Computer Systems Fit for the Legal Profession?

Delacroix, Sylvie

DOI:
10.2139/ssrn.3158132
10.1080/1460728x.2018.1551702

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Delacroix, S 2018, 'Computer Systems Fit for the Legal Profession?', Legal Ethics.
https://doi.org/10.2139/ssrn.3158132, https://doi.org/10.1080/1460728x.2018.1551702

Routledge
Taylor & Francis Group

Legal Ethics

ISSN: 1460-728x (Print) 1757-8450 (Online) Journal homepage: http://www.tandfonline.com/loi/rlet20

Computer systems fit for the legal profession?

Sylvie Delacroix

To cite this article: Sylvie Delacroix (2018): Computer systems fit for the legal profession?, Legal Ethics, DOI: 10.1080/1460728x.2018.1551702

To link to this article: https://doi.org/10.1080/1460728x.2018.1551702

Published online: 12 Dec 2018.

LEGAL ETHICS
https://doi.org/10.1080/1460728x.2018.1551702

Routledge
Taylor & Francis Group

Computer systems fit for the legal profession?

Sylvie Delacroix [a,b]

[a]Law, University of Birmingham, Birmingham, UK; [b]The Alan Turing Institute, London, UK

ABSTRACT

This essay aims to contribute robust grounds to question the Susskinds' influential, consequentialist logic when it comes to the legitimacy of automation within the legal profession. It does so by questioning their minimalist understanding of the professions. If it is our commitment to moral equality that is at stake every time lawyers (fail to) hail the specific vulnerability inherent in their professional relationship, the case for wholesale automation is turned on its head. One can no longer assume that, as a rule, wholesale automation is both legitimate and desirable, provided it improves the quality and accessibility of legal services (in an accountable and maximally transparent way). The assumption, instead, is firmly in favour of designing systems that better enable legal professionals to live up to their specific responsibility. The rest of the essay outlines key challenges in the design of such profession-specific, 'ethics aware' decision-support systems.

KEYWORDS

Professions; vulnerability; expert systems; prediction tools; wholesale automation; augmentation; professional responsibility; automation bias; Susskind; moral equality; expertise; machine learning

The fast expanding reach and prowess of computer systems has for a while now led some to ponder when, if at all, computers might replace humans and in what capacities. Others seek – more wisely – to grasp the reach and depth of the transformations that are already well underway: new tools commonly end up changing not just the nature of the problems they are meant to solve, but also the tool-users themselves. As such, the considerable changes that have already been brought about by the development of 'smart' technologies in the last fifty years are not that remarkable, save for one thing: their sheer speed.

Today, the habits generated by technologies introduced a decade ago are so deeply ingrained that many would not be able to contemplate a life without them. Mobile, connected devices have not only changed the way we make friends, say. They are also changing our very understanding of what friendship stands for, what we can expect from our friends, and what they can expect from us. Could the same be said of the way computer systems are increasingly being deployed in professional contexts? Are these systems about to change our very understanding of what the legal profession stands for, what we can expect from it and what it can expect from us?

The Susskinds' *The Future of the Professions* is an important book, not least because it forces us to tackle the implications of recent progress in our ability to augment or

CONTACT Sylvie Delacroix s.delacroix@bham.ac.uk

automate central, 'professional' tasks. The Susskinds rightly denounce the professions' head-in-the-sand response to the issues at stake: recent advances in our ability to extract knowledge from professions-relevant data (thanks in part to novel natural language processing techniques) have already started to revolutionise the way professionals work. Rather than drag our feet or stare in disbelief we should, according to the Susskinds, actively embrace the chance to make professional expertise more affordable and accessible. The latter, consequentialist mantra, is repeated throughout the book. Its simplicity is made possible by an important assumption: there is no particular value which the professions, in contrast to other expert service providers, 'stand for'.

If the *Future of the Professions'* influence is left to grow unchallenged, the Susskinds may well be proven right. In the same way technological developments are changing our very understanding of friendship, the systematic, efficiency-driven deployment of automated systems within the professions may well turn the professions into 'mere' expert service providers for good. In the meantime, to challenge the latter conclusion requires a level of critical engagement that is currently lacking. This paper hopes to remedy this: while it shares the conviction that computer systems will play an essential role within the legal profession, and that this could transform it for the better, this paper unpacks key hurdles on the way to the latter, normative conclusion. Before reviewing possible uses (and abuses) of such systems within the legal profession (Section 2), Section 1 outlines a critical understanding of the *raison-d'-être* underlying the professions as an institution.

1. A normative understanding of the professions: computer systems' design and deployment constraints

One may, as a social scientist, acknowledge the professions as a historically rooted and constantly evolving institution and limit oneself to recording those transformations and possibly predicting future ones. A superficial reading would lead one to argue that it is precisely what the Susskinds have endeavoured to do, based on a timely analysis of the likely impact of the professions' widespread reliance on increasingly capable machines. The problem is: there is no such thing as a purely descriptive account of institutions. The very delineating of that institution's reach necessarily relies on some conceptual analysis. Most importantly, the Susskinds' explicitly normative judgment as to the positive impact of technology-induced transformations presupposes some kind of functional analysis of the 'professions' as an institution. And it is the Susskind's minimalist and strictly instrumental analysis of the professions that underlies their normative conclusions.

Instead of the superficial reading mentioned above, the Susskind's normative conclusions can be seen as the result of a – piecemeal – genealogical account of the professions. A genealogy necessarily marries historical investigation and functional analysis. In asking 'why do we have this or that institution?', a genealogy presupposes that the object it studies can meaningfully be treated as *functional*, that is, as serving an end other than itself.

Along this line, Susskind and Susskind refer, among other things, to Terence Johnson's historically informed critique of professionalism[1] to debunk the still widely influential, traditional account of the professions as 'devoted to the service of the public, above and

[1]T Johnson, 'Imperialism and the Professions: Notes on the Development of Professional Occupations in Britain's Colonies and the New States' (1972) 20 *The Sociological Review* 281.

beyond material incentives'.[2] Terence Johnson indeed denounces professionalism as a mechanism for protecting occupational power through a mystification process. Far from serving the public interest, the professions, on that account, only serve to consolidate certain occupations' – lucrative – monopoly over the provision of particular services. The knowledge asymmetry that triggers the need for such services is exploited (rather than compensated) so as to make any critical assessment of their services beyond reach. Johnson's historically informed critique clearly contributes to the Susskinds' 'failure' verdict: the professions do not serve the public interest that their justificatory rhetoric claims to serve, given their failure to deliver services that are affordable, of good quality, and accountable. The Susskinds' normative conclusion is to embrace the radical transformation promised by increasingly widespread reliance on automated systems within the professions.

