Ethics, computer systems and the professions

Sylvie Delacroix
UCL Laws and Computer Science[1]

Table of Contents

[1] The work leading to this paper was funded by the Leverhulme Trust. I am grateful for the comments and insights of Maria Lee, Kathy Charmaz, Richard Moorhead, Jonathan Montgomery, Andrea Sangiovanni, Michella Antonelli, Jurgen Van Gael, Neil Lawrence and the participants of the recent IALS workshop on expertise.

Tomorrow's "professional workshop" is more likely than not to rely heavily on automated systems, whether they are conceived as decision-aids or whether they are meant to replace professionals in some or most of the tasks they accomplish. Can such systems be designed in such a way as to improve the situational awareness that conditions a professional's meeting her ethical responsibility (and hence counterbalance the combined effects of rigidified cognitive processes and repeated exposure to particular situations[2])? This paper's endeavour to answer that question leads to two distinct theses.

First, this paper demarcates the professions from other types of expert service providers. In doing so, it makes a key claim: of those expert services whose safe delivery is in the public interest, healthcare, legal / financial services and education stand out because they give rise to a very particular type of vulnerability, one that potentially threatens the moral equality of those seeking those services (the particular case of those shaping the architecture conditioning our virtual interactions will also be discussed). The first section of this paper delineates the ethical demands stemming from this vulnerability, which is contrasted to the kind of vulnerability that characterises all lay-expert relationships given their inherent knowledge asymmetry. This paper argues that the specific type of responsibility entailed by the former can and should ground an understanding of the "professions" that has been (re)defined around it (and is hence restricted in scope).

This paper's second thesis claims that the success criterion for emerging uses of artificial intelligence in the professions should not *just* be whether they improve the affordability, quality and accountability of the professions' services (as the Susskinds' recent *The future of the professions*[3] suggests). On those three counts, a lot of computer systems[4] are likely to be successful. Yet it will be nothing short of a catastrophe if such systems fail to assist professionals (whom they will be working alongside with) in meeting their particular ethical responsibility. Aside from reducing a professional's cognitive load, computer systems can be designed in such a way as to challenge routine perceptions and modes of thought, hence improving situational (and ethical) awareness. Yet the instrumental rationality that overwhelmingly presides over computer scientists' lightning speed progress (a rationality that is openly at work in the Susskinds' arguments) makes such design choices unlikely. This paper aims to foster much-needed public engagement with the design and shape of automated systems within the professions by introducing key technical concepts and outlining both the challenges and potential uses of such systems. The latter requires a robust understanding of the particular nature of the professions' ethical responsibility, which is unpacked in the section that follows.

[2] These effects are discussed in detail in a separate paper considering the impact of habit, habituation and intuitive processes on the way we make ethical judgments in a professional context.

[3] (R. Susskind & Susskind, 2015)

[4] The phrase "computer system" rather than "expert system" is meant to reflect the fact that an increasing number of automated decision processes designed to support (or take over aspects of) professional work have very little in common with the "expert systems" that were designed in the 1960's.

I. Delineating the professions: institutional history, scope and conceptual analysis

In its adjective form, the term "professional" is easy enough to define as the opposite of amateurism. Things get more complicated when it comes to grasping what is entailed by the plural noun "the professions". This is in part because of its loose colloquial use today, to cover any occupation that has been professionalised by the introduction of a code of ethics, formal training and accreditation. Reference to these formal characteristics as a lowest common denominator defining the professions tends to go hand in hand with a functional analysis. The latter was key to the birth of the sociology of the professions as a field. One of the foundational studies by Carr-Saunders and Wilson highlighted the professions' role in "preserv[ing] and pass[ing] on a tradition" to resist the "crude forces which threaten steady and peaceful evolution".[5] This functional analysis gave rise to a dominant tradition within the sociology of the professions whose aim is to try and capture the professions' key, differentiating characteristics[6] (these studies typically emphasise their formal self-regulation and training programme, a commitment to a rather vague notion of "public service", the nature of their knowledge base etc.).

Critiques[7] of such trait-based and functionalist accounts of the professions have sought to dispute the assumption that professionalism is best understood as an inherent characteristic of particular occupations. This paper proceeds from a similarly critical stand towards this assumption, yet instead of shifting attention towards the wider socio-economic context[8] of professionalism claims, this analysis turns the functionalist assumption upside down. The characteristics by reference to which the professions are best delineated are not those of the occupations they encompass, but rather those of the relationship between the professions and individuals who need their services.

This perspective reversal throws a different light on professionalism's changing institutional context. The shift towards increasingly large and complex organizations[9] encompassing (and

[5] (Saunders & Wilson, 1933, p. 497) In a similar vein, Durkheim's analysis of the professions (Durkheim, 1957) is part and parcel of his analysis of the disenchantment characteristic of Modernity and its concomitant "normative void" (which the professions are meant to address).

[6] (Hickson & Thomas, 1969) (Greenwood, 1957)

[7] (Johnson, 1972) for instance denounces professionalism as a means of reinforcing and controlling the power concomitant with being the exclusive provider of a particular kind of service. In a related critique, Freidson highlights the stakes inherent in the professions' retaining control over "the social and economic methods of organising the performance of their own work" (Freidson, 1970, p. 185)

[8] "But professionalization was at best a misleading concept, for it involved more the forms than the contents of professional life. It ignored who was doing what to whom and how, concentrating instead on association, licensure, ethics code. In fact, not only did it miss the contents of professional activity, but also the larger situation in which that activity occurs." (Abbott, 2014, pp. 1-2)

[9] Skeptics liken this shift to the "industrialization" of legal practice, which is seen as inevitably leading to professionalism retreating to a "shrinking territory", to use John Flood's expression (Flood, 2008).

thus structuring) professional work[10] has often been seen as unavoidably endangering the normative commitments at the heart of the professions:

> *"The traditional assumption (Aronowitz, 1973; Burris, 1993; Leicht and Fennell, 2001; Oppenheimer, 1973) has been that [the shift of professional activity within the confines of increasingly large and complex organizations] would inevitably erode professionalism, as the new organizational context of work would at minimum expose professionals to managerial pressures and at worst recreate 'factory like conditions' (Oppenheimer, 1973, pp. 213–14), triggering processes of deprofessionalization and proletarianization."*[11]

In contrast, a user-centred conceptualisation of the values inherent in the professions considers such institutional transformations (of which the technological innovations unpacked in the rest of this paper are but one aspect) to be a source of both challenges and opportunities.[12] In striving to unpack the normative commitments that stem from the characteristics of lay-professionals relationships (rather than from some "transcendent" values[13]), the account of the professions defended here also resists increasingly minimalist definitions. Influenced by so-called "conflict or power framework" critiques[14], this minimalism can notably be seen at play in the Susskind's *The future of the professions*, according to which "the professions are our current solution to the challenge in society of supporting people who need access to practical expertise".[15]

The following section resolutely goes against this minimalist trend, and argues instead for a robust restriction in the scope of the professions. This restriction is driven by the need to answer conceptually (and normatively) the fact that, unlike other occupations that are in the public interest, the provision of legal and financial advice, healthcare and education is concomitant with a very particular type of vulnerability (which in turn gives rise to a specific type of responsibility).

---

10 The "dissection and stratification" of legal work that is concomitant with the widespread resort to corporate management techniques is deemed by (Sommerlad, 1995) to "deal a substantial blow to the professional's image as an independent moral agent in the public sphere".

11 (Muzio, Brock, & Suddaby, 2013, p. 702)

12 The opportunities brought about by the shift of professional work to organizational settings is notably highlighted by (Evetts, 2011, pp. 416-417): "Other aspects of organizational change, including credentialism, governance and external forms of regulation, would seem to produce some benefits (for example of transparency and control of more extreme professional powers) while, at the same time, resulting in detrimental effects such as increased bureaucracy, form-filling and paperwork […]There are other opportunities arising from the combination of the logics of professionalism and the organization which may prove advantageous. One of these is the incorporation of human resource management (HRM) from the organization into professional employment practices, processes and procedures."

