European Journal of Law and Technology, Vol 11 No.3 (2020)

EJLT European Journal of Law and Technology

Building a 'Personal Data' App

Pauline McBride[1]

Abstract

This article recounts the author's experience of building a legal expert app using the Neota Logic platform and shares reflections on that process. In particular, it describes how the process of building the app involves a different kind of engagement with legal rules and outlines the benefits of such engagement for lawyers. It offers reflections on how the process of building an app delivers insights relevant to the intersection of legal practice and legal theory, speculates about the role that machine learning might play in the design of legal expertise automation apps and touches on the educational value of building a legal expert app.

1. Introduction

In this article I recount my experience of building a legal expert app in Neota Logic[2] and share my reflections on the process.

In Section 1 I explain how apps are constructed using the Neota Logic platform. At Section 2 I discuss my experience in building the app while at Section 3 I focus on the choices that had to be made in creating the app. Sections 4 to 7 set out a series of reflections occasioned by building the app. In particular, I describe how the process of building the app involves a different kind of engagement with the relevant legal rules and outline the benefits of such engagement for lawyers (Section 4). I reflect on how the process delivers insights relevant to the intersection of legal practice and legal theory (Section 5), speculate about the role that machine learning might play in

[1] School of Law, Queen's University, Belfast BT7 1NN, p.mcbride@qub.ac.uk. The author is a solicitor, qualified to practice in Scotland.

[2] I was able to access the Neota Logic platform for self-directed learning while attending a short course 'Legal Technology and Innovation' run by the University of Ulster in 2019. This course was intended to support continuing professional development. I am grateful to Mark Potkewitz and Jane Hollway of the University of Ulster, who ran this interesting course, and to Neota Logic for the opportunity to use their platform.

the design of legal expertise automation apps (Section 6) and touch on the educational value of building a legal expert app (Section 7).

2. Building an app in Neota Logic

2.1 The Neota Logic platform

Neota Logic is a 'no-code' automation platform.[3] It allows professionals to automate and deliver services through bespoke apps. Though not sector-specific, the Neota platform is well-suited to the needs of the legal sector. It facilitates document automation and the delivery of expert advice by automated means.

Applications are built in Neota using a modular approach. Authors of an app define the problem to be addressed, dividing the problem into discrete issues. By means of the platform, authors may specify a series of questions to be put to users of the app, the sequence in which those questions should appear, the text that should be presented along with the questions, and the feedback to be supplied to the user, whether in the form of a report delivered within the app or sent by email to the user. In addition, the platform provides authors with a series of reasoning tools. These tools allow authors to incorporate calculations or construct if/then mappings within the app. If/then mappings may be constructed by various means including by way of decision trees.[4]

Neota provides users of the platform with online training. Thanks to the training it is relatively straightforward to build an app in the platform.

2.2 A 'Personal Data' app

I explored the functionality offered by the Neota platform by building an app that would help users determine whether information would be classed as personal data for the purposes of the data protection regime.

Two factors were important in my choices as to the nature of the app to be built. The app is intended as a legal expert app.[5] I have appropriate expertise in data protection.

---

[3] Neota Logic, 'No-Code Development' <https://www.neotalogic.com/platform/no-code-development/> (accessed 3 July 2020).

[4] In Neota Logic, decision trees and other if/then mappings are the reasoning engines at the heart of the expert system. Lupica at al. suggest that decision trees are a typical feature of legal expertise apps. Lois R Lupica, Tobias A Franklin, and Sage M Friedman, 'The Apps for Justice Project: Employing Design Thinking to Narrow the Access to Justice Gap' (2017) 44 Fordham Urb L J 1363 <https://ir.lawnet.fordham.edu/ulj/vol44/iss5/> (accessed 1 July 2020).

[5] According to Greenleaf the 'Components of a legal expert system include the 'inferencing mechanism', the 'knowledge base', the application developer interface, the user interface and

The guidance[6] produced by the Information Commissioner's Office[7] as to the meaning of personal data is accompanied by a flowchart of questions that might be used to determine whether information is personal data.[8] The availability of the flowchart made it straightforward to assess the ease with which a decision tree could be generated using the Neota Logic platform.

2.3. Designing the app

In designing an app, consideration has to be given to the way in which the user will interact with the app, the information that will be provided to and collected from the user and the overall look and feel. All of these aspects are important. However, if the app is to operate as a means of delivering legal expertise, the design of the app's reasoning engine (in this case, a decision tree) is key.

Decision trees model a set of rules for providing decisions (or results) according to the values of a series of variables.[9] In Neota Logic the values of the variables of the decision tree may be captured by means of information input by users in response to questions. The results can be presented to users either as text set out within the app or by means of reports that can be downloaded via the app or issued by email.

For the purposes of the Personal Data app, the rules to be modelled by the decision tree are those set out in the Guidance provided by the ICO.[10]

The ICO's Guidance suggests that eight questions are relevant in determining whether information should be classed as personal data. Two of these, whether an

---

the user-supplied problem facts.' Graham Greenleaf, 'Legal Expert Systems — Robot Lawyers? (An Introduction to Knowledge-Based Applications to Law)' (August 12, 1989) Proc Australian Legal Convention, Darling Harbour, Sydney, August 1989 <https://ssrn.com/abstract=2263868> (accessed 2 July 2020) 5.

[6] ICO, 'What is personal data – A quick reference guide: Data Protection Act 1998' <https://ico.org.uk/media/for-organisations/documents/1549/determining_what_is_personal_data_quick_reference_guide.pdf>
(accessed 3 July 2020). Although this guidance was issued under the 'old regime', the new regime preserves the requirements that personal data should 'relate to' an individual and that the individual be identified or identifiable. The current guidance contains no flowchart but affirms that the factors set out in the flowchart are relevant. ICO, 'What is the meaning of 'relates to'?' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/what-is-personal-data/what-is-the-meaning-of-relates-to/#pd1> (accessed 3 July 2020).