Now, if instead of starting from the 'devotion to public service' functional hypothesis (which informs the Susskinds' normative conclusions), one were to start from a different answer to the 'why do we have this institution' question, one that highlights the need for particularly stringent norms of ethical integrity within certain occupations, one would get a different story. While it is not possible, within this essay, to back this up with the required historical investigations, it is likely that a genealogical critique driven by such a functional interpretation would lead to different conclusions. It might be that the historical processes that brought about the professions as an institution are only partly related to the ideal of ethical integrity as it features today in the professions' justificatory rhetoric. Yet it is the case that, within certain occupations, there are specific ethical challenges that are qualitatively different from those entailed by the knowledge asymmetry that characterises the provision of all expert services.

In a bid to expose the risks concomitant with the progressive conflation of *professional* responsibility with that of expert service providers in general, I highlight in a separate paper[3] the specific vulnerability inherent in the lay-professional relationship. The difference between the latter and the vulnerability concomitant with the hire of a mountain guide (or car mechanic) is not one of degree: when our life is at stake on the mountain side we are probably as vulnerable as can be. Independently of this primary vulnerability, however, the lay-professional relationship can provide fertile ground for inferiorising treatment that is wrong not because it violates some norm of fairness but rather because it threatens our commitment to moral equality: our equal moral worth as individuals independently of any contingent traits or status.

Be it through objectification or infantilisation, the vulnerability inherent in the circumstances that prompt recourse to a professional can all too easily be exploited in a way that compromises a lay person's ability to meaningfully contribute to the way she projects her sense of self, both socially and through her body. To retain a sense of 'authorship' over one's process of self-construction indeed requires that there be, to a minimal degree, some movement of to and fro between the process of definition of our 'self' *from without* (human encounters or environmental constraints) and *from within* (the way we appropriate these encounters or constraints). This to and fro movement is never easy. In some circumstances, it can be become particularly challenging: just like the person who has been diagnosed with a

[2]MS Larson and MS Larson, *The Rise of Professionalism: A Sociological Analysis*, Vol 233 (University of California Press 1979).
[3]See S Delacroix, *A Vulnerability-Based Account of Professional Responsibility* (2018) <https://papers.ssrn.com/abstract=2840864>.

grave illness, the person who is accused of murder will struggle to retain the sense that she may contribute anything to the self she projects, other than the institutionally imposed 'murderer' label. Outside criminal law, other circumstances such as sudden poverty or divorce proceedings can also affect one's sense of authorship over that process of self-definition: one may 'no longer know how to continue, given who one was'.[4]

Interestingly, this vulnerability-based account finds an echo in what Susskind and Susskind call the 'disempowerment' charge:

> our professions, as presently organized, often discourage self-help, self-discovery, and self-reliance; and they can unnecessarily inhibit or even alienate individuals who, once equipped with better insight, would benefit from engaging and participating more directly in their problems.[5]

Susskind and Susskind frame the above concern as a psychological one. Yet -as Sangiovanni brilliantly suggests in his *Humanity without Dignity*[6] – one may argue that when disempowerment prevents a person from being able to play an active role in the deployment of her sense of self (as is the case in paradigmatic instances of social cruelty, such as slavery, rape or torture), a fundamental value – moral equality – is under threat.[7]

If such a key value is indeed at stake in the way the professions operate (this is not a premise I seek to demonstrate a priori[8]), one cannot help but be concerned by the extent to which the special degree of responsibility it entails (and the non-utilitarian framework it demands) is bulldozed out of the range of relevant considerations. To those who worry about 'the loss of trustworthy institutions', meant to 'protect ourselves from exploitation by unscrupulous quacks', Susskind and Susskind retort:

> [The professions'] members claim that they are not simply reliable but are also people of upstanding character and motivated by non-selfish interests. For many observers and providers, this strong sense of trust is an indispensable feature of professional work. It is important that professionals are of outstanding moral character, and put the interests of the recipients of their work ahead of their own. […] The trust objection suggests that the professions, and our ability to trust in them in the strong sense, are the only way to resolve our fundamental challenge (that we all have problems for which we do not personally have the expertise to resolve). Yet we think this is mistaken. Our primary need is only for a reliable outcome.[9]

The underlined sentence in the above passage encapsulates a fundamental problem in Susskind and Susskind's argument: it is expertise in general – not the professions – that is our answer to what Susskind and Susskind call our 'fundamental challenge' (i.e. that none of us has the knowledge necessary to be able to deal with every one of our needs or problems). This should be obvious – so far nobody is suggesting that hairdressers, carpenters or indeed

[4]A Sangiovanni, *Humanity Without Dignity: Moral Equality, Respect and Human Rights* (Havard UP 2017).
[5]R Susskind and D Susskind, *The Future of the Professions: How Technology Will Transform the Work of Human Experts* (Oxford University Press 2015).
[6]Sangiovanni (n 4).
[7]The conceptual link between our commitment to moral equality and the type of responsibility that stems from the specific vulnerability inherent in the lay-professional relationship is articulated at length in Delacroix (n 3).
[8]Depending on the way professionals operate (exacerbating or moderating the particular vulnerability concomitant with the need for legal or other services), this key value- moral equality- is threatened (or not). Why? Because exploiting someone's vulnerability to compromise her ability to deploy her sense of self is a form of social cruelty (as per Sangiovanni), that is different only in degree from the most paradigmatic instances, such as rape, slavery or torture. So my argument does not start from the premise that the professions have a commitment to moral equality, but posits instead that the professions' way of operating cannot help but have an impact on the extent to which this fundamental value is upheld.
[9]Susskind and Susskind (n 5), my emphasis.

mountain guides should be counted as members of the professions. So why do the Susskinds repeatedly seek to level down the difference between the professions and experts by emphasising that they both answer the same 'knowledge problem'? Strictly speaking, they do both answer that problem, but to repeatedly articulate our concept of the professions solely by reference to that common denominator is a sure way of ridding it of any substance.

It may be unfair to argue that this precisely the Susskind's agenda.[10] Yet one gets the sense that, for the Susskinds, the claim to ethical integrity that most deem to be an essential part of our concept of the professions is but a contingent, historically rooted claim. While it still plays a role in the professions' justificatory rhetoric,[11] that claim can be shown to have increasingly little in the way of empirical evidence to back it up.[12]

There is indeed no lack of empirical evidence to support the Susskinds' key verdict – reiterated throughout *The Future of the Professions*: our professions are failing. They are 'by and large, […] unaffordable, under-exploiting technology, disempowering, ethically challengeable, underperforming, and inscrutable'.[13] This is a hefty and seemingly comprehensive charge-list. Yet when one looks at the narrative behind each of these charges, one finds that they mostly (except for the – notable – disempowering aspect) fall under a broadly utilitarian outlook on the professions. That outlook can be summarised under point 1 below:

> 1: the professions do not serve the public interest they claim to serve, given their failure to deliver services[14] that are affordable,[15] of good quality,[16] and accountable.[17]

[10]The Susskinds devote substantial parts of their book to discussing various accounts of the professions.