13 Freidson for instance links "professionalism's claim to special status" to a "claim to allegiance to some transcendent value, whether that be Truth, Beauty, Enlightenment, Justice, Salvation, Health, or Prosperity" (Freidson, 1999, p. 127)

14 See note 5.

15 (R. Susskind & Susskind, 2015, p. 250)

1.1. Conceptualising the professions by reference to their specific vulnerability-based responsibility

1.1.1. A Commitment to moral equality

Endeavouring to climb some unknown mountain without resorting to the expertise of a mountain guide is reckless. Embarking upon a mountain expedition with a supposed mountain guide who turns out to be a fraud is perilous too. Much the same goes for repairing (recent) cars, building houses or scuba-diving. The knowledge asymmetry that prompts the very need for expertise necessarily leaves those resorting to it in a position of vulnerability, which is exacerbated by the difficulty inherent in establishing an expert's credentials. Given the public interest in a wide range of expert services being delivered safely and reliably, a series of safeguards are typically put in place. They govern, among other things, experts' accreditation and the remedies available should experts' services prove wanting.

Of those expert services whose safe delivery is in the public interest, healthcare[16], legal[17] and financial services and education stand out because they are concomitant with a very particular type of vulnerability. The difference between the latter and the vulnerability at play when relying on a mountain guide (or car mechanic) is not one of degree: when our life is at stake on the side of the mountain we are probably as vulnerable as can be. The difference lies in the fact that the role of the guide does not affect our development of those interests and concerns that are closest to our sense of self. With educators, by contrast, the materials with which we develop a sense of self are collected and given shape. With doctors, bankers and lawyers, shame, intimacy, fragility are to the fore: whether we are struggling to preserve our health or our social standing and recognition (which a divorce, sudden poverty, prosecution etc. can all endanger), our sense of *owning*[18] the way we

[16] Healthcare is meant to include all expert services aimed at supporting or improving our health (hence counsellors, psychologists, midwives, nurses, osteopaths etc. are all included). As a short for "healthcare provider", I sometimes use the term "doctors" to refer in fact to all those involved in healthcare.

[17] For our purposes, legal services should be understood loosely. Because of their power to radically affect and shape family life and fundamental liberties, there are grounds to include both social care workers and police officers (arguments against including them would point at the fact that in both cases those particular powers are only ever temporary, for they need to be enacted by lawyers).

[18] Such sense of ownership (or authorship) need not entail a smooth and coherent sense of self. As Sangiovanni puts it: "All it requires is that our ambivalences, regrets, dependencies, and upheavals are integrated into our self-conception—but there is no embargo on their being integrated *as* ambivalences, regrets, dependencies, and upheavals. Our sense of self could, furthermore, be of an episodic sort that rejects structured narratives; it could even be grounded in a repudiation of a unified self in favour of a loose succession of selves (and self-conceptions). The important thing is that such a self-conception (or series of discordant or disunified self-conceptions), whatever it is (they are), be genuinely felt as the product of a self-conceiver, as someone who is a (part) author of their life" (Sangiovanni, 2017)

project ourselves, both socially and through our body is typically fragilised. Doctors or lawyers taking charge in the face of events such as a grave illness or unemployment (or an educator guiding a child through some milestones) may all too easily reinforce the feeling that we are only ever "worked upon"[19], determined by events whose roots and implications we cannot comprehend (and hence attempt to make *ours*).

In the medical domain, Parsons' analysis of illness (as a "disturbance of the total person") brings out some aspects of the particular vulnerability inherent in the doctor-patient relationship. Parsons indeed emphasises the extent to which illness affects a person's ability to define her self-conception through interaction or "role performances", which are henceforth compromised (and can thus add up to a perceived loss of "esteem"[20]):

> *"[T]he situation of illness very generally presents the patient and those close to him with complex problems of emotional adjustment. It is, that is to say, a situation of strain […] suffering, helplessness, disablement and the risk of death, or sometimes its certainty, constitute fundamental disturbances of the expectations by which men live. They cannot in general be emotionally `accepted' without the accompaniments of strain with which we are familiar and hence without difficult adjustments […] [F]or the `normal' person illness, the more so the greater its severity, constitutes a frustration of expectancies of his normal life pattern. He is cut off from his normal spheres of activity, and many of his normal enjoyments. He is often humiliated by his incapacity to function normally. His social relationships are disrupted to a greater or less degree."*[21]

Parsons' reference to "humiliation"[22] (concomitant with "an incapacity to function normally") hints at the extent to which the epistemic inequality that presides over all forms of expert services can all too easily morph, within the provision of professional services, into moral inequality. How so? Sangiovanni brilliantly articulates the conceptual link between the exploitation of another's vulnerability and moral (in)equality via the notion of social cruelty. What makes it wrong to treat others as inferiors -and hence what grounds our

---

[19] "Having a firm sense of self gives rise to a concomitant sense of ourselves as autonomous, or self-governing: our choices, actions, values, commitments, concerns are our *own*. Losing our sense of self gives rise, in turn, to the sense that we are not in control, that we are being determined by events, that we are not ourselves. In one mode, we create and are created; in the other, we are only ever worked upon" (Sangiovanni, 2017).

[20] In the context of chronic illness, (Charmaz, 1983) highlights the fact that the persistent compromising of one's ability to accrue esteem is concomitant with an eroded sense of social personhood, or "loss of self".

[21] (Parsons, 2012, pp. 310-311)

[22] Parsons also emphasises that a patient often has no choice but to accept violations of personal and bodily integrity: "[I]t should be noted that the burdens the physician asks his patients and their families to assume on this advice are often very severe. They include suffering […] risk of death, permanent or lengthy disablement, severe financial costs and various others. In terms of common sense it can always be said that the patient has the obvious interest in getting well and hence should be ready to accept any measures which may prove necessary.". (Parsons, 2012, p. 310) The reciprocity argument hinted at in the above quote is expanded upon in (Gerhardt, 1987): from a Parsonian perspective, the patient's refraining from reciprocating those personal and bodily violations is concomitant with an exemption from those sanctions typically imposed on those who fail in (or do not conform to) their social roles.

commitment to moral equality- is not some mysterious "value-bestowing capacity possessed to an equal extent by each one of us" (such as dignity[23]) but rather a rejection of social cruelty. The latter is defined as "the unauthorized, harmful and wrongful use of another's vulnerability to attack or obliterate their capacity to develop and maintain a sense of self".[24] On this basis, one may argue that the professions have a particular type of ethical responsibility because educators, bankers, lawyers and doctors[25] are all -in our society- in a position to significantly alter our sense of self (and the knowledge asymmetry considerably diminishes our ability to partake in this alteration[26]). Such an account contrasts with traditional attempts to ground professional responsibility in some "transcendent" value, whether it be a rather vague notion of "public good"[27] or Freidson's reference to "Truth, Beauty, Enlightenment, Justice, Salvation, Health, or Prosperity".[28]

1.1.2. The Professions' shifting domain

1.1.2.1. Socio-cultural variation

This vulnerability-based account of the professions necessarily entails that the scope of the latter will vary historically and across societies (and will be the object of ongoing contestation): in societies where the development of one's sense of self is widely acknowledged to be dependent on a spiritual dimension, spiritual advisers and / or religious representatives will uncontroversially belong to the professions. In a society that is committed to a separation between the religious and public spheres (as in France today), spiritual guidance would not be deemed to be in the public interest (even if a large part of the population comes to build a strong link between their sense of self and a religious affiliation). Similarly, in a less materialistic society (one that does not foster a strong connection between one's sense of self and wealth), bankers and financial advisers may not be included within the scope of the professions.