[7] The Information Commissioner's Office (ICO) is the UK's regulatory authority for enforcement of the data protection regime.

[8] The flowchart is set out in Appendix 1.

[9] Michael Poulshock, 'Rule-Based Legal Information Systems' (July 1, 2010) <https://blog.law.cornell.edu/voxpop/2010/07/01/rule-based-legal-information-systems/> (accessed 3 July 2020).

[10] n 6.

individual is identifiable from the data, and whether the data 'relates to' an identifiable individual are drawn directly from the legislation. The remaining six questions are intended to tease out the meaning of 'relates to' in the context of the definition of personal data.

These questions become the backbone of the app. They are the variables that are relevant to the decision as to whether information is personal data. Users of the app are prompted to provide responses to the questions, that is, to provide the values of the variables (in this case by means of yes/no/unsure responses). The app delivers guidance to the user as to whether the user's information is personal data or not, based on the user's responses to the questions. The decision tree specifies rules-based if/then mappings between the values of the variables and the results.

For these purposes I created the Personal Data app decision tree, set out in Figure 1, closely modelling the rules embodied in the ICO's Guidance and flowchart.

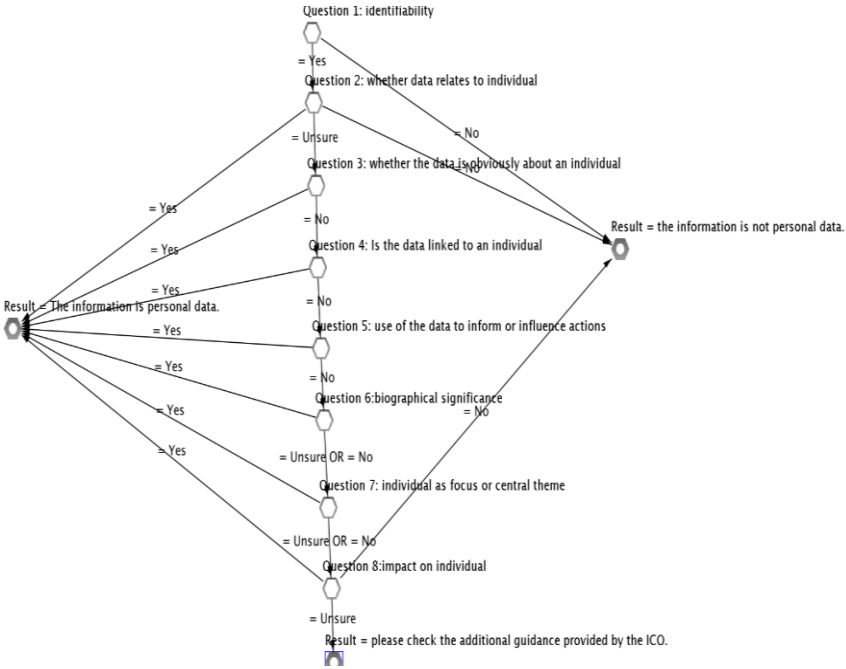

Figure 1: The Personal Data app decision tree created in Neota Logic

3. Choices in the process of building the app

3.1 Choices in building the app

The process of building a legal expert app obliges the author of the app to carefully consider the user's expectations, ease of use of the app, how the app should be designed so as to meet legal requirements,[11] and how to effectively limit liability.[12] However, the design of the decision tree, in particular, necessitates a close analysis of the rules the app is intended to apply. It does so because the design of the decision tree involves choices about the variables that are relevant to the decision-making process, the options for the values of those variables, and the range of possible results. Just as important, the process obliges the author of the app to reflect on the legal advice function, in particular *how* legal advice should incorporate and model an analysis of uncertainty or risk.

3.2 Choices in the design of the decision tree

Consider the ICO's flowchart.[13] It sets out eight questions that are relevant for determination of the question as to whether particular information is personal data. Questions 1 and 2 address the legislative provisions. Questions 3 to 8 are, in a sense, secondary.[14] They are intended to assist in the process of answering question 2, teasing out the implications of that question. In designing the decision tree for the app, an author *could* choose to incorporate only questions 1 and 2, ignoring the 'secondary' questions. It is plain, however, that the ancillary questions provided by the ICO are a useful aid, and that the app is the better for including questions 3 to 8 as variables. Indeed, it is these ancillary questions that embody legal expertise; a combination of guidance derived from case law and the regulator's own guidance on the meaning of 'personal data'.

---

[11] For example, having regard to the data protection regime, if the app is provided free of charge, it might be unnecessary and therefore unlawful to collect information about the name of the user.

[12] As regards limitation of liability, the designer of the app should consider whether liability should be limited by means of a contract between the app provider and the user or by means of a disclaimer. In either case, the app should be designed so as to secure an appropriate consent or acknowledgement from the user. For a discussion of risks in relation to expert systems see, for example, Joseph A Cannataci, 'Law, Liability and Expert Systems' I & Soc (1989) 3 AI and Society 169; William A Hyman, Waymon L Johnston, Steven Spar, 'Legal Liability and System Safety Applied to Expert Systems' (1989) 16(3) Computers and Industrial Engineering 355; George S Cole, 'Tort Liability for Artificial Intelligence and Expert Systems' (1990) 10 Computer L J 127.

[13] The flowchart is set out in Appendix 1.

[14] They are secondary in the sense that these factors are not explicitly set out in legislation. They embody interpretations of the term 'relates to'.