[11]'When we consider why the professions established their reputations for trustworthiness in the first place, they likely did so to meet this primary concern. Put another way, they established a reputation for trustworthiness not as an end in itself, but as a useful way to signal their reliability to others'.

[12]B Keogh, *Review into the Quality of Care and Treatment Provided by 14 Hospital Trusts in England: Overview Report* (NHS 2013); T Lagu and others, 'A Mixed-Methods Analysis of Patient Reviews of Hospital Care in England: Implications for Public Reporting of Health Care Quality Data in the United States' (2013) 39 *Joint Commission Journal on Quality and Patient Safety* 7; KMJMH Lombarts and others, 'Measuring Professionalism in Medicine and Nursing: Results of a European Survey' [Public Library of Science] (2014) 9 *PLoS ONE* e97069; R Moorhead, 'Lawyer Specialization–Managing the Professional Paradox' (2010) 32 *Law & Policy* 226; R Moorhead, 'Precarious Professionalism: Some Empirical and Behavioural Perspectives on Lawyers' (2014) 67 *CLP* 447; R Moorhead, A Sherr and A Paterson, 'Contesting Professionalism: Legal Aid and Nonlawyers in England and Wales' (2003) 37 *Law & Society Review* 765; R Moorhead and others, *Quality and Cost: Final Report on the Contracting of Civil, Non-Family Advice and Assistance Pilot* (The Stationary Office 2001); MJ O'Fallon and KD Butterfield, 'A Review of the Empirical Ethical Decision-Making Literature: 1996–2003' (2005) 59 *Journal of Business Ethics* 375; A Paterson and A Sherr, 'Quality, Clients and Legal Aid' (1992) 142 *New Law Journal* 783; A Sherr, R Moorhead and A Paterson, *Lawyers-the Quality Agenda Vol. 1 Assessing and Developing Competence and Quality in Legal Aid; the Report of the Birmingham Franchising Pilot* (HMSO 1994).

[13]Susskind and Susskind (n 5). In the legal domain, see e.g. SJ Harper, *The Lawyer Bubble : A Profession in Crisis* (Basic Books 2013).

[14]Because of its intrinsic link to the affordability and quality of the services delivered, the failure to exploit up-to-date technologies charge can be incorporated into the affordability and quality charges.

[15]'Most people and organizations cannot afford the services of first-rate professionals; and most economies are struggling to sustain most of their professional services, including schools, court systems, and health services' Susskind and Susskind (n 5).

[16]'The fifth problem with the professions is that they underperform. This is not to suggest that the professions invariably achieve low levels of attainment. Rather, we maintain that in most situations in which the professions' help is called for, what is made available may be adequate, good, or even great, but rarely is it world-class'; ibid. In the legal domain, there is a growing body of empirical literature, reviewed in detail – mostly in the British context – in Moorhead, 'Precarious Professionalism: Some Empirical and Behavioural Perspectives on Lawyers' (n 12), that paints an even bleaker picture than that suggested by Susskind and Susskind (n 5).

[17]'Recipients of professional services, often by the nature of the arrangement, are able, neither to evaluate the substance of the guidance they receive nor to judge whether a given profession is best placed to undertake the work. Sometimes, of course, the problem being solved or the work being undertaken is so complex that no lay person could hope to grasp what is going on. But there are occasions, no doubt, when there is intentional obfuscation, to justify high fees, perhaps, or for straightforward self-aggrandizement. Where there is opacity and mystification, there will be mistrust and a lack of accountability'; Susskind and Susskind (n 5).

There is little doubt that the growing availability and sophistication of automated systems could have a positive impact on the professions' ability to better honor the so-called 'grand bargain' that 'grants professionals both their special status and their monopolies over numerous areas of human activity'.[18] There is clear potential for those systems to dramatically improve both the affordability and quality of the services delivered by the professions. Yet Susskind and Susskind's unquestioning adherence to a utilitarian framework means their analysis misses the extent to which the increased availability of automated systems has the potential to reinforce, rather than alleviate, another way in which the professions may be said to be showing signs of failing:

> 2: The professions do not live up to the ideal of ethical integrity that plays a key role in their self-conception and justification of relative self-regulation

In large part because their outcome-focused analysis leads them to deem the professions' ideal of ethical integrity to be a contingent (rather than conceptual) feature, Susskind and Susskind brush off rather lightly the possibility that computer systems might worsen (rather than alleviate) the second way (encapsulated in '2') in which the professions are failing.[19] While they do refer to ethics in the charge-list mentioned above, their way of formulating that concern does not depart from their overall utilitarian outlook and merely prolongs the affordability aspect via a concern for distributive justice: 'if we have the technological means to spread expertise in society far more widely at much lower cost, we believe we should strive to make this happen'.[20] Indeed, who wouldn't?

Yet within the professions it is not just 'expertise' that automated systems will spread. It is also the ethical challenges that stem from the vulnerability inherent in the lay-professional relationship. Today those challenges are as pressing as ever: one might argue that, as Western societies' concern for moral equality has grown, so has the saliency of the professions' particular ethical responsibility. Sadly though, there is little evidence that this increased saliency has in fact led to growing ethical awareness within the professions. Hence the normative conclusions that stem from this 'alternative' genealogical critique would, overall, be remarkably similar to the Susskinds', with an important proviso: the success criterion for emerging uses of computer systems in the professions should not 'just' be whether they improve the affordability, quality and accountability of the professions' services. On those three counts, a lot of automated systems are likely to be successful.

Interestingly, the Susskinds acknowledge this outcome-independent line of argument by referring to Sandel's 'moral limits' objection:[21] could it be that we feel uncomfortable about the idea of an increasing number of professional 'tasks' being handled by computer

[18] ibid.

[19] Unlike affordability or quality concerns, which lend themselves to an outcome driven approach, the professions' (relative) failure to live up to their ideal of ethical integrity is notably difficult to pin down. Recent empirical studies (mostly in the fields of law and medicine, less so in education) paint a rather worrisome picture when it comes to assessing the extent to which 'the professions' live up to various interpretations of the ideal of ethical integrity that plays such a role in both their self-understanding and the 'grand bargain' at the root of their relative monopoly and self-regulation privileges. BG Garth, 'Rethinking the Legal Profession's Approach to Collective Self-Improvement: Competence and the Consumer Perspective' (1983) *Wisconsin Law Review* 639; HP Gunz and SP Gunz, 'The Lawyer's Response to Organizational Professional Conflict: An Empirical Study of the Ethical Decision Making of In-house Counsel' (2002) 39 *American Business Law Journal* 241.

[20] Susskind and Susskind (n 5).