---

[23] Interestingly, a reference to dignity was also prevalent in early studies of the professions, such as that of Saunders and Wilson, but it applied the other way round: Carr-Saunders and Wilson indeed believed that the professions seen as a liberating force from slavish dependence upon the state. As a growing number gained admission to the professions, a correspondingly larger number would share "the institutional base" from which to enjoy "freedom, dignity and responsibility" (Saunders & Wilson, 1933, p. 503)

[24] (Sangiovanni, 2017)

[25] Here "lawyers" and "doctors" should be understood to include the wide range of professionals whom we may have to entrust with the care of our health or legal (and social) status.

[26] "This need for recognition is borne of our nature as self-presenting beings: we do not simply participate in the (dual) process of reflection and creation as Robinson Crusoes. We define and redefine our self-conception—which includes, recall, our place in a network of roles and relationships—in communication and interaction with others similarly engaged selves." (Sangiovanni, 2017)

[27] "A manifest and demonstrable commitment to the public good is what legitimately creates the profession as a political force: professionalism depends upon serving the public interest and also upon proving to itself and to others that it does genuinely serve the public interest. From this comes the idea that only a community of experts can be relied upon to manage the power associated with that knowledge in the public interest. This community of experts is necessary to protect the public interest from the forces of market, state, and self-interest. It is a plausible claim for law" (Moorhead, 2014b, pp. 450-451)

[28] Freidson, 1999, p. 127

1.1.2.2. Shrinking domain

Unlike traditional references to "[a] manifest and demonstrable commitment to the public good"[29] or some transcendent values, this vulnerability-based account of the professions also proves helpful in considering the way in which some occupations may, in the future, cease to be professions. What if, say, medicine was so advanced that it only ever required at most one "day off" in order to successfully cure or address any ailments? In such a desirable scenario, one may argue that healthcare providers would be no different from any other experts, since being ill wouldn't prompt the very specific type of vulnerability discussed above.

Following the same line, there are fields within current medical and legal practices that are not concomitant with the type of vulnerability described above (whether one considers examples such as mergers and acquisitions or haematology). Specialists in such fields may also be deemed to be no different from other experts, with two practical implications. Aside from the fact that their responsibility is of a different nature from their -medical or legal- colleagues, one may also argue that, were future computer systems able to *replace* them in all of their tasks, this would not raise the same issues as those arising from the eventual replacement of *professionals* by computer systems, which are discussed below.

1.1.2.3. Technology-enabled expansion

The advent of specialised computer systems that are capable of replacing both experts and professionals within some fields, thus opening up the enticing prospect of wide, low cost availability to the lay public, raises several important questions. One of them applies equally to both expert and professionals, and can be phrased thus: how does one address the problem inherent in the fact that the successful use of such systems requires its users to have correctly labelled their problem as one that can be addressed by that particular system? Can such systems reliably acknowledge their own limitations and alert users to the need for human advice when appropriate?

> *"Do we embrace IT and the information society and develop legal information products and systems which will guide us far more extensively than would otherwise be possible but may fail, on occasions, both to notify us that more complex issues may be at play and that human specialist advice is needed? Or do we reject the new technology on the grounds that even although the law will be invoked far less*

[29] "A manifest and demonstrable commitment to the public good is what legitimately creates the profession as a political force: professionalism depends upon serving the public interest and also upon proving to itself and to others that it does genuinely serve the public interest. From this comes the idea that only a community of experts can be relied upon to manage the power associated with that knowledge in the public interest. This community of experts is necessary to protect the public interest from the forces of market, state, and self-interest. It is a plausible claim for law" (Moorhead, 2014b, pp. 450-451)

frequently, on those fewer occasions we can nevertheless then have complete confidence in the reliable, expert disposal of our problems?"[30]

This is an important question, and early on Susskind unambiguously advocated the first of the two options presented above: the benefits of increased accessibility and affordability far outweigh the possible wrong-headed uses of such systems from time to time. I would argue for a third option: as the real world-relevance of this "automation adequacy" problem grows, so will the need for a new "triage" type of profession, one that would not be conceivable without the systematic support of some ambitious, cross-disciplinary expert system. Many of those who most need healthcare, legal or educational help happen to be both the most vulnerable and worst-equipped to identify that need in the first place. Merging the expertise of GPs, CAB advisers, social care workers and educators need not be an inconceivable dream *if* -at the moment it is still a big "if"- a cross-disciplinary team were to harness the resources of machine learning and systematically record the raw, life narratives presented in the context of GP, CAB etc consultations (as well as the categories applied to such narratives). The challenges raised by the need to secure both the data recorded and its appropriate processing would be considerable, but no less considerable than in other sensitive healthcare applications. Such a system could help the professional "listener" to engage with each person's narrative without prejudice and to raise questions that she may not otherwise raise. Consideration of the most appropriate type of specialist help would come as a second step, facilitated (but not dictated) by the support of this cross-disciplinary "expert" system.[31]

Now, to increase the accessibility and affordability of such a sorely needed "triage" profession, it may be tempting to develop an automated, "chat box" version[32] of such a "cross-disciplinary listener". From a purely utilitarian perspective, it would be hard to find fault with such automation[33], considering the acute need for such "triage" services to be highly accessible. Given the no less acute vulnerability that will inevitably characterise the relationship between such a professional "listener" and those in need of its services, however, I would argue that automation, however enticing, is particularly perilous in this case. Given the rapid progress of affective computing, it is only a matter of time before we are able to converse with "embodied conversational agents" who are able to pick up on subtle, non-verbal behaviour cues to learn about our emotional states and themselves use

[30] (R. E. Susskind, 1998, p. 282)

[31] Such a system would capitalise on the recent advances in both speech-recognition and natural language processing.

[32] When immersive virtual reality becomes affordable enough, a 3D, avatar version would be particularly attractive from that perspective, given its ability to mobilise powerful emotions and instinctive reactions not otherwise mobilised in 2D (or audio) interactions.

[33] While (R. Susskind & Susskind, 2015) clearly adopts such a utilitarian perspective, (R. E. Susskind, 1998) remains agnostic. In a two-pages section entitled "Moral Limitations", it mentions Weisembaum's concern for "obscene" computer applications, "including those that `propose to substitute a computer system for a human function that involves interpersonal respect, understanding, and love' (page 269)" but stops short of taking any particular ethical stand: "It is my purpose here to encourage that ethical debate [about extending information technology into the administration of justice] but not to engage in it directly" (R. E. Susskind, 1998, p. 69)

both verbal and non-verbal cues to *signify* empathy far better[34] than most professionals would. The catch is in the "signify": there is a world of difference between signifying and experiencing empathy.[35] The latter is arguably at the root of our experiencing ethics as an imperative that trumps instrumental considerations. For a robot (or virtual avatar), by contrast, that ethical imperative can never be more than a goal among others[36] - no matter how high it features in the hierarchy of that machine's goals.

The above hint at virtual reality raises a different way in which the domain of the professions may be expanding: what about those in charge of designing the architecture that shapes our virtual interactions, whether they be "merely" through social networks or through fully immersive virtual environments? Recent events have thrown a rather sharp light upon the impact of the former, prompting calls for the professionalization of machine-learning[37] as an occupation. Given the extent to which social networks have for many become an essential forum of social interaction (and hence a medium through which they (re)define their self-conception), there is little doubt that those re-writing the algorithms that determine whether -and which- friends or news are "suggested" (for instance) have a particular type of ethical responsibility. The latter takes its root in a vulnerability similar in kind to what has so far been discussed in the context of healthcare, education and legal / financial services. The difference with the latter is twofold: first, there is an extent to which we simply haven't quite caught up with the fact that this relatively novel occupation provides services that are no less in the public interest than those of lawyers, educators or doctors. Second, and most importantly, there is an increasingly wrongheaded tendency to think of professional services as involving -to a minimal degree[38]- some form of "personal" relationship, which contrasts with the remote connection between those who develop such algorithms and those whose relationships they shape. In short, I would argue it is high time machine learning as an occupation be deemed part of the professions, and that the only

[34] See notably (R. Susskind & Susskind, 2015, p. 170)

[35] Levinas' ethics is built upon empathetic imagination (see also Husserl's *Einfühlung* or Smith's "sympathy"). Kearney defines empathetic imagination as the ability to be receptive to the other: "can we be responsible for the other if we are not first receptive to the other?...if we can't hear its call, if we cannot empathize?" (Kearney, 1998), p. 232.