The process of designing the decision tree also draws attention to the fact that the ICO's Guidance and flowchart specify two possible answer values (a binary, yes/no option) for some of the question variables but three possible values (yes/no/unsure) for others. The decision tree for the Personal Data app replicates this choice. However, it is difficult to understand why the ICO should suppose that those faced with the problem of determining whether their data is personal data might be uncertain about their answers to questions 2,[15] 6 and 7 but not the other questions. It would seem to make more sense for the user to be given the option of selecting 'unsure' as the value for any of the question variables.

One might imagine that the Guidance and flowchart would make provision for only two possible results: personal data or not. In fact, the Guidance and flowchart suggest three further possible results: 'see the detailed ICO guidance';[16] 'the data is likely to be personal data' (in relation to questions 6 and 7) and 'the data is unlikely to be personal data' (in relation to question 8). I catered for the first of these in the app but not the second or third.

In principle, in framing a flowchart or a decision tree, one may choose to specify results that signify likelihood rather than certainty. However, it is not obvious why the ICO hedges its bets in relation to questions 6 and 7. After all, questions 6 and 7 draw upon the criteria articulated by the Court of Appeal in *Durant*.[17] Auld LJ has this to say about the criteria that may be relevant in the determination of whether information is personal data:

> It seems to me that there are two notions that may be of assistance. The first is whether the information is *biographical in a significant sense* ... The second is one of *focus*. The information should have the putative data subject as its focus rather than some other person with whom he may have been involved or some transaction or event in which he may have figured or have had an interest ... [18]

While many complained that the Court of Appeal adopted too narrow an interpretation of personal data, *Durant* is still authority for the proposition that data should be regarded as personal data where either of the two tests (biographical significance, focus) are met.[19] In designing the Personal Data app the writer opted to

---

[15] Note that while the ICO's flowchart suggests a binary answer to question 2, the body of the guidance makes it clear that the values of answers to question 2 include an 'unsure' option. It is this answer that leads the decision-maker through questions 3-8.

[16] This result is specified in the Guidance though omitted from the flowchart. It is incorporated in the Personal Data app as 'please check the additional guidance provided by the ICO'.

[17] *Durant v FSA* [2003] EWCA Civ 1746.

[18] *Durant* (n 17) [28] (italics added).

[19] *Durant* has not been expressly overruled. However, the lower courts have (implicitly or explicitly) accepted the view that the criteria in *Durant* are too narrow. Nevertheless, the courts have accepted that the criteria set out in *Durant* are relevant in determining whether data is

follow the lead of the Court of Appeal, specifying that information is personal data where the user provides a 'yes' response to either of questions 6 and 7.

The position in relation to question 8 is less clear-cut. The ICO makes it clear that questions 3 to 8 provide a *non-exhaustive* list of questions that might illuminate the answer to question 2 namely, whether the data 'relates to' the individual. It follows that information may 'relate to' an individual and qualify as personal data even where questions 3 to 8 elicit a negative response. Arguably therefore it makes sense to cater for a 'this information might be personal data' result. Note however that this is not what the ICO has done. Instead the ICO has opted for an 'unlikely to be personal data' result. One might speculate that although the ICO is prepared to accept that questions 3 to 8 do not capture every scenario in which information will 'relate to' an individual, the ICO considers that, in practice, the chances of that being so are low. It is not clear how the ICO might justify such a determination. In any event, it is questionable whether a 'this information is unlikely to be personal data' result is of practical use to those concerned to determine whether their information is personal data. In designing the app, the writer made no provision for such a result. If, indeed, questions 3 to 8 do not wholly exhaust the scenarios embraced by question 2, it might be better, and simpler, to include a caveat to that effect in the text displayed to the user within the app as well as in any report provided to the user.

3.3 Choices: a summary

The process of building the decision tree for the Personal Data app focused attention on the content and import of the relevant legal rules. It revealed that the ICO itself made certain choices in formulating the flowchart. It showed that far from being a simple mapping exercise (fact-rule combinations to binary outputs) the task of designing a decision tree involves choices, abstractions of the process of providing legal advice.

4. Reflections on building the app

4.1 A different kind of engagement with legal rules

The process of building an app that is intended to capture the implications of legal rules necessitates a deep engagement with the rules. It also obliges the lawyer-developer to engage with those rules in a manner that is untypical.

---

personal data. *R (on the application of Kelway) v Upper Tribunal (Administrative Appeals Chamber)* [2013] EWHC 2575 (Admin); *Rudd v Bridle* [2019] EWHC 893 (QB)[2019].

Lawyers do think deeply and systematically about the meaning and application of legal rules.[20] When analysing the implications of particular legal rules for situations faced by their clients they may implicitly or explicitly engage in if/then argumentation.[21] However, lawyers do not typically explicitly construct algorithmic representations of the relevant rules and their implications, and will often rely on experience and heuristics[22] to address uncertainty both with regard to the relevant facts and the import of the relevant rules.[23]

Algorithmic representations, of the kind embodied in flowcharts or decision trees, not only model decision rules but the relationship between the various rules. They also embody and project a particular model of uncertainty whether in relation to the presence or absence of relevant facts or the outcome based on a set of facts.[24] By creating an explicit model, a lawyer-developer must 'Identify the points of uncertainty and the type or range of alternative outcomes at each point.'[25] This kind of engagement with the legal rules occupies a middle ground between an analysis of the import of the rules in the abstract, and an analysis that is specific to a particular fact situation. It attempts to determine, ex ante, where uncertainty lies.

