also the systems in footnote 20.

[43]   See the papers cited in footnote 1 *supra*.

[44]   https://www.bbc.co.uk/programmes/b0540h85 (last accessed 21st September 2019).

from the data in other fields that one should not generalise too hastily from success in other fields to success in law. Law is a social construct, not a natural phenomenon. In fact, GOFAIL remains alive as well, and computer scientists continue to collaborate with lawyers on successful applications which outperform the best that the new AI can offer[45]. So Erich (and I) can rest secure in the knowledge that GOFAIL is likely to last longer than our careers.

---

[45] For a very recent example, see AL-ABDULKARIM, L., ATKINSON, K., BENCH-CAPON, T., WHITTLE, S., WILLIAMS, R., & WOLFENDEN, C. (2019). Noise induced hearing loss: Building an application using the ANGELIC methodology. *Argument & Computation*, 10:1 5–22, which represents a collaboration between the Computer Science Department of the University of Liverpool, and the law firm Weightmans. The ANGELIC methodology is described in AL-ABDULKARIM, L., ATKINSON, K., & BENCH-CAPON, T. (2016). A methodology for designing systems to reason with legal cases using Abstract Dialectical Frameworks. *Artificial Intelligence and Law*, *24*(1), 1–49. That paper reports a success rate of greater than 90% across a range of applications.