[36] Because our evaluative stances are very much built upon emotions such as joy, fear or repulsion, which in turn take their root in bodily sensations (and our memory thereof - See Damasio's Somatic Marker theory), the evaluative stances that a sophisticated automated system may, at some point, possibly come to develop autonomously will necessarily be alien to us, embodied humans (just as nursery stories will likely remain alien to such systems).

[37] Neil Lawrence for instance points out at the "widespread social effects" of machine-learning models to suggest that the latter ought to be "validated" according to standards set by machine learning as a *professional* community: "Statistics is already familiar with the vital role that a profession plays in regulating claims. In the pharmaceutical industry a statistical trial is the standard by which a new drug is validated. Traditionally, in machine learning, we have paid less attention to the validation and explanation of our models because our models *seemed* less consequential. A prediction of a model to re-rank a search result or our social media news feed. However, these methods, while each making a decision of perhaps little consequence, are now being applied to millions or billions of people. The sum of many inconsequential decisions could certainly end up being highly consequential. Whatever the effect of filter bubbles and fake news on recent elections, the very fact that such algorithms can have such widespread societal effects is of significant importance." (Lawrence, 2016)

[38] Even when education, healthcare or legal services are provided "remotely".

reason it hasn't been so is the worrying lack of public awareness when it comes to the impact of the architecture shaping our virtual interactions (and our vulnerability therein).

1.1.3. Responsibility and Trustworthiness

For now -and within the UK- lawyers, bankers, doctors, educators and those shaping our virtual interactions can all be said to be endowed with an additional level of responsibility (over and above that of all experts in virtue of their epistemic superiority) that stems from the particular kind of vulnerability they are confronted with. 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".[39] Yet Susskind and Susskind frame the above concern as a "psychological" one and do not delve into the ethical implications flowing from it. When "disempowerment" means that a patient, the client of a lawyer or a pupil is prevented from being able to play an active role in the deployment of her sense of self, a fundamental value -moral equality- is under threat.

If such a key value is indeed at stake in the way the professions operate, one cannot help but be concerned by the way the special degree of responsibility it entails (and the non-utilitarian framework it demands) is bulldozed out of the range of relevant considerations by Susskind and Susskind's take on one of the key concerns raised against the prospect of widespread reliance on increasingly capable machines in the professions ("both when it displaces human beings and when it enables less expert human beings to perform at the level of specialists"[40]) . To those who worry about "the loss of trustworthy institutions", meant to "protect ourselves from exploitation by unscrupulous quacks"[41], 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*

[39] (R. Susskind & Susskind, 2015, p. 35)

[40] (R. Susskind & Susskind, 2015, p. 232)

[41] Ibidem

<u>all have problems for which we do not personally have the expertise to resolve)</u>. Yet we think this is mistaken. Our primary need is only for a reliable outcome."[42]

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"[43]- that is our answer to what Susskind and Susskind call our "fundamental challenge" (i.e. that none of us has the knowledge necessary to being 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 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.[44] Yet one gets the sense that, for the Susskinds, the claim to ethical integrity that some 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[45], that claim can be shown to have increasingly little in the way of empirical evidence to back it up.[46]

1.2. Spoiling the story: why the need for a concept? Marrying concept and history? The art of genealogy.

The above paragraph opens up a methodological question that has been lurking behind much of the discussion so far: As a historically rooted and organically developed institution, to what extent is it helpful to insist on a conceptual analysis of the professions? Does it make sense to restrict the scope of the professions to healthcare, legal/ financial services and education only[47] (against the contemporary expansionist trend discussed in the

---

[42] (R. Susskind & Susskind, 2015, pp. 236-237)

[43] It should be clear by now that the professions delineate a particular, sub-group of "experts".

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

[45] "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" (R. Susskind & Susskind, 2015, p. 237)

[46] (Keogh, 2013; Lagu, Goff, Hannon, Shatz, & Lindenauer, 2013; Lombarts, Plochg, Thompson, Arah, & on behalf of the, 2014; Moorhead, 2010, 2014a; Moorhead, Sherr, & Paterson, 2003; Moorhead et al., 2001; O'Fallon & Butterfield, 2005; Paterson & Sherr, 1992; Sherr, Moorhead, & Paterson, 1994)

[47] The disparities between "the ethical and legal ideas" that are promoted today by the legal, healthcare and education components of the professions should give one pause, whether one opts for a genealogical frame of inquiry or not. How can it make sense to bundle those three fields of expertise (healthcare, law and education)

introduction) in virtue of a shared feature deemed to be salient "from the outside"? Why pick the special kind of vulnerability entailed by the provision of those particular services (a feature that is not widely discussed by any of those professions' justificatory discourses)? And most importantly, why pick any feature at all as a conceptually salient trait determining the legitimate, practical reach of that concept? To conceptually restrict the scope of the professions by stipulating that only those confronted with the special type of vulnerability described above "qualify" may sound like a desperate attempt to foist a political and / or ideological agenda upon an institution that has (and will continue to) develop(ed) organically.

One could, by contrast, humbly acknowledge the professions as a constantly evolving institution and, as a social scientist, limit oneself to recording those transformations and possibly predicting future ones (which might radically transform that institution)? 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 the technology-induced transformations they foresee 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 gain from being seen as the result of a -piecemeal- genealogical account[48] of the professions. A genealogy necessarily marries historical investigation and functional analysis. The former highlights the historical processes that brought about a particular institution. The moving forces underlying these processes may have nothing to do with the official rhetoric legitimising that institution. In that case a genealogical account will prove damaging to that institution's perceived legitimacy, but this disobliging consequence is only the by-product of a genealogy's true critical ambition. The latter is to debunk ahistorical accounts of institutions[49], which contribute abstract hypotheses that consciously or unconsciously

---

despite the significant differences in -among other things- the extent to which they do in fact self-regulate? Can one nevertheless claim that they should be considered apart from other types of expertise (such as banking, accountancy or scientific research) and together delineate the sphere of the "professions"? The answer, in short, consists in arguing that the disparate legal regimes (and legitimising discourses behind them) are a contingent by-product of historical power struggles that are not only obsolete but likely to work at cross-purposes with tomorrow's emerging needs (think of the strategic importance of education in a world marked by the rise of increasingly far-reaching automated systems). Today the differences between each of these professions are not only hard to justify, they also actively undermine any chance to effectively address the challenge that unites them: given the particular vulnerability of those relying on each of those professions' expert services, what safeguards can be put in place to make sure that members of the professions live up to the special responsibility entailed by it?

[48] For full developments on the concept of genealogy as a method of inquiry whose tools include history (among others), see …

[49] While some institutions remain stable under genealogical explanation, others face a necessary 'makeover', requiring re-examination of their meaning or legitimation basis. The full power of a genealogy typically reveals

share in those institutions' self-understanding. This debunking ambition requires a functional account. In asking 'why do we have this or that institution?', a genealogy indeed presupposes that the object it studies can meaningfully be treated as *functional*, that is, as serving an end other than itself. The study of the driving forces underlying the history of a given institution aims at unveiling the unexpected or hidden ambitions and needs which that institution serves.

Susskind and Susskind refer, among other things, to Terence Johnson's historically informed critique of professionalism[50] to debunk the still widely influential, traditional account of the professions as "devoted to the service of the public, above and beyond material incentives".[51] 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 Susskind and Susskind "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.[52] The Susskinds' normative conclusion is to embrace the radical transformation promised by increasingly widespread reliance on computer systems within the professions. This technology-driven transformation indeed has the potential, they argue, of remedying each of the three "failure counts" mentioned above.