4.2 The benefits of a different kind of engagement with legal rules

4.2.1 A user-centred perspective

---

[20] Greenleaf acknowledges the range of skills that lawyers must possess in order to carry out legal work including 'logical reasoning', 'interpretative skills' and 'research skills'. Greenleaf (n 5).

[21] Lipshaw suggests that lawyers convert 'real-world narratives' into 'a series of if-then propositions.' Jeffrey Lipshaw, *Beyond Legal Reasoning: a Critique of Pure Lawyering* (Routledge 2017) 4.

[22] Susskind notes the reliance of experts on heuristics and experience. Richard Susskind, Expert Systems in Law and the Data Protection Adviser (1987) 7(1) *Oxford Journal of Legal Studies* 145. See also Edwina L Rissland 'Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning' (1990) 99(8) The Yale Law Journal 1957, 1966.

[23] Darlington notes that expert systems may have to handle two kinds of uncertainty: user uncertainty as to the answer to questions posed by the system and uncertainty as to the conclusion of a rule. Keith Darlington, *The Essence of Expert Systems* (Prentice Hall 2000) 79. Stevens et al. alert to the twin problem of factual and legal uncertainty in modelling legal expert systems. Charles Stevens, Vishal Barot and Jenny Carter, 'The Next Generation of Legal Expert Systems - New Dawn or False Dawn?' in M Bramer, M Petridis, and A Hopgood (eds), *Research and Development in Intelligent Systems* XXVII SGAI 2010 (Springer-Verlag 2010).

[24] Darlington (n 23) 78-93.

[25] John F Magee, 'Decision Trees for Decision Making' Harvard Business Review July-August 1964 https://hbr.org/1964/07/decision-trees-for-decision-making (accessed 1 July 2020). My choices in modelling uncertainty In the case of the personal data app were constrained by the fact that I intended to rely on the Information Commissioner's flowchart as the basis for the decision tree for the app. Nevertheless, I made choices both with regard to the points of uncertainty (which questions generated binary yes/no responses and which yes/no/unsure responses) and the range of alternative outcomes (this information is personal data/this information is not personal data/consult the ICO detailed guidance).

The process of constructing the decision tree for the app, by its nature, encourages a user-centred perspective. It obliges the lawyer-developer to reflect on how easy or otherwise it may be to give a binary yes/no response to the questions that drive the decision making workflow. In this respect it compels the lawyer-developer to step into the shoes of the client. Such a user-centred perspective may deliver a deeper appreciation of the way in which clients experience the impact of relevant rules and/or guidance and contribute to a critical assessment of those rules.

4.2.2 A glimpse into the mind of the regulator

Many regulators in the UK and other jurisdictions produce flowcharts intended to provide guidance on regulatory compliance.[26] By exploring a flowchart with an eye to its automation in the form of an app, the lawyer-developer must devote attention to the *regulator's* assumptions and choices about 'the points of uncertainty and the type or range of alternative outcomes at each point.'[27] In the case of the personal data app this revealed some surprising assumptions and choices on the part of the regulator, most obviously that the ICO appears to have reservations about the guidance produced by the Court of Appeal in *Durant*[28] as to the meaning of personal data, while seeming to have no similar reservations about its own guidance.[29] Information about the assumptions relied on by regulators is useful to advisers especially where those assumptions are not explicitly and transparently revealed in textual guidance. The exercise may also afford a basis for a critique of the flowchart (and accompanying

[26] For example, the Pensions Regulator provides flowcharts to assist users in determining whether there has been a breach of the law relating to pension schemes. The Pensions Regulator, 'Complying with the duty to report breaches of the law' <https://www.thepensionsregulator.gov.uk/en/document-library/code-related-guidance/complying-with-the-duty-to-report-breaches-of-the-law> (accessed 1 July 2020). The Department for the Environment Food and Rural Affairs (DEFRA) provides several flow charts and decision trees relating to the legislative requirement to apply a 'waste hierarchy'. DEFRA 'Guidance on applying the waste hierarchy to hazardous waste' <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/69457/pb13687-hazardous-waste-hierarchy-111202.pdf> (accessed 1 July 2020). The Independent Bankers Association of Texas (IBAT) provides various decision trees relevant to regulatory compliance. IBAT, 'Decision Trees & Charts' <https://www.ibat.org/decision-treesflow-charts> (accessed 1 July 2020).

[27] n 25.

[28] n 17.

[29] The ICO offers a 'the data is likely to be personal data' conclusion where the user supplies an affirmative response to questions 6 and 7 in the ICO flowchart. These questions, in essence, ask if the criteria set out in *Durant* are met. In contrast, if the user supplies an affirmative response to questions 3, 4 and 5, the ICO flowchart offers a 'the data is personal data for the purposes of the DPA' conclusion. Questions 3, 4 and 5 are based on criteria supplied by the ICO, drawing on its interpretation of the relevant legislative provisions.

guidance) provided by the regulator where, as here, the assumptions and choices baked in to the regulator's flowchart appear to underplay or mask uncertainty.[30]

5. Musings at the intersection of legal practice and legal theory

Legal theorists have variously applauded and denigrated attempts to model legal decision making through the use of algorithms. Susskind writes that

> the successful construction of expert systems in law will be of profound theoretical and practical importance to all those whose concern is the law.[31]

Leith, by contrast, is sceptical as to the commercial viability and the soundness of the theoretical underpinnings of legal expert systems.[32] The process of building an app offers the lawyer-developer the opportunity to reflect on the claims of theorists, and to contribute to the debate about the strengths and limitations of legal expert apps.