[21] MJ Sandel, *What Money Can't Buy: The Moral Limits of Markets* (Macmillan 2012); Susskind and Susskind (n 5).

systems for reasons that are similar in kind to those that underlie our repugnance at body organs being traded like ordinary goods? Sandel seeks to capture what underlies our concern about the proliferation of market norms (which, in the context of the present discussion, would displace 'professional norms') by referring, among other things, to two key objections. Sandel's 'inequality objection' – '[i]n short, if inequality is large enough, markets may lead to a lack of adequate or "meaningful consent" in the choices people make'[22] – is quickly dealt with, the Susskinds pointing out that automated systems will improve access to affordable expertise and will only affect the provision of expertise (not payment for it).

Most interesting is the Susskind's answer to Sandel's 'corruption objection', which they formulate as a 'trade-off' –here their answer is worth quoting in full:

> Let us turn now to the Corruption Objection. There are two basic reasons why we might also resist this—either because we do not think that the professions in fact have a special moral character, or because we do not think that this character is degraded in the market. But suppose instead that both are true—that the professions do have this character and that it is degraded in some way if their work is done according to market norms. In that case, there is a trade-off—we must strike a balance between the value we place on protecting this moral character and the value we place on the pursuit of greater access to affordable practice expertise. The Corruption Objection is clear on how to resolve this trade-off—the pursuit of the latter comes at the price of the former, but that price is too high and ought to be resisted. In contrast, we believe, for two reasons, that a diminution in the moral character of professional work is a price worth paying. First, the professions, unlike many other occupations, are responsible for many of the most important functions and services in society. It was recognition of the importance of their work that drove the initial 'grand bargain' (see section 1.4). Secondly, levels of access and affordability to the practical expertise that the professions provide fall well short of acceptable.[23]

One may want to pause and disentangle the two different ways in which the notion of 'price' intervenes in the passage above. First it surfaces implicitly in the Susskinds' reference to Sandel's argument – i.e. there are things whose nature is perverted (and hence their value to us is undermined) by any endeavour to place a price on them. The Susskinds 'find Sandel's arguments to be compelling in general'.[24] Yet they resist the application of such arguments to the displacement of professional norms by market norms because 'a diminution in the moral character of professional work is a price worth paying' (see above). Here the word 'price' conveys the fact that, because the two values at stake cannot be reconciled, one of them must give way. The problem is that while the nature of one of the values at stake is pretty clear – increasing accessibility to the professions (which itself must stem from a concern for equality of opportunity) – the other is not. The Susskinds only refer to 'a diminution in the moral character of professional work', without much indication of what might ground that moral character.[25] Given their wide-ranging, minimalist understanding of the professions as 'our answer to the limited knowledge problem', it is far from clear what, in their account, warrants granting that moral character to the professions.

[22]Susskind and Susskind (n 5).
[23]ibid.
[24]ibid.
[25]In fact, the Susskinds frequently remind us that 'it is important not to exaggerate this dimension of professional activity': 'Moreover, "moral" tasks may well feature more prominently in professional work than they do in other sectors. Again, though, it is important not to exaggerate this dimension of professional activity. It would be disingenuous to suggest that all professional work involves matters of the gravest ethical significance'; ibid.

That moral character is fleshed out, by contrast, in the vulnerability-based account of the professions hinted at above,[26] and explicitly tied to a key value: moral equality. Given the very particular type of vulnerability at stake, the very shape (and depth) of our commitment to moral equality is determined in part by the way our professions meet, on a daily basis, the demands entailed by this vulnerability (it might be that, as a matter of fact, our commitment to moral equality is all too often left in rather bad shape by our professions –see note 8). Now let's imagine – for the sake of the argument – that the Susskinds are happy to endorse the above. If it's our commitment to moral equality that explains the professions' moral character, then such is also the value that, in their 'trade-off', gives way in favour of the Susskind's concern for a different kind of equality: equality of access to professional expertise. If so, one may start to worry about the extent to which such a trade-off makes sense. In his 'The idea of equality',[27] Williams eloquently depicts the way in which 'equality of respect' (i.e. moral equality) and equality of opportunity will often end up 'pulling in different directions', urging us to nevertheless resist the 'temptation to abandon some of its elements'. For it is tempting

> to claim, for instance, that equality of opportunity is the only ideal that is at all practicable, and equality of respect a vague and perhaps nostalgic illusion; or alternatively, that equality of respect is genuine equality, and equality of opportunity an inegalitarian betrayal of the ideal – all the more so if it is thoroughly pursued, as now it is not.[28]

The good news is that there is no need, in this particular instance, to 'abandon' anything, and the Susskinds' trade-off between the professions' moral character and their accessibility is only live because of a false premise: that the professions' 'moral character' will be degraded by the introduction of ever more capable computer systems. That premise is false on one condition: that these systems be designed and introduced in a way that frees the professions to take the full measure of the responsibility that is theirs in virtue of the very particular type of vulnerability they are confronted with. This is discussed in Section 2.2, while Section 2.1 critically examines the normative assumptions commonly held by wholesale automation enthusiasts.

2. Possible (ab)uses of computer systems within the professions

2.1. When data trumps rules (and principles?): automation's potential scope

As a methodology that is well suited to automating those tasks that rely heavily on 'tacit knowledge' Machine Learning (ML) is destined to have a large impact on professional, value-loaded contexts. Only 15 years ago, the possibility of replacing professionals in such tasks as medical diagnosis was deemed implausible, given its 'non-routine' character. The latter characterisation – routine v. non-routine tasks[29] – was used[30] to distinguish those tasks 'that can be accomplished by following explicit rules' (and hence lend

[26]This account is fully articulated in Delacroix (n 3).
[27]B Williams, 'The Idea of Equality' in B Williams (ed), *Problems of the Self: Philosophical Papers 1956–1972* (CUP 1973).
[28]ibid.
[29]According to Polanyi, non-routine tasks rely on tacit knowledge about which 'we can know more than we can tell'; M Polanyi, 'The Logic of Tacit Inference' (1966) 41 *Philosophy* 4.
[30]D Autor, FS Levy and RJ Murnane, 'Upstairs, Downstairs: Computers and Skills on Two Floors of a Large Bank' (2002) 55(3) *Industrial and Labor Relations Review* 432.

themselves to automation) versus those 'for which the rules are not sufficiently well understood to be specified in computer code and executed'.[31] Today, this rules-based demarcation (traditional, rule-based expert systems need explicit rules) is both obsolete and misleading. Because ML algorithms operate on the basis of a fundamentally different methodology from that underlying expert systems, the extent to which a task can be distilled into rules (whether explicit or tacit[32]) has become irrelevant. What matters, instead, is the accessibility and quality of the recorded data pertaining to that task: the more abundant the data, the more robust the correlations, which in turn determine the performance of the ML algorithm.