Now, if instead of starting from the "devotion to public service" functional hypothesis that informs Susskind and Susskind's critique (and 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 article, 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 slightly 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 the occupations outlined above (education, legal and financial services, and healthcare), there are specific ethical challenges that are qualitatively different from those entailed by the responsibility which is concomitant with the epistemic superiority that characterises all experts.

---

itself when it is directed at institutions whose self-understanding requires concealing the impact of these historical processes. The stronger their claim to authority, the more reluctant these institutions are to reveal their historical contingency, "both in the sense that they are what they are rather than some others, and also in the sense that the historical changes that brought them about are not obviously related in a grounding or epistemically favourable way to the ethical [or legal etc.] ideas they encouraged" (Williams, 2000, p. 155)

[50] (Johnson, 1972)

[51] (Larson & Larson, 1979)

[52] This failure verdict will be discussed in more detail in section 2.1

Today those challenges are as pressing as ever. In fact 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 computer systems are likely to be successful. Yet it will be nothing short of a catastrophe if computer systems fail to assist professionals (whom they will be working alongside with[53]) in preserving the situational awareness necessary to countering the effects of habituation and being equal to their particular ethical responsibility. Or to put it simply: how to not "get [too] used to it", and avoid the plague aptly described by Chesterton: "[s]trictly 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".[54] Different ways of designing computer systems to meet this challenge will be discussed in section 2.3.

II. Computer systems fit for the professions: specific needs and constraints

2.1. Transforming the professions: failures, needs and opportunities

There is no lack of empirical evidence to support one of the key verdicts underlying Susskind and Susskind's *The Future of the Professions*: our professions are "failing", for they are "by and large, […] unaffordable, under-exploiting technology, disempowering, ethically challengeable, underperforming, and inscrutable".[55] This is a hefty and seemingly comprehensive charge-list.[56] 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:

---

[53] The reasons why, in healthcare, law and education I do not ever foresee a total replacement of professionals by computer systems will be expanded upon in section 2.4.

[54] (Chesterton, 1955)

[55] (R. Susskind & Susskind, 2015). In the legal domain, see e.g. (Harper, 2013)

[56] The other three items on their "charge list" -the professions are "under-exploiting technology, disempowering and ethically challengeable (all interpreted in a strictly utilitarian fashion)-, will be discussed in the next section.

1: the professions do not serve the public interest they claim to serve, given their failure to deliver services[57] that are affordable[58], of good quality[59], and accountable[60].

Now there is little doubt that the growing availability and sophistication of computer systems is going to 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".[61] If, that is, the latter is understood along strictly utilitarian lines, à la Susskind and Susskind. There is clear potential for automated systems to dramatically improve both the affordability and quality of the services delivered by the professions.[62] Yet Susskind and Susskind's unquestioning adherence to a utilitarian framework means their analysis misses the extent to which the increased availability of such 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 utilitarian, 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

---

[57] 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.

[58] "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" (R. Susskind & Susskind, 2015, p. 33)

[59] "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." (R. Susskind & Susskind, 2015, pp. 35-36) In the legal domain, there is a growing body of empirical literature, reviewed in detail -mostly in the British context- in (Moorhead, 2014a), that paints an even bleaker picture than that suggested by Susskind and Susskind above. (Keogh, 2013; Lagu et al., 2013; Lombarts et al., 2014)

[60] "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" (R. Susskind & Susskind, 2015, p. 36)

[61] (R. Susskind & Susskind, 2015, p. 9)

[62] They may also improve their accountability, if certain safeguards are built into the design of such systems (see below)

professions are "failing".[63] 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 reiterates 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".[64]

The Susskinds' way of answering what they call the "trust objection" (which comes closest to encapsulating the concern for ethical integrity which they otherwise bypass) is most revealing:

> *"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."*[65]

The above quote has the merit of being candid. The instrumental rationality[66] that is openly at work here, and that often underlies the uncritical endorsement of various forms of efficiency-maximising technologies is not always easy to pin down. The devastating conclusions that can stem from the routine but systematic application of such instrumental rationality by dutiful professionals is notably illustrated in Edwin Black's[67] chilling account of IBM's role in the Holocaust. A precursor to the computer, the "IBM Hollerith" punch card machine (used to tabulate and alphabetize census data) indeed greatly facilitated the identification, segregation and extermination of millions of Jews.[68] In a similar vein, Bauman describes the extent to which "the spirit of instrumental rationality, and its modern, bureaucratic form of institutionalization […] made the Holocaust-style solutions not only possible, but eminently 'reasonable' - and increased the probability of their choice".[69]

---

[63] 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. (Garth, 1983; Gunz & Gunz, 2002)

[64] (R. Susskind & Susskind, 2015, p. 36).

[65] (R. Susskind & Susskind, 2015, pp. 236-237)

[66] 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).

[67] (Black & Wallace, 2001)

[68] In France (where the Resistance tampered with Hollerith machines) the rate of Jewish deaths is claimed to be one-third that of Holland, where this technology was well applied.

[69] This increase in probability is more than fortuitously related to the ability of modern bureaucracy to co-ordinate the action of a great number of moral individuals in the pursuit of any, also immoral, ends" (Bauman, 1989, p. 18).

Now one does not need to consider examples as extreme as the Holocaust to witness the effects of instrumental rationality let loose. The crucial question, for the purpose of this paper, is the extent to which the introduction of automated systems is at all likely to amplify the dangers inherent in a technology-enabled "cloak of instrumental rationality":

> *"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"*[70]

What needs to be considered, in other words, is the extent to which the distance between the morally abhorrent and the ordinary is likely to be narrowed by increased reliance on ever more capable machines. Social psychology studies on the effect of so called "automation bias" suggest that "automated devices can fundamentally change how people approach their work, which in turn can lead to new and different kinds of error".[71] 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[72] proceeded on the basis of randomized controlled trials[73], such as Skitka et al.'s study.[74] 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[75] – "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).[76] They also refer to what they call "social loafing, diffusion of responsibility[77] and possible belief in the relative authority of computers and automated decision aids":

---

[70] (Dillard, 2003, p. 14)

[71] (Linda J. Skitka, Mosier, & Burdick, 2000)

[72] With a few exceptions, see notably (Campbell, Sittig, Guappone, Dykstra, & Ash, 2007) for a study based on fieldwork.

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

[74] (Linda J Skitka, Mosier, & Burdick, 1999)

[75] The term "cognitive miser" comes from (Crocker, Fiske, & Taylor, 1984).

[76] (Linda J Skitka et al., 1999, p. 992)

[77] "Given that people treat computers who share task responsibilities as a `team member', and show many of the same in-group favouritism effects for computers that they show with people (Nass, Fogg & Moon, 1996), it may not be surprising to find that diffusion of responsibility and social loafing effects also emerge in human-

> *"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 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".*[78]

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 professions (as defined in this paper) 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, healthcare and education 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.

Now there are ways in which particular kinds of computer systems could be designed to address not only the "automation bias" discussed above but also, more generally, the specific challenges arising from the professions' vulnerability-based ethical responsibility. Given the explicitly instrumental logic that underlies the Susskinds' account of the professions, however, such constructive developments are made less likely. The following section's endeavour to delineate possible professions-specific, computer system applications that would take into account the challenges mentioned above is certainly not meant to be exhaustive. Its chief ambition is to stimulate the much needed public debate on this issue in a way that is alive to the "potential ethical implications"[79] which Lord Neuberger (referring to the Susskinds' discussion[80]) emphasised recently as "one of the

---

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." (Linda J Skitka et al., 1999, p. 992)

[78] (Linda J Skitka et al., 1999, p. 993)

[79] (Neuberger, 2016)

[80] "The Susskinds point out that this potential development has ethical as well as employment implications and they call for a public debate on the issue. There are many who are sceptical about the Susskinds' predictions, but there is no doubt but that they could be right. The legal profession should, I suggest, be preparing for the problems and opportunities which would arise from such an enormous potential area of development, and one of the most difficult challenges will be to consider the potential ethical implications" (Neuberger, 2016)

most difficult challenges" arising from the likely deployment of more and more capable computer systems within the professions.