It would be for the market to determine whether the Personal Data app is commercially viable but it has value: it incorporates the regulator's guidance as to the import of the legal rules around the meaning of personal data;[33] it automates the regulator's flowchart; it is easier to use than a flowchart or textual guidance; it can be made accessible online; at least in some cases it will provide users with a straightforward answer to whether their data is personal data.[34]

---

[30] Amato offers a critique of draft guidance issued by the US Food and Drug Administration in relation to clearance decisions for medical devices. Amato suggests that the flow chart accompanying the guidance was the 'foundation' for that guidance. Stephen F Amato, 'Regulatory strategies for biomaterials and medical devices in the USA; classification, design and risk analysis' in Stephen F Amato and Robert M Ezzell Jr (eds) *Regulatory Affairs for Biomaterials and Medical Devices* (Woodhead Publishing 2015) 32.

[31] Richard Susskind, 'Expert Systems In Law: A Jurisprudential Approach To Artificial Intelligence And Legal Reasoning' (1986) 49(2) MLR 168.

[32] Philip Leith, 'The Rise and Fall of the Legal Expert System' (2010) 1(1) European Journal of Law and Technology.

[33] It incorporates the guidance even though it makes different choices around modelling of uncertainty from those made by the ICO in its own flowchart.

[34] Stevens et al. suggest a range of benefits for the law firm that adopts legal expert systems including

- 'Faster delivery of legal advice'
- 'Liberation of fee earning time otherwise spent in the labour-intensive and repetitive tasks of taking instructions, carrying out legal research and giving advice'
- 'Increased productivity'
- 'Reduced dependence on transitory human expertise '
- 'Reduction in human error'

Despite the reservations of Leith and others, legal expert systems are 'widespread in use in a number of contexts'.[35] Many prestigious law firms have developed and provide legal expert apps using the Neota Logic platform.[36]

At the same time, the experience of building the Personal Data app also flags the potential limitations of such apps. In the absence of the ICO's guidance and flowchart, it might have been difficult to construct the decision rules for the app. In particular, it might have been difficult to frame questions designed to elicit information from the user as to whether or not data is personal data.

It is also difficult to know in what proportion of cases the app would deliver real value for a client. That is, it is hard to know how frequently clients might struggle to answer any of the questions with certainty, or how often clients' in-app responses to those questions would prompt the app to issue an equivocal answer to whether their data was personal data. The real commercial value of the app may be for law firms in alerting prospective clients to the fact that legal issues are rarely straightforward![37]

Relatedly and perhaps most importantly, in the context of building the app, it is not clear how the exercise of modelling uncertainty *should* be carried out. That is, it is unclear what considerations might justify giving users of the app a binary choice of yes/no answers to questions in some cases but not in others, or of the app issuing unequivocal answers in some cases but not in others. As a result, while this app appears to serve a useful purpose, Leith's criticism concerning the lack of theoretical underpinnings may yet have some bite.

6. A role for machine learning?

The feasibility of creating decision trees (with or without the assistance of a platform such as Neota Logic) depends to a considerable extent on the expertise possessed by or available to the person creating the tree.[38] However, even where the author of an app has access to relevant expertise, the process of designing a decision tree may

Stevens et al. (n 23).

[35] Kevin D Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (CUP 2017) 10.

[36] Legal expertise apps are widely used in commercial contexts. For a range of apps developed using Neota Logic see Neota Logic, 'Neota App Gallery" https://www.neotalogic.com/neota-logics-client-app-gallery/ (accessed 1 July 2020).

[37] It would be interesting to know how many legal expert apps are offered to clients or prospective free of charge for this reason.

[38] Andreas Müller and Sarah Guido, *Introduction to Machine Learning with Python* (O'Reilly Media 2017) 2. Müller and Guido provide a useful description of the skills, tools and software required for machine learning.

prove difficult on account of uncertainty whether in relation to the relevance of the facts or the import of the legal rules.
In such cases, machine learning applications might be of assistance.[39] Machine learning applications can be used to infer decision trees from datasets.[40] In theory, decision trees derived in this way might be used as the basis for the decision tree reasoning engine at the heart of a legal expert app.[41]

For example, Figure 2 shows a visualisation of the decision tree generated by a machine learning application from a dataset which I created.[42] The decision tree, like the decision tree created in Neota Logic, is intended to model rules for determining whether information is personal data. It differs from the decision tree created for the Personal Data app in that it specifies two rather than three results.[43] The machine learning application correctly inferred that of the eight variables in the dataset (the questions set out in the ICO's flowchart) only questions 1 and 2 (labelled 'a' and 'b' respectively on the graph) are determinative, questions 3 to 8 being subsidiary to question 2. It treats the remaining variables as redundant.

[39] Murthy notes several advantages of machine learning decision tree classification including circumvention of the 'bottleneck of acquiring knowledge from a domain expert'. Sreerama K Murthy, 'Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey' (1998) 2 Data Mining and Knowledge Discovery 345, 346.
[40] George Seif, 'A Guide to Decision Trees for Machine Learning and Data Science' Nov 30, 2018 <https://towardsdatascience.com/a-guide-to-decision-trees-for-machine-learning-and-data-science-fe2607241956> (accessed 1 July 2020). For a more detailed explanation of the process by which machine learning may be used to infer decision trees see J R Quinlan, 'Induction of Decision Trees' Machine Learning 1986 (1) 81.
[41] Shapiro outlines a role for machine learning in generating decision trees that may be used in expert systems. Alen David Shapiro, 'The Role of Structured Induction in Expert Systems' (DPhil thesis, University of Edinburgh 1983). Bonner et al. maintain that 'Sets of rules in the form of decision trees generated by machine learning techniques are a good starting point for further tasks aiming at developing a knowledge base.' Richard F Bonner, Violetta Galant, Mieczyslaw L Owoc, 'On Features of Decision Trees as a Technique of Knowledge Modelling' (Workshop on Computer Science and Information Technologies CSIT'99, Moscow, Russia, 1999).
[42] Using pandas, the writer created a dataset in which each item of data was represented in a series of yes/no responses to each of the eight questions set out in the ICO's flowchart. The data was labelled accordingly, as personal data or not. The decision tree was generated using methods available in scikit-learn and the graph generated by graphviz. The dataset is available on request.
[43] That is, the information is/is not personal data.