As an example, a recent medical application allowing for wholly automated skin cancer diagnosis[33] will have been trained on a large dataset structured as example pairs (x, y) where 'x' corresponds to the images containing skin lesions – the pixels – and 'y' identifies whether it is cancerous or not – the disease label. The aim of the system's learning process is to find a function $f:X \rightarrow Y$ that matches the example pairs. Such a function need not – and in fact *does not* – reflect the rules – tacit or otherwise – which dermatologists follow when assessing skin lesions.[34]

Of course, the pertinent datasets are not always neatly labelled example pairs that are relatively free of any syntactic ambiguity, as in the example above. The set of data X can be generated by the artificial agent's interaction with the environment. In that case the aim of the learning process is to come up with an action-selection policy that minimises some measure of long-term cost. Systems combining supervised learning from human expert games (where x is the game strategy, and y the game result) with reinforcement learning from games of self-play have recently made headlines given their ability to outperform human experts.[35] Outside the world of games, a system designed to predict the outcome of cases tried by the European Court of Human Rights has also received a lot of attention, given its impressive accuracy (79 percent).[36] That system was based on a binary classification task – is a specific article of the Convention violated or not? – similar to that of the skin cancer application, except that instead of having images as

[31]F Levy and RJ Murnane, 'The Skill Content of Recent Technological Change: An Empirical Exploration' (2003) 118 *The Quarterly Journal of Economics* 1279.

[32]In contrast, D Autor, *Polanyi's Paradox and the Shape of Employment Growth*, Vol 20485 (National Bureau of Economic Research 2014) insists on holding on to this concept of 'non-routine task', by – misleadingly– explaining recent successes in applications relying on tacit knowledge thus: '[R]ather than teach machines rules that we do not understand, engineers develop machines that attempt to infer tacit rules from context, abundant data, and applied statistics.' DH Autor, 'Why Are There Still So Many Jobs? The History and Future of Workplace Automation' (2015) 29 *The Journal of Economic Perspectives* 23.

[33]A Esteva and others, 'Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks' (2017) 542 *Nature* 115 describes a system that can accurately predict whether a picture of a skin lesion is cancerous or not. The performance of that system was found to be 'on a par' with that of a group of 21 board-certified dermatologists.

[34]It is important to understand that this independence from the professional's thought processes applies to all of machine learning's recent forays into the professions. This methodological point has significant implications. It has for instance long been assumed that, if successful, the automation of key aspects of doctors' work would spell the end of the Courts' deference towards the expertise underlying clinical judgment, as it would prove that the relevant knowledge can in fact be applied 'mechanically' and is hence routinized'. Since the success of Machine Learning, this assumption has been proven wrong. At best, all a judge would find, were she to open the 'black box' at the heart of Machine Learning applications, would be a large collection of seemingly random correlations (since learning algorithms proceed independently of any modelisation of the processes underlying the task at stake).

[35]D Silver and others, 'Mastering the Game of Go with Deep Neural Networks and Tree Search' [Nature Publishing Group] (2016) 529 *Nature* 484.

[36]N Aletras and others, 'Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective' (2016) 2 *PeerJ Computer Science* e93.

input, this system relied on recent progress in natural language processing to classify textual input (extracted from published ECHR cases).

These advances have enabled the birth of a so-called 'science of judicial predictions'. Unlike 'amateur' prediction models, 'which are typically assessed ex post to infer causes', the algorithmic model developed by Katz et al.[37] to predict the decisions of the US Supreme Court over nearly two centuries (despite changes in the Court composition and socio-cultural contexts) for instance managed to anticipate whether the court would 'reverse' the status quo or not with 70.2% accuracy. A similar endeavour, which focussed on the French Court de Cassation rulings (with more diverse outcome variables[38]) managed to predict the court ruling based on the case description with impressive accuracy.[39]

While these court cases predictions do bring benefits (particularly for those businesses whose risk models in part depend on the outcome of such cases), they also come with dangers. The most evident risk is that of inherent conservatism: cases with a low success prediction are unlikely to be heard in court, in turn making organic changes within case law less likely. The latter changes indeed often depend upon an accumulation of previous, unsuccessful cases that trigger a growing number of dissenting voices (both within and without the judiciary). While there may be ways of developing tools that not only predict the chances of success in court, but also the likelihood that a particular case will eventually contribute to some organic evolution within case law, there will be little commercial incentives for the latter tools. The other type of risk concomitant with such prediction tools is less tangible, but could nevertheless contribute to a shift in the aspirations we associate with law: those who deem prediction accuracy to be the most promising aspect of recent technological advances within the legal profession indeed often assume that the success of a legal system can and ought to be measured according to the extent to which such a system reduces uncertainty. From that perspective, if those advances ultimately allow us to automate (rather than merely predict) court rulings, we should embrace them: how better to foster 'the rule of law [which is] preferable to that of any individual'[40] than by substituting algorithmic predictability for fickle human judgments? As Pasquale puts it: 'One literal way of achieving the oft-quoted ideal "a rule of law, not of men" is to dispense altogether with persons implementing or interpreting law'.[41]

[37]DM Katz, MJ Bommarito II and J Blackman, 'A General Approach for Predicting the Behavior of the Supreme Court of the United States' (2017) 12(4) *PLOS ONE* e0174698 <https://doi.org/10.1371/journal.pone.0174698>. Remarkably, the same authors went on to test the accuracy of crowdsourcing 'as an alternative to expert-based judgment or purely data-driven approaches' to predicting future Supreme Court decisions, and reached an impressive 80.8% accuracy M Katz, MJ Bommarito and J Blackman, 'Crowdsourcing Accurately and Robustly Predicts Supreme Court Decisions' (2017) <https://ssrn.com/abstract=3085710>.

[38]Cassation, Cassation sans renvoi, Cassation partielle, Cassation partielle sans renvoi, cassation partielle cassation, cassation partielle rejet cassation, rejet, irrecevabilité (non-lieu à statuer, non-lieu à recevoir, qpc seule irrecevabilité, were excluded because of their rarity).

[39]The accuracy score varies depending on the number of outcome variables that are selected (8 or 6): 'We observe an apparent 6 percentage points decrease in average scores when the classifier is trained on the dataset with more classes'. OM Sulea and others, 'Predicting the Law Area and Decisions of French Supreme Court Cases' (2017) <arXiv:1708.01681>.

[40]Aristotle, *The Politics* (Cambridge University Press 1996) Book III, Ch 16, p 88.

[41]FA Pasquale, 'A Rule of Persons, Not Machines: The Limits of Legal Automation' (2018) *George Washington Law Review* forthcoming refers to 'automators of law [who] tend to see their work as one more step toward elevating the legal system above the fallibility of any particular person within it' and cites JC Smith, 'Machine Intelligence and Legal Reasoning – The Charles Green Lecture in Law and Technology' (1998) 73 *Chicago-Kent Law Review* 277 in that context.