2.2. Possible professions-specific computer systems applications: challenges and constraints

In this paper I use the term "expert system" to refer to any form of artificial intelligence designed to handle, support or replace (aspects of) expert work. There are very different kinds of tools that can be used to develop such expert systems. The outline below is not exhaustive, and is meant as a guide through concepts whose technicality tends to discourage much-needed public engagement.

2.2.1. Rule Based Systems:

The advent of "expert systems", first developed in the 1970's, was based on a key insight: "intelligent systems derive their power from the knowledge they possess rather than from the specific formalisms and inference schemes they use."[81] Such systems require input from experts in their respective fields (whether they be oncologists, insolvency lawyers or otherwise), working alongside so-called knowledge engineers. The latter can either construct a decision tree manually (based on "if-then rules") or induce a decision tree, based on a collection of labelled instances.[82] The performance of such an expert system will in large part depend on the adequate selection of particular classifying features (or "attributes")- and their discriminative power: in a large range of medical domains, say, blood pressure and body temperature tend to have a high discriminative power. Those classifying features will in turn form the "nodes" in the decision tree.

[81] Formulated by Feigenbaum (Feigenbaum, 1977) and paraphrased by (Hayes-Roth, Waterman, & Lenat, 1983) , pp 6-7.

[82] "Hunt et al. (1966) used their Concept Learning System (CLS) for building decision trees in medical diagnosis and prognosis. They state (p. 170): `In medicine fairly large files of records may be obtained in the course of routine hospital administration or from a special survey. Such records are often examined in order to plan an intensive, and perhaps expensive, specialized investigation. A drawback to this research strategy is that it is difficult to organize large files of records to reveal complex interactions in a manner that can be understood by the human investigator. Some help can be obtained by using computer oriented techniques of in- formation retrieval, such as program to print selected two- and three-way tables plotting one variable against another. The investigator still must nominate the variable in which he is interested, since such programs have no way of discovering interesting patterns on their own. A CLS program, on the other hand, is designed to do precisely this'." (Kononenko, 2001, p. 4)

One of the key advantages of such rule-based systems is their relative degree of transparency. This means that in principle domain experts (whether they be doctors, lawyers or educators) should be able to challenge and / or modify a decision tree to account for advances in their respective fields. In practice, however, such decision trees tend to be embedded in larger, complex systems, and the decision trees themselves often require heuristic add-ons -relying on a variety of statistical methods- to optimise their performance.[83] While their relative accountability makes such systems particularly attractive, they can be hazardous when applied to data that is sparse or non-deterministic (i.e. the data's particular meaning or implications are themselves subject to interpretation). This limitation, combined with the considerable investment required from experts whose availability is limited by definition (expert systems tend to be particularly attractive within highly specialised, complex fields), has contributed to a shift, since the 1990s, towards different kinds of tools.

2.2.2. Case-Based Reasoning Systems

When the knowledge acquisition task that conditions rule-based systems is either impossible or cumbersome (for instance in less-well understood domains that do not lend themselves to knowledge formalisation), case-based reasoning systems can be developed, even on the basis of a limited amounted of experience (as new cases can simply be added to the case-base). Such case-based methodology is appropriate in domains that are likely to see novel or exceptional cases (without the latter a domain is better modelled with a rule-based system) and where there is some logical relationship between old and new cases.[84] Aside from Medicine and Engineering, Law is a domain that lends itself to this approach.[85]

These systems can be combined with artificial neural networks (see below). In such "hybrid" systems[86], the knowledge extracted by such neural networks notably guides the cases retrieval.[87] Recent advances in Natural Language Processing[88] and Machine learning for instance make it possible to automate, in some cases, the analysis of legal materials.[89]

---

[83] Such add-ons may for instance be introduced to take into account the dependency between different attributes (or classifying features).

[84] (Kolodner, 1992; Pal, Dillon, & Yeung, 2012)

[85] (Ashley, 1992)

[86] (Maxwell & Schafer, 2010) propose a "hybrid" method "achieved via split query expansion with two branches. One branch applies knowledge-engineered or other currently applied expansion techniques, and the other applies linguistically based expansion on query predicates". That second "branch" "utilises information from events, states and attributions, extracted automatically from predicate-argument structures in text using statistical Natural Language Processing".

[87] (Hsu & Ho, 1998)

[88] (Hovy, 2014) investigates a largely unsupervised approach to learning "interpretable, domain-specific entity types from unlabelled text". The results suggest that it is possible to learn domain-specific entity types from unlabelled data.

[89] Aletras and his team recently made the headlines by designing a system (relying on the supervised learning method described below) capable of predicting the outcome of cases tried by the European Court of Human

2.2.3. Artificial Neural Networks:

This group of tools owes its name to the fact that it was inspired by the study of humans' central nervous systems. As a metaphor to understand the key characteristics of artificial neural networks as a method, however, it is of limited help, save for the fact that the process through which the "weights" -see below- of a neural networks are adapted in response to training data may in some sense be related to the way our own intelligence is deemed to work.

Now, to understand the specificity of neural networks as a method, a useful starting point is to consider the way networks that are designed to recognise handwriting might work. Confronted with an image (containing handwriting), the neural network's first layer of "neurons", called "input neurons" will be activated by the pixels in that image. The activated input neurons send data (via "synapses") to a second layer of neurons, and then again (via more synapses) to a third layer of output neurons.[90] Whether or not the output neurons correctly determine which character was read will depend on the adequacy of the parameters stored in the synapses. These parameters are called "weights"; they reflect a function that is updated in the course of a learning process. One of the advantages of this method (compared to probabilistic methods) is that the algorithm that adapts the weights in response to training data tends to be more robust (and "scalable", so that it can be trained on very large amounts of data).

There are three major learning paradigms whose suitability largely depends on the task at hand.

1. Handwriting recognition is one of the tasks that lends itself to a ***supervised learning*** approach: one might feed a system with a set of example pairs (x, y) where "x" corresponds to the images containing handwriting and "y" identifies which character is being read. The aim of the learning process is to find a function $f: X \rightarrow Y$ that matches the example pairs.
2. In ***unsupervised learning***, by contrast, some unlabelled data x is given, and the aim is to infer a function that reveals some hidden structure within the data. Based on this hidden structure, a system might for instance detect anomalies (detection of fraudulent activity is a common application). The success of such a method largely depends on whether the system was correct in its assumption that the majority of instances in the unlabelled data was "normal". (Removing

Rights based solely on the textual content of published judgments (Aletras, Tsarapatsanis, Preoţiuc-Pietro, & Lampos, 2016).

[90] In more complex systems there will be extra layers of neurons.

anomalous data from the given set –in which case we have switched to a supervised learning model- will increase the accuracy of the system.)

3. In ***reinforcement learning***, the set of data x is not given. Instead it is generated by an agent's interaction with the environment. The aim of the learning process is to come up with an action-selection policy that minimises some measure of long-term cost. The key challenge is to determine the latter. 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.[91] Yet the difficulty inherent in determining the long term-cost function that is to inform the action-selection policy of an expert system increases dramatically when one switches from games to real-life conundrums. This increase in difficulty largely stems from the fact that constructing the real-world objective (and the cost function that comes with it) is necessarily -in real life- an ethically loaded endeavour; the characterisation of the events that are to structure the learning phase cannot but be a value-laden interpretive exercise. Whereas the real-world objective in the game of GO (or in systems designed to discover new, stable atomic structures, say) can be described in an algorithm (in which case the teacher's job can be automated to some extent), the real-world objective that presides over professions-specific systems cannot.