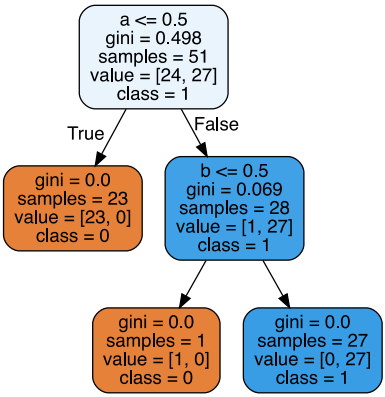

Figure 2: A visualisation of the decision tree generated from the writer's 'personal data' dataset

In this case, since the machine learning model is trained on a dataset whose features or variables are the criteria offered by the ICO for determining whether data is personal data, the model does little more than infer rules that are already known, offering, at best, a means of cross-checking the suitability of decision trees created manually.[44]

However, tasked with classifying data as falling within or outside some legal definition or rule, machine learning applications can do much more than infer decision trees that merely reflect the content of the relevant legal rules or guidance. The strength of machine learning lies in its ability to detect statistically relevant patterns in data.[45] Given sufficient labelled training data, an appropriate selection of features or variables, and a well-tuned machine learning model, the model could, in theory, infer a decision tree indicating which of the selected features is relevant to a particular legal

[44] Hunter criticises attempts to make machine learning applications (specifically neural networks) 'restate doctrinal rules' observing that this 'seems an inappropriate use of the technology' since it relies on statistics rather than rules. Dan Hunter, 'Out of their minds: Legal theory in neural networks' (1999) 7 Artificial Intelligence and Law 129, 135. See also Lyria Bennett Moses, Janet Chan, 'Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools' [2014] UNSWLawJl 25.

[45] Hunter (n 44) 135; Moses et al (n 44).

outcome.[46] For example, a machine learning model might infer that certain types of data, for example names, addresses, credit card information and so on are likely to be categorised as personal data.[47]

This is not to suggest that decision trees generated by machine learning applications will invariably be appropriate for use in a legal expert app. The suitability of a decision tree generated by machine learning will depend a range of factors including the quality and completeness of the dataset used to train the machine learning algorithm. Minor changes in the dataset used to train the machine learning algorithm may result in the generation of very different rules.[48] Decision trees inferred by machine learning algorithms do not always tally with the intuitions of experts.[49] Quinlan describes this factor as 'the chief obstacle to the use of induction for building large expert systems.'[50] In addition, 'uncertain data'[51] presents 'a great challenge for classification algorithms'[52] and so for the generation of decision trees from datasets containing such data.

Nevertheless, the possibility of a role for machine learning in generating decision trees for use in a legal expert app is appealing because it offers the prospect of a different approach to the problem of uncertainty. Whereas typically lawyers model uncertainty by reference to experience and heuristics, machine learning applications model uncertainty by reference to data, relying on statistics and probabilities to capture correlations.[53] These approaches need not be antagonistic. Machine learning

[46] Katz sees a role for the use of machine learning to predict legal outcomes. Daniel Martin Katz, 'Quantitative Legal Prediction—Or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry' (2013) 62 Emory Law Journal 909.

[47] Deep learning, one form of machine learning, has been used in data privacy classification. See for example, Griffin R Bishop, Harutyun Sadoyan, Leo Grande, Samuel John Pridotkas 'Deep Learning for Data Privacy Classification' (2018) <https://digitalcommons.wpi.edu/mqp-all/6616> (accessed 1 July 2020). However, in contrast to decision trees deep learning models are not readily interpretable.

[48] Ruey-Hsia Li, Geneva G Belford , 'Instability of decision tree classification algorithms' Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining 570. For example, the omission of a single sample from my dataset resulted in the generation of a very different decision tree. The sample that was omitted was the only sample where the value of the first question (identifiability) variable was positive and the value of all other question variables negative.

[49] Quinlan (n 40) 103, 104.

[50] Quinlan (n 40) 103, 104.

[51] Data is 'uncertain' when it is unclear how it should be classified.

[52] SB Kotsiantis, 'Decision trees: a recent overview' (2013) 39 Artificial Intelligence Review 261, 273.

[53] Ashley (n 35) 107. Denvir sees a role for machine learning in providing lawyers with a means for the 'quantification of uncertainty via probabilistic methods.' Catrina Denvir, 'Scaling the Gap' in Catrina Denvir (ed), *Modernizing Legal Education* (CUP 2019) 76, 77.

might be used to draw out and model the experience of and heuristics employed by lawyers.[54]

7. A role for legal expert apps in legal education?

Writing in 2013 Chambliss speculated that US law schools, in anticipation of market liberalisation, might 'scale up legal advising through a combination of legal technicians and legal expert systems'.[55] Whether in anticipation of market liberalisation or not, several universities worldwide have partnered with Neota Logic in order to offer students the opportunity to create legal expert apps on the platform. The University of Melbourne's Law School devotes an entire course to the creation of such apps.[56]

Such an approach seems far-sighted. Many prestigious law firms[57] have developed legal expert apps: an ability to create such apps is likely to be attractive to such firms. However, an opportunity, whether as part of pre- or post-qualifying legal education, to create and use such apps is also desirable for the reasons touched on in this article. Building the app not only teaches app development skills, it also deepens knowledge of the relevant rules, fosters a user-centred approach to advice, encourages critical appraisal of the rules and facilitates engagement with debates (as relevant to legal practice as legal theory) not only about the merits and demerits of legal expertise apps but the extent to which law can be formalised. The process is a valuable form of experiential learning or learning by doing.