In an endeavour to highlight the perils inherent in the literal (and reductive) understanding of the rule of law presupposed by wholesale automation enthusiasts, Pasquale articulates 'what is lost when society cedes more aspects of the authoritative articulation of rights and duties to computational processes'.[42] He does so through a detailed survey of both modest (from automated tax preparation to contesting parking tickets) and less modest 'substitution through legal automation' endeavours. While the former, low-stakes substitutive legal automation is, on balance,[43] deemed a 'laudable phenomenon',[44] Pasquale emphasises the far-reaching long-term costs inherent in proposals to accelerate what he calls 'the robotization of law'. In this respect, he joins a growing number of voices – Mireille Hildebrandt[45] notable among them – who warn us of the way in which:

> we can no longer take the Rule of Law for granted as an affordance of our information and communication technology (ICT) infrastructure, due to the rapid and radical integration of algorithmic decision-systems and other types of data-driven intelligence into the administration of justice […] If the technological embodiment of modern law and its offspring, the Rule of Law, is changed, the law itself will change – potentially beyond recognition.[46]

This essay is an invitation to step back: just as Section 1 considered an alternative interpretation of the *raison d'être* underlying the professions (one that is not easily compatible with the Susskinds' consequentialist mantra when it comes to assessing the legitimacy of wholesale automation), the next Section 2.2 stems from a shift in focus. If, instead of maximising legal certainty, one aims to empower legal professionals to live up to their ethical responsibility, the difficult questions when it comes to designing computer systems for the legal profession are less about the degree of autonomy they should be endowed with and more about how to achieve true human-computer complementarity.

2.2. Achieving human-computer complementarity through decision-support systems

There will be cases (think parking fines) where there is little downside to the vital increase in affordability and accessibility that automation brings, provided transparency, accountability and privacy are preserved – as far as possible.[47] Yet aside from administrative legal work (case management is one example), such clear-cut cases of unproblematic, wholesale automation are not that common: laudable as it may be, the drive to democratise legal

[42]Pasquale (n 41).

[43]Pasquale highlights the often-underestimated extent to which automation (whether in the tax domain or otherwise) often ends up licensing higher levels of legal complexity – and reduced transparency.

[44]Pasquale (n 41).

[45]M Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Elgar 2015).

[46]M Hildebrandt, 'The Force of Law and the Force of Technology' in M McGuire and T Holt (eds), *The Routledge Handbook for Technology, Crime and Justice* (Routledge 2017).

[47]The extent to which full transparency is in fact both achievable and desirable varies according to particular applications. A Weller, 'Challenges for Transparency' (2017 ICML Workshop on Human Interpretability in Machine Learning (WHI 2017)) outlines several reasons to doubt the desirability of transparency, particularly when the latter allows users to 'game' the system. Accountability concerns (and the need for 'explainable AI') have given rise to a fast-growing research field that is beyond the scope of this review. See among others B Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 *Big Data and Society* 1; M Veale, M Van Kleek and R Binns, 'Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making' (2018) <arXiv preprint arXiv:180201029>; S Wachter, B Mittelstadt and L Floridi, 'Transparent, Explainable and Accountable AI for Robotics' (2017) 2 *Science Robotics* 1.

expertise by distilling it into mass-market, problem solver apps can conceal issues that demand human input. As an example, an app that allows those who have recently been dismissed from their job to avail themselves of their right to severance pay (which may be opaque due to complex legislation) is commendable. Yet without a proactive referral system (to employment lawyers, but also potentially other types of social or psychological support) such an app would, in many ways, be deemed to fail its users: the vulnerability that is concomitant with finding oneself jobless indeed cannot be addressed by algorithms, no matter how much empathy such apps may be able to display.

At the 'lay' end, apps of the kind mentioned above may be deemed invaluable, empowering tools[48] for those who would otherwise be left unaware of and unable to exercise their rights. Yet wherever our commitment to moral equality is at stake (given the special kind of vulnerability described above), such apps ought to be conceived as gateways or 'triage devices' directing to appropriate human advice, rather than replacing it altogether.

At the legal professional's end, automated systems could, in principle, be designed so as to allow legal professionals to develop greater emotional and situational awareness and see beyond 'the usual man in the usual place'[49] when meeting, interviewing or defending a client. Ethical lapses within professional practice indeed most often stem from a failure to discern ethically relevant considerations which may only be distantly connected to the problem in relation to which a professional is consulted. Whether it comes to the need to take into account the vulnerability of a client's family member,[50] say, or considering the impact of a company's merger upon the environment and members of the local community, an ability to see beyond one's immediate query does condition the ethical awareness which a professional needs if she is to live up to her particular responsibility.

Professions-specific automated systems can and should be designed with a view to fostering such perspective widening. To that end, they may usefully leverage recent research on the factors that impact upon individuals' differential creativity.[51] Among the characteristics used to assess such creativity,[52] fluency and flexibility are of particular relevance

[48] In the field of consumer protection, apps that allow for the automatic detection of 'unfair clauses [that] are currently hidden within long and hardly readable ToS'; M Lippi and others, 'Automated Detection of Unfair Clauses in Online Consumer Contracts' (2017) 302 *Legal Knowledge and Information Systems Frontierst in Artificial Intelligence and Applications* 145 may be valuable as a stop-gap measure, but in the longer term those apps may end up standing in the way of a much-needed, fundamental reform in the way informed consent is obtained.

[49]

> [T]he horrible thing about all legal officials, even the best, about all judges, magistrates, barristers, detectives, and policemen, is not that they are wicked (some of them are good), not that they are stupid (several of them are quite intelligent). It is simply that they have got used to it. Strictly they do not see the prisoner in the dock; all they see is the usual man in the usual place. They do not see the awful court of judgment; they only see their own workshop. (GK Chesterton, 'The Twelve Men' in GK Chesteron (ed), *Tremendous Trifles* (Sheed & Ward 1955))

[50] For a study examining the impact of expertise and cognitive load upon a GP's ability to pick up signs of child-safeguarding concerns, see X Pan and others, 'A Study of Professional Awareness Using Immersive Virtual Reality: The Responses of General Practitioners to Child Safeguarding Concerns' (2018) *Frontiers in Robotics and AI* <https://doi.org/10.3389/frobt.2018.00080>.

[51] DL Zabelina and others, 'Patterning and Nonpatterning in Creative Cognition: Insights from Performance in a Random Number Generation Task' (2012) 6(2) *Psychology of Aesthetics, Creativity, and the Arts* 137.

[52] The recently published report sponsored by the French Parliament 'For a meaningful artificial intelligence'; C Villani, *For a Meaningful Artificial Intelligence: Towards a French and European Strategy* (Villani Mission on Artificial Intelligence 2018) highlights at several points the need to look ' into the complementarity between humans and artificial intelligence: if we are to assume that, for most jobs, individuals will have to work with a machine, then it is vital to find a complementarity set-up that does not alienate staff but instead allows for the development of truly human capabilities, such as creativity, manual dexterity, problem- solving abilities, etc.'

when it comes to countering the effects of professional routine. Yet so far the fast-developing research on artificial creativity has not ventured as much as it could into potential professions-specific, ethics-oriented applications.[53] If we can have systems that foster the creativity of mathematicians[54] and musicians, why not lawyers? Some would retort that lawyers are not supposed to think too creatively: their professional practice is meant to be structured around well-defined procedures, rules and principles, and decision-support systems are there, if anything, to help them abide by those procedures while reducing cognitive load. The latter understanding of professional responsibility is, sadly, as common as it is flawed. Far from a 'moral sums game' at which one may excel, professional ethics can only be *practiced*: the dynamic and fallible nature of the values at stake requires constant, renewed engagement. Since the latter can all too easily get compromised under the combined weight of time pressure and management constraints, professions-specific decision-support systems ought to draw upon the growing number of 'creativity focused' applications in domains ranging from art to business.[55] They also need to take into account our weaknesses and biases. While the impact of so-called 'automation bias' is considered in (i), our normative laziness and the 'loafing effect' is addressed in (ii).