2.3. Challenges inherent in computer systems designed for the professions

The success of machine learning techniques in scientific fields (and within games) stems in large part from the fact that unchanging and reliable laws that are generally applicable to one's data can in principle be identified. These laws condition the learning phase. The attribution of labels for instance reflects the teacher's understanding of the laws structuring the domain in question, whether it be quantum mechanics laws determining the energy level of atomic structures or otherwise. As a work in progress that we are all in the process of re-shaping as we learn to live together, the laws and standards governing social interactions and what we owe to each other[92] (obviously going beyond legal requirements) are, by contrast, the object of constant contestation and evolution, and something of fundamental value is lost when such norms are made less contestable (because they are "enshrined" in some opaque and rigid computer system). To try and nevertheless subsume some constant and universal laws that can be relied on "uncontroversially" by the system's teacher during the learning phase (and subsequently mostly forgotten because their work is made invisible), can only be done if one is prepared to not only compromise the essential contestability of the ethical standards we live by but also, potentially, ignore important features of our moral

[91] (Silver et al., 2016)

[92] If it ever were possible, the day such laws stopped being re-constructed by us, frozen into some automated computer system, would be a dire one for humanity.

landscape, for some of the things we value most are not easily subsumable under constant and universal laws.

In the specific context of designing automated decision-making processes in a way that allows for accountability and compliance with "key standards of legal fairness", Kroll and others emphasise, among other things, "that computer scientists should focus on creating algorithms that are *reviewable*, not just compliant with the specifications that are generated in the drafting process"[93] (to which I would add "and not just compliant with existing legal standards", for legal regulation is unlikely to ever match the speed of technological change). Interestingly, this commitment to reviewability (to reflect changes in a society's normative landscape), while laudable, may prove tricky to implement for reasons that have nothing to do with computer science, and everything to do with the fact that we are rather lazy normative animals. Having been given the chance to relax and sit back to enjoy the benefits of simplified practical reasoning (thanks to automated decision-making processes), we may not be as easily stirred into "critically re-assessing" a system's normative assumptions as we might think. The more performant such automated decision-making processes become, the more our "normative muscle" may go limp (and unlikely to be mobilised again).[94] The difficulty inherent in preserving the ability to review (and oversee) the normative assumptions informing computer systems has significant implications for their design within the professions, whether they are conceived to support educators, lawyers, bankers, healthcare providers (or indeed those designing the architecture shaping our virtual interactions).

The problem is that the technical language currently underlying discussions of computer systems' design issues (relating to both the construction of the real-world objective that structures the learning phase and the labelling process and/or selection of classifying features) discourages public engagement. This in turn lessens incentives for designers to build into such systems mechanisms that allow for continuous public oversight[95] of the values and assumptions that have influenced the design of the system and its learning phase. Now one might hope that one can bypass the difficulties mentioned above by designing highly specialised expert systems that are meant to tackle specific tasks within well-defined domains, as is the case of most professions-specific expert systems today.

---

[93] (Kroll et al., 2016)

[94] This "habituation" aspect is discussed in detail in my forthcoming …

[95] Here the concept of "oversight" is preferred to that of "transparency". Indeed, (Kroll et al., 2016) point out that it is "far from obvious" what form, if any transparency should take: "Perhaps the most obvious approach is to disclose a system's source code, but […] [t]he source code of computer systems is legible only to experts, and even they often struggle to understand what it will do: […] In systems that […] machine learning, the decisional rule itself may emerge from the specific data under analysis, in ways that no human can explain". They also remind us that full transparency can in some cases be counter-productive (opacity may be needed to "prevent tax cheats or terrorists from gaming the system") or legally barred (because of privacy concerns, for instance).

2.3.1 Existing professions-specific computer systems

In the field of education, AI's latest developments have yet to be leveraged to their full extent. In their recent report - *Intelligence unleashed: an argument for AI in education*[96], Luckin and others put forward a compelling case for the rigorous and systematic development of AI within education, whereby "hundreds and then thousands of individual AIEd components, developed in collaboration with educators, conformed to uniform international data standards, and shared with researchers and developers worldwide […] enable system-level data collation and analysis that help us learn much more about learning itself and how to improve it"[97] – and tailor it.

In contrast to education, healthcare is probably the field that has so far benefited most from recent advances in machine learning. The ability to learn from vast amounts of patient records, clinical observations etc. has enabled the diagnosis (and treatment) of conditions hitherto unrecognised. Automated systems of various kinds (whether they interpret medical imaging, help form an accurate diagnosis – or prognosis-, or prescribe drugs) are now so pervasive that they have become the norm in the way medicine is practiced in most developed countries.

In the legal domain, automated systems tackle tasks ranging from automated document classification (see "Predictive Coding", developed for the American discovery process) to producing a range of legal documents such as contracts, divorce or incorporation papers. Harnessing recent progress in natural language processing, "Ross Intelligence", a more generalist legal research tool (built upon the IBM Watson platform), is capable of extracting relevant information from the vast amounts of (unstructured) legal reports to accurately answer lawyers' queries (which do not need to rely on carefully selected keywords). The more queries they get (and hence the larger their user base is) the better these systems become.

2.3.2. The axiological dimension of professions-specific computer systems

Whether such systems provide intelligent tutoring, determine the optimal treatment for a particular type of illness or handle matrimonial property disputes, some value judgments inevitably need to be made. The following paragraphs aim at remedying this axiological dimension's lack of exposure by highlighting some of its key aspects.

[96] (Luckin, Holmes, Griffiths, & Forcier, 2016)

[97] (Luckin et al., 2016, p. 12)

First, and most importantly, a real-life narrative needs to be "labelled" as a problem that lends itself to resolution by a particular type of computer system. This labelling requires the professional to decide what element(s) are relevant to the particular problem at stake, whether there are any other -potentially more important, but non-conceptualised- problems that need to be considered etc. This labelling process requires not only great skill but also a kind of judgment that is not altogether dissimilar to the type of "human wisdom" – or *anthrôpinè Sophia*- I discuss elsewhere (in the context of considering the possibility and merits of expertise in ethics[98]). For the humility demanded by *sophia* -requiring acute awareness of what one does not know- is key to countering the well-known propensity to "treat every problem as if it were a nail […], if the only tool you have is a hammer".[99] The Susskinds acknowledge the issues underlying the fact that "[m]any of the problems that [professionals] tackle are in fact defined by the solutions that the professions themselves have developed. So when we say, for example, that a client has a tax or accounting problem or that a patient has a dental or surgical concern, these very characterizations of the concerns are framed in terms of the categorizations and capabilities of professional providers".[100] Yet the Susskinds discuss those issues not so much to highlight the extent to which the labelling process is itself value-laden, but rather to criticise the inherent conservatism in the way change and improvement is contemplated for the professions.[101]

As for the extent to which future "expert systems" could competently handle "ethical issues", the Susskinds are rather upbeat about the prospect:

> *"[F]uture 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."*[102]

Now one could, in principle, endeavour to distil as many of the ethical principles or issues that might pertain to a particular professional domain as possible (at a given time) into a set of formalised constraints, which the "expert system" would have to take into account. These formalised ethical constraints would have to be acknowledged as a necessarily imperfect and incomplete instantiation of the ethical principles at stake. As such they would have to remain transparent and easily contestable so as to allow for frequent updating by the system's users (which is not

[98] See …

[99] (Maslow, 1966)

[100] (R. Susskind & Susskind, 2015, p. 42)

[101] "Although our professions are failing in significant ways, they are not incentivised to work differently" (R. Susskind & Susskind, 2015, p. 43)

[102] (R. Susskind & Susskind, 2015, p. 280)

impossible, but a tall order[103]). It is even possible that such a system would "lead to [better] outcomes when measured against common moral requirements (for example, to minimize harm to noncombatants)".[104] Even if one shares the Susskinds' optimism regarding the feasibility of such integration of ethical constraints (my scepticism mainly stems from the fact that one would have to iron out their inherent contestability[105]), the question remains as to whether one should in fact entrust an automated system with, say, the tutoring of a child, the handling of a divorce or the treatment of a patient, quite independently of the quality of their outcomes (note that this question assumes a model of automated systems designed to *replace* professionals even in the most significant of their tasks, a model which I do not deem desirable).