8. Conclusion

My experience of building the app suggests that the exercise delivers benefits for lawyers (and prospective lawyers)

- as a means of motivating deep engagement with the relevant legal rules

[54] Langley and Simon describe research relating to the use of machine learning to produce a new set of diagnostic rules for an expert system. The rules generated by the machine learning system offered more accurate results than those 'hand-crafted' by an expert. The method 'used causal knowledge … gleaned from the expert, to constrain the rule-induction process' and the dataset on which the machine learning model was trained had been labelled by the expert. Pat Langley, Herbert A Simon, 'Applications of Machine Learning and Rule Induction' (1995) 38(11) Communications of the ACM 58.

[55] Elizabeth Chambliss, 'Law School Training for Licensed Legal Technicians - Implications for the Consumer Market' (2014) 65 SCL Rev 579.

[56] The course is titled 'Law Apps'. The University of Melbourne 'Law Apps', https://law.unimelb.edu.au/students/jd/enrichment/pili/subjects/law-apps (accessed 1 July 2020).

[57] Allen and Overy, Clifford Chance, Freshfields, Herbert Smith Freehills, Weightmans are among the law firms that have developed apps using Neota Logic.

- as a way of compelling the lawyer-developer to step into the client's shoes for the purposes of assessing and modelling uncertainty
- as a source of insights and reflection

Of the various insights and reflections engendered by the process of building the Personal Data app three stand out. All three are related.

The first relates to the difficulty in modelling uncertainty in building the app. It is not clear how uncertainty *should* be modelled: the lawyer-developer has choices in modelling uncertainty both in relation to the relevant facts and the import of the relevant legal rules.

This suggests that even now there is a need for further examination of issues that lie at the intersection of legal practice and legal theory: about the extent to which law in action can be modelled as a set of rules; about uncertainty in applying legal rules to instant cases; about how uncertainty can or should be modelled in legal expert apps; about the assumptions that play into choices about modelling uncertainty.

The second is that building an app that is based on a regulator's flowchart highlights the regulator's approach to uncertainty (at least so far as expressed in the flowchart). This was a revelation. Aside from illuminating those choices, such insights might conceivably offer a basis for a challenge to the regulator's approach.

The third is the reflection that machine learning models might be of assistance in generating decision trees for a legal expert app, offering a data-driven approach to modelling legal uncertainty. Such speculation is not new, but the ready availability of open source machine learning models makes it possible to experiment with such models.[58]

Of course, in building a functioning app you also learn to do just that. It is challenging and, dare I say it, it is fun. It is no surprise then that universities have been keen to provide students with the opportunity to create legal expert apps: much is learned in the process.

Bibliography

Amato, S.F., 'Regulatory strategies for biomaterials and medical devices in the USA; classification, design and risk analysis' in Stephen F Amato and Robert M Ezzell Jr (eds) *Regulatory Affairs for Biomaterials and Medical Devices* (Woodhead Publishing 2015)

[58] Platforms such as kaggle, H_2O.ai, Colaboratory and fast.ai make machine learning accessible with free courses and software libraries.

Ashley, K.D., *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (CUP 2017)

Bennett Moses, L., Chan, J., "Using Big Data for Legal and Law Enforcement Decisions: Testing the New Tools" [2014] UNSWLawJl 25

Bishop, G. R., Sadoyan, H., Grande, L., & Pridotkas, S. J. Deep Learning for Data Privacy Classification (2018) <https://digitalcommons.wpi.edu/mqp-all/6616> (accessed 1 July 2020)

Bonner, R.F., Galant, V., Owoc, M.L. 'On Features of Decision Trees as a Technique of Knowledge Modelling' (Workshop on Computer Science and Information Technologies CSIT'99, Moscow, Russia, 1999)

Cannataci, J A., 'Law, Liability and Expert Systems' I & Soc (1989) 3 AI and Society 169

Chambliss, E., 'Law School Training for Licensed Legal Technicians - Implications for the Consumer Market' (2014) 65 SCL Rev 579

Cole, G.S., 'Tort Liability for Artificial Intelligence and Expert Systems' (1990) 10 Computer L.J. 127

Darlington, K., *The Essence of Expert Systems* (Prentice Hall 2000)

Denvir C., 'Scaling the Gap' in Catrina Denvir (ed), *Modernizing Legal Education* (CUP 2019)

Greenleaf, G.,'Legal Expert Systems — Robot Lawyers? (An Introduction to Knowledge-Based Applications to Law)' (August 12, 1989) Proc Australian Legal Convention, Darling Harbour, Sydney, August 1989 <https://ssrn.com/abstract=2263868> (accessed 2 July 2020)

Hunter, D., 'Out of their minds: Legal theory in neural networks' (1999) 7 Artificial Intelligence and Law 129

IBAT, 'Decision Trees & Charts' <https://www.ibat.org/decision-treesflow-charts> (accessed 1 July 2020

ICO, 'What is personal data – A quick reference guide: Data Protection Act 1998' <https://ico.org.uk/media/for-organisations/documents/1549/determining_what_is_personal_data_quick_reference_guide.pdf>
(accessed 4 July 2019)

Katz, 'Quantitative Legal Prediction—Or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry' (2013) 62 Emory Law Journal 909

Kotsiantis S. B. 'Decision trees: a recent overview' (2013) 39 Artificial Intelligence Review 261

Langley, P., Simon, H.A. 'Applications of Machine Learning and Rule Induction' (1995) 38(11) Communications of the ACM 58.