2.2.1. Instrumental rationality and automation bias

When discussing what they call the 'Trust Objection' (in relation to the deployment of automated systems within the professions), the Susskinds argue:

> Our primary need is only for a reliable outcome. Of course, we do not want the people and systems that meet this need to be dishonest or criminal. But neither do we necessarily need them to be motivated by an altruistic regard for others. That would be too onerous a requirement. Our primary concern need not be with altruism or the achievement of the highest ethical ideals but to make sure that our problems are resolved reliably, efficiently, and effectively.[56]

The above quote has the merit of being candid. The instrumental rationality[57] that is openly at work here often underlies the uncritical endorsement of various forms of efficiency maximising technologies. Of course, there is nothing wrong with an endeavour to maximise efficiency per se. What needs to be considered, rather, is the extent to which

[53]For a notable study in that direction, see J Inthorn, ME Tabacchi and R Seising, 'Having the Final Say: Machine Support of Ethical Decisions of Doctors' in Simon Peter van Rysewyk and Matthijs Pontier (eds), *Machine Medical Ethics* (Springer 2015):

> The DSS can ask the right questions, can suggest different ethical perspectives […] and it can certainly inspire creativity. Creativity can be simulated in the system by stretching the given parameter ranges, using the perspectives of other actors, or even putting the problem description in another context. This can help the user to find solutions that are not limited by a restricted frame of mind that focuses on the situation at hand but frequently misses ideas on how to extend or modify decision spaces by integrating multiple perspectives and normative questions into decision making processes.

[54]'Heuristics for transforming conceptual spaces, including the space of heuristics, have been applied in a number of programs (leant 1983). One of these, whose task is to generate new mathematical concepts [is called 'the automatic mathematician']. [It] might be developed for other domains, in which the knowledge and judgment of human users could aid, and be aided by, the application of the transformational and evaluative heuristics'. M Boden, *Creativity and Art: Three Roads to Surprise* (OUP 2010).

[55]F Adam and others, *Creativity and Innovation in Decision Making and Decision Support* (Ludic Publishing LTD 2006).

[56]Susskind and Susskind (n 5).

[57]When it informs the assessment of actions, 'instrumental rationality' assesses actions solely by reference to how effective they are in achieving their specified end (hence without the need to judge the legitimacy of that end).

the rapid growth in the deployment of professions-specific systems is at all likely to amplify the dangers inherent in a technology-enabled 'cloak of instrumental rationality':[58]

> The enabling technology can mesmerise the actors, shielding or displacing the moral issues present. It appears that technology is the updraft that allows and facilitates a dramatic spread of an ideology legitimised by the unquestioned reign of instrumental rationality.[59]

This unquestioning attitude towards technology has notably been associated with what social psychology studies call 'automation bias'. These studies suggest that 'automated devices can fundamentally change how people approach their work, which in turn can lead to new and different kinds of error'.[60] Because errors that stem from having allowed incorrect automated input to override a correct, 'human' – i.e. non-automated – judgment (those errors are classified as 'automation bias') are both difficult to track down and only anecdotally reported, studies of automation bias have so far mostly[61] proceeded on the basis of randomised controlled trials,[62] such as Skitka et al.'s study.[63] The latter compared error rates in a simulated flight task with and without a computer that monitored system states and made decision recommendations. When the automated aid was inaccurate (missing a key event for instance), participants in the non-automated condition outperformed those in the automated condition.

Of particular interest are the causal factors that Skitka et al. hypothesised might contribute to the commission and omission errors associated with the presence of automated decision aids. Among these, Skitka et al. identify cognitive miserliness[64] – 'most people will take the road of least cognitive effort, and rather than systematically analyse each decision, will use decision rules of thumb or heuristics' (automated systems will act as the latter).[65] They also refer to what they call 'social loafing, diffusion of responsibility[66] and possible belief in the relative authority of computers and automated decision aids':

> Finally, people may respond to computers and automated decision aids as decision- making authorities. Obedience can be defined as people's willingness to conform to the demands of

[58] 'Morrow's assessment of the potential for disconnect between technology and morality, that '[t]he story of evil in the world is so often a matter of hardware outperforming conscience: Can outruns Should. Or rather, Can outruns Should Not': L Morrow, *Evil: An Investigation* (Basic Books 2003) 56, corroborates this concern' GS Reed and N Jones, 'Toward Modeling and Automating Ethical Decision Making: Design, Implementation, Limitations, and Responsibilities' (2013) 32 *Topoi* 237.

[59] JF Dillard, 'Professional Services, IBM, and the Holocaust' (2003) 17 *Journal of Information Systems* 14.

[60] LJ Skitka, K Mosier and MD Burdick, 'Accountability and Automation Bias' (2000) 52 *International Journal of Human-Computer Studies* 701.

[61] With a few exceptions, see notably EM Campbell and others, 'Overdependence on Technology: An Unintended Adverse Consequence of Computerized Provider Order Entry' (AMIA Annual Symposium Proceedings, Vol. 2007, 94, American Medical Informatics Association 2007) for a study based on fieldwork.

[62] These randomized controlled trials may not be ideally suited to understanding the impact of automated decision aids in real-life circumstances.

[63] LJ Skitka, KL Mosier and M Burdick, 'Does Automation Bias Decision-Making?' (1999) 51 *International Journal of Human-Computer Studies* 991.

[64] The term 'cognitive miser' comes from J Crocker, ST Fiske and SE Taylor, 'Schematic Bases of Belief Change' in J Richard Eiser (eds), *Attitudinal Judgment* (Springer 1984).

[65] Skitka, Mosier and Burdick (n 63) 992.