Law Society of England and Wales, 'LawTech Delivery Panel' (11 October 2018) <https://www.lawsociety.org.uk/support-services/lawtech/articles/lawtech-delivery-panel/>
(accessed 4 July 2019)

Law Society of Scotland, 'Law Society set to drive innovation in legal technology' (10 October 2018)
<https://www.lawscot.org.uk/news-and-events/news/lawscottech-launch/>
(accessed 4 July 2019)

Leith, P., 'The Rise and Fall of the Legal Expert System' (2010) 1(1) European Journal of Law and Technology.

Li, R., Belford, G.G., 'Instability of decision tree classification algorithms' KDD 2002 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (ACM 2002) 570 <http://portal.acm.org.ezproxy.lib.gla.ac.uk/toc.cfm?id=775047>

Lipshaw, J., *Beyond Legal Reasoning: a Critique of Pure Lawyering* (Routledge 2017)

Lupica, L.R., Franklin, T.A., and Friedman, S.M., 'The Apps for Justice Project: Employing Design Thinking to Narrow the Access to Justice Gap' (2017) 44 Fordham Urb L J 1363 <https://ir.lawnet.fordham.edu/ulj/vol44/iss5/> (accessed 1 July 2020).

Magee, J.F., 'Decision Trees for Decision Making' Harvard Business Review July-August 1964 https://hbr.org/1964/07/decision-trees-for-decision-making (accessed 1 July 2020)

Müller, A., Guido, S., *Introduction to Machine Learning with Python* (O'Reilly Media 2017)

Murthy, S.K., 'Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey' (1998) 2 Data Mining and Knowledge Discovery 345

Neota Logic, 'Neota App Gallery" https://www.neotalogic.com/neota-logics-client-app-gallery/ (accessed 1 July 2020)

Neota Logic, 'No-Code Development' <https://www.neotalogic.com/platform/no-code-development/> (accessed 5 July 2019)

Poulshock, M., 'Rule-Based Legal Information Systems' (July 1, 2010) <https://blog.law.cornell.edu/voxpop/2010/07/01/rule-based-legal-information-systems/> (accessed 3 July 2020).

Quinlan, J.R., 'Induction of Decision Trees' 1986 (1) Machine Learning 81

Rissland, E.L., 'Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning' (1990) 99(8) The Yale Law Journal 1957

Seif, G., 'A Guide to Decision Trees for Machine Learning and Data Science' Nov 30, 2018 <https://towardsdatascience.com/a-guide-to-decision-trees-for-machine-learning-and-data-science-fe2607241956> (accessed 4 July 2019)

Shapiro, A.D., 'The Role of Structured Induction in Expert Systems' (DPhil thesis, University of Edinburgh 1983)

Stevens, C., Barot V., and Carter, J., 'The Next Generation of Legal Expert Systems - New Dawn or False Dawn?' in Bramer, M., Petridis, M., and Hopgood A., (eds) *Research and Development in Intelligent Systems* XXVII SGAI 2010 (Springer-Verlag 2010)

Susskind, R., 'Expert Systems In Law: A Jurisprudential Approach To Artificial Intelligence And Legal Reasoning' (1986) 49(2) MLR 168.

Susskind, R., 'Expert Systems in Law and the Data Protection Adviser' (1987) 7(1) Oxford Journal of Legal Studies 145.

The Pensions Regulator, 'Complying with the duty to report breaches of the law' <https://www.thepensionsregulator.gov.uk/en/document-library/code-related-guidance/complying-with-the-duty-to-report-breaches-of-the-law> (accessed 1 July 2020)

The University of Melbourne 'Law Apps', https://law.unimelb.edu.au/students/jd/enrichment/pili/subjects/law-apps (accessed 1 July 2020).

Tsang, S., Kao, B., Yip, K.Y., Ho, W., and Lee, S.D. 'Decision Trees for Uncertain Data' (2011) 23 (1) IEEE Transactions on Knowledge and Data Engineering 64

Appendix 1

Flowchart

1.	**Can a living individual be identified from the data, or, from the data and other information your possession, or likely to come into your possession?**	
	Yes	Go to question 2.
	No	The data is not personal data for the purposes of the

2.	**Does the data 'relate to' the identifiable living individual, whether in personal or family life, business or profession?**	
	Yes	The data is 'personal data' for the purposes of the DPA.
	No	The data is not 'personal data' for the purposes of the DPA.

3.	**Is the data 'obviously about' a particular individual?**	
	Yes	The data is 'personal data' for the purposes of the DPA.
	No	Go to question 4.

4.	**Is the data 'linked to' an individual so that it provides particular information about that individual?**	
	Yes	The data is 'personal data' for the purposes of the DPA.
	No	Go to question 5.

5.	**Is the data used, or is it to be used, to inform or influence actions or decisions affecting an identifiable individual?**	
	Yes	The data is 'personal data' for the purposes of the DPA.
	No	Go to question 6.

6.	**Does the data have any biographical significance in relation to the individual?**	
	Yes	The data is likely to be personal data for the purposes of the DPA.
	No	Go to question 7.
	Unsure	Go to question 7.

7.	**Does the data focus or concentrate on the individual as its central theme rather than on some other person, or some object, transaction or event?**	
	Yes	The data is likely to be personal data for the purposes of the DPA.
	No	Go to question 8.
	Unsure	Go to question 8.

8.	**Does the data impact or have the potential to impact on an individual, whether in a personal, family, business or professional capacity?**	
	Yes	The data is 'personal data' for the purposes of the DPA.
	No	The data is unlikely to be 'personal data'.