[66]

> Given that people treat computers who share task responsibilities as a 'team member', and show many of the same in-group favoritism effects for computers that they show with people (Nass, Fogg and Moon 1996), it may not be surprising to find that diffusion of responsibility and social loafing effects also emerge in human-computer interaction. To the extent that some tasks are shared with computerized or automated decision aids people may well diffuse responsibility for those tasks to those aids, and feel less compelled to put forth a strong individual effort. (ibid 992)

> an authority, even if those demands violate people's sense of what is right […] Given that computers and automated decision aids are introduced into many work environments with the articulated goal of reducing human error, they may well be interpreted to be smarter and more authoritative than their users. To the extent that people view computers and automated decision aids as authorities, they may be more likely to blindly follow their recommendations, even in the face of information that indicates they would be wiser not to.[67]

The latter two factors (diffusion of responsibility and deference to authority) are of particular importance for our present concerns. For the decision aid systems that may plausibly be used in the legal profession differ in some important ways from those used for plane navigation. When a decision needs to be made based on the latter, both the parameters that ought to inform the decision and the options underlying it are well defined. For a wide range of legal matters, by contrast, the parameters that contribute to both the framing and the solution of a problem are the product of a value-laden interpretation. The responsibility (and apparent precariousness) entailed by this inevitable axiological component can be hard to bear. In that context, any opportunity to 'pass the moral buck' is particularly attractive, especially when the 'buck' is passed to a system that does not deal in ambiguities and raw intuitions, thus conveniently ironing out dimly perceived inconsistencies or unarticulated ethical concerns.

2.2.2. A special kind of moral philosopher? Beware what you wish for.

According to the Susskinds, it is possible that:

> future systems (modelled, for example, on traditional, rule-based expert systems) could articulate and balance moral arguments, identify consistencies and illogicalities, point out assumptions and presuppositions of given lines of debate, and identify conclusions that can validly be drawn from some set of premises. Such systems would be a special kind of moral philosopher, capable of clear and structured reasoning about ethical issues.[68]

The emphasis on 'clear and structured reasoning', pointing at a procedural rather than substantive understanding of ethical expertise has the advantage of avoiding the naïve (and all too common) assumption that currently dominates discussions of what computer scientists call 'the value-alignment problem'. Its discussion indeed often proceeds from the assumption that moral values are essentially static: once the values that are relevant to a particular application have been identified (a challenge in itself), one may proceed with their neat incorporation into a system that is designed to simplify our practical reasoning.

Aside from its naivety (given ongoing, constantly evolving ethical disagreements), I have highlighted elsewhere[69] the danger inherent in such an assumption, which may well turn into a self-fulfilling prophecy: what if we do indeed end up with a set of static moral values? Rather than reflecting some fanciful state of collective 'ideological and ethical contentment', such a standstill would be brought about because of a novel type of collective disability, triggered by lack of normative exercise.

[67] ibid.
[68] Susskind and Susskind (n 5).
[69] S Delacroix, *Taking Turing by Surprise? Designing Autonomous Systems for Morally-Loaded Contexts* (2019) <https://papers.ssrn.com/abstract=3025626>.

Whenever decision-support systems succeed in enabling us to step back and relax – somehow trusting machines to have gotten our 'moral sums' right –, they cannot but compromise the critical engagement necessary to live up to one's ethical responsibility. Now, the answer is not to ditch any form decision-support, but rather to design the latter differently. Of particular interest, in terms of method, are systems that keep end-users within the learning loop. Sometimes referred to as 'interactive machine learning' or 'IML',[70] this method demands regular input on the part of end-users (as well as their monitoring the result of the learning process), in turn requiring the designers of such systems to pay a lot more attention to the extent to which particular design choices are likely to foster (rather than diminish) the extent to which end-users retain a critical, reflective stance during their interaction with computer systems.[71] Aside from potentially improving the system's learning performance, this interactive method also has the potential to keep moral torpor at bay by encouraging an 'ethical feedback loop' that carves a continuous, active role on the part of the professional community whom the system is designed for.

3. Conclusion

Tomorrow's 'professional workshop' is more likely than not to rely heavily on automated systems. There may come a time when these automated systems' superior reliability seemingly extends to *most* of the work associated with a particular profession. Given the likely affordability gains concomitant with widespread automation, it will be tempting to consider the latter's desirability (and legitimacy) as self-evident. This essay hopes to have contributed robust grounds to question this consequentialist logic, whose influence is felt well beyond the Susskinds' book (and academia).

Section 1 challenges the minimalist understanding of the professions that conditions much of the Susskinds' normative conclusions. If the specific responsibility of legal professionals stems from more than 'ideological rhetoric', and is concomitant with the fact that it is our commitment to moral equality that is at stake every time lawyers (fail to) hail the specific vulnerability inherent in their professional relationship, the case for wholesale automation is turned on its head. One can no longer assume that, as a rule, wholesale automation is both legitimate and desirable, provided it improves the quality and accessibility of legal services.[72] The assumption, instead, is firmly in

70

> Although humans are an integral part of the learning process (the provide labels, rankings etc.), traditional machine learning systems used in these applications are agnostic to the fact that inputs/outputs are from/for humans. In contrast, interactive machine learning places end-users in the learning loop (end users is an integral part of the learning process), observing the result of learning and providing input meant to improve the learning outcome. Canonical applications of IML include scenarios involving humans interacting with robots to teach them to perform certain tasks, humans helping virtual agents play computer games by giving them feedback on their performance. (Wendell Wallach and C Allen, *Moral Machines: Teaching Robots Right from Wrong* (OUP 2008))

[71] A number of human-computer interaction studies have recently started to focus on the value inherent in fostering a reflective attitude on the part of technology users. Along this line, see ED Mekler and K Hornbaek, 'Momentary Pleasure or Lasting Meaning?: Distinguishing Eudaimonic and Hedonic User Experiences' (Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (2016)) as well as P Slovák, C Frauenberger and G Fitzpatrick, 'Reflective Practicum: A Framework of Sensitising Concepts to Design for Transformative Reflection' (Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (2017)).

[72] And preserves a minimum degree of transparency, accountability and privacy – the fast-growing research concerning the feasibility – and desirability- of the latter constraints in the context of various kinds of automation is beyond the scope of this review (see note 47).

favour of designing systems that better enable legal professionals to live up to their specific responsibility.

The second section outlines key challenges in the design of such profession-specific, 'ethics aware' decision-support systems. Aside from reducing professionals' cognitive load, decision-support systems can and should be designed to counter the effects of routinisation, raise awareness of seemingly peripheral considerations and, most importantly, better listen to and engage with the person seeking professional expertise. Our growing understanding of the non-cognitive underpinnings of professional judgment (in part thanks to virtual reality simulations[73]) – combined with novel, creativity-focussed AI research – has the potential to radically alter the way we design decision-support systems meant for the morally-loaded contexts that pervade most of the legal profession. This potential will only be realised, however, if the legal profession as a whole proactively engages in a long overdue debate about the values it stands for (as a profession) and the extent to which current system design choices may hamper or foster those values.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This work was supported by Leverhulme Trust: [Grant Number 07134DT PLP2010/0096].

ORCID

Sylvie Delacroix http://orcid.org/0000-0002-8517-7782

[73]Reliance on immersive virtual reality to study 'live' and ecologically valid professional judgments – albeit in a controlled environment – has great potential as a tool to study the 'skilled intuitions' that are known to play a major role in professional judgment. See Pan and others (n 50).