The Susskinds acknowledge this outcome-independent line of argument by referring to Sandel's "moral limits" objection[106]: could it be that we feel uncomfortable about the idea of an increasing number of professional "tasks" being handled by computer systems for reasons that are similar in kind to those that underlie our repugnance at body organs being traded like ordinary goods (or the right to immigrate to the US being "bought" for $500000)? Sandel seeks to capture what underlies our concern about the proliferation of market norms (which, in the context of our 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"[107]- is quickly dealt with, the Susskinds pointing out that "expert 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—*

[103] In the Golden age of "Rule-based systems", some hoped that their relative simplicity would allow their users to update them without any input from IT specialists (this turned out to be over-optimistic).

[104] (R. Susskind & Susskind, 2015, p. 282)

[105] The system would have to allow for the possibility that some of those ethical principles will be subject to contestation. Quite how such a system would reflect the contested nature of some ethical concepts (and allow for potentially conflicting updates to its principles) remains to be seen.

[106](Sandel, 2012; R. Susskind & Susskind, 2015)

[107] (R. Susskind & Susskind, 2015, p. 241)

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".[108]

Now 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".[109] 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.[110] 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.

That moral character is fleshed out, by contrast, in the account of the professions developed in section 1.1, 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

[108] (R. Susskind & Susskind, 2015, p. 243)

[109] (R. Susskind & Susskind, 2015, p. 242)

[110] 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" (R. Susskind & Susskind, 2015, p. 291)

shape by our professions. Yet the outrage at the cruelty[111] instantiated in scandals such as Winterbourne's testify to its endurance – and salience within the professions.

Now let us imagine –for the sake of the argument- that the Susskinds are happy to endorse the above. If it is our commitment to moral equality that ultimately grounds 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"[112], 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".[113]

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 expert 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. The obstacles that are to be overcome to that end are many – and varied. Aside from reducing professionals' cognitive load, expert systems can and should be designed to counter the effects of routinisation, raise awareness of seemingly peripheral considerations (and one's fallibility[114]) and, most importantly,

---

[111] In his forthcoming *Humanity without dignity: moral equality, respect and human rights*, Andrea Sangiovanni successfully argues that it is a rejection of social cruelty -defined as "the unauthorized, harmful and wrongful use of another's vulnerability to attack or obliterate their capacity to develop and maintain a sense of self" that stands at the centre of our commitment to moral equality (rather than "human dignity").

[112] (Williams, 1973)

[113] (Williams, 1973, p. 114)

[114] Klein's pre-mortem method is a good example of an approach that could relatively easily be turned into an automated decision-aid which would be likely to benefit healthcare providers, educators and lawyers alike. To quote Klein and Kahneman: "We agree that the introduction of algorithms and other formal decision aids in organizations will often encounter opposition and unexpected problems of implementation. Few people enjoy being replaced by mechanical devices or by mathematical algorithms, and many devices and algorithms function less well in the real world than on the planning board (Yates, Veinott, & Patalano, 2003). Even decision aids and procedures that leave the authority of the decision maker intact— decision analysis is a salient example—are often resisted, for both good and bad reasons […] Despite our different attitudes toward formal methods, we agree on the potential of semi-formal strategies. An example is the premortem method (Klein, 2007) for reducing overconfidence and improving decisions". (Kahneman & Klein, 2009)

better listen to and engage with the person seeking professional expertise. Possible applications are outlined below. No matter how sophisticated they become, none of them is ever meant to *replace*, individually or collectively, members of the professions (though their number will dwindle).

III. Conclusion

The fast-expanding reach and prowess of Artificial Intelligence (AI) 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, and to reflect upon their desirability.

There is little doubt that the growing availability of automated systems within the professions is likely to improve the affordability and accessibility of professional services. But there is a danger that it will compromise, rather than foster, the extent to which the professions live up to the ethical integrity that is concomitant to their particular responsibility. To grasp the latter, one must ask what differentiates professionals from other experts. The answer, I argue, lies in the characteristics of the relationship between the professions and individuals who need their services (rather than the characteristics of the occupation itself). At the heart of this relationship is a very particular type of vulnerability, which healthcare providers, educators, bankers, lawyers (and those designing the architecture that shapes our virtual interactions) all encounter. Unlike an aptly vague reference to "public service", an endeavour to ground the responsibility of the professions in the concerns that stem from this particular type of vulnerability -including, most prominently, a concern for moral equality- has significant conceptual *and* normative implications. For it not only leads to re-delineating the scope of the professions in a way that is both culturally and historically dependent; it also calls into question the wisdom of an instrumental, outcome-focused success criterion for future, professions-specific computer systems.

Given the pace of recent progress in AI, designing computer systems that can increase both the accessibility and quality of professional services is not the challenge that it once was. Yet in the race to increase the performance of such systems, a narrow focus on easily measurable outcomes is all too likely – and risky. If it is to do more than rhetorical work, the professions' particular ethical responsibility needs to be made a determining factor in the design of AI systems: this means -among other things- designing computer systems that improve the situational (and ethical) awareness of professionals by systematically challenging routine perceptions and modes of thoughts (and hence countering the effects of professional habituation).

Bibliography

Abbott, A. (2014). *The system of professions: An essay on the division of expert labor*: University of Chicago Press.

Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. *PeerJ Computer Science, 2*, e93. doi:10.7717/peerj-cs.93

Ashley, K. D. (1992). Case-based reasoning and its implications for legal expert systems. *Artificial Intelligence and Law: an international journal, 1*(2-3), 113-208.

Bauman, Z. (1989). *Modernity and the Holocaust*: Cornell University Press.

Black, E., & Wallace, B. (2001). *IBM and the Holocaust: The strategic alliance between Nazi Germany and America's most powerful corporation*: Crown Publishers New York.

Campbell, E. M., Sittig, D. F., Guappone, K. P., Dykstra, R. H., & Ash, J. S. (2007). *Overdependence on technology: an unintended adverse consequence of computerized provider order entry.* Paper presented at the AMIA.

Charmaz, K. (1983). Loss of self: a fundamental form of suffering in the chronically ill. *Sociology of health & illness, 5*(2), 168-195.

Chesterton, G. K. (1955). The Twelve Men *Tremendous Trifles* (pp. 57-58). New York: Sheed & Ward.

Crocker, J., Fiske, S. T., & Taylor, S. E. (1984). Schematic bases of belief change *Attitudinal judgment* (pp. 197-226): Springer.

Dillard, J. F. (2003). Professional services, IBM, and the holocaust. *Journal of Information Systems, 17*(2), 1.

Durkheim, E. (1957). *Professional Ethics and Civic Morals*. London: Routledge.

Evetts, J. (2011). A new professionalism? Challenges and opportunities. *Current sociology, 59*(4), 406-422.

Feigenbaum, E. A. (1977). *The art of artificial intelligence. 1. Themes and case studies of knowledge engineering*. Retrieved from

Flood, J. (2008). Partnership and professionalism in global law firms: resurgent professionalism? In D. Muzio, S. Ackroyd, & J.-F. Chanlat (Eds.), *Redirections in the Study of Expert Labour: established professions and new expert occupations* (pp. 52-74). New York: Palgrave MacMillan.

Freidson, E. (1970). *Professions of medicine: as study of the sociology of applied knowledge*. New York: Dodd, Mead & Co.

Freidson, E. (1999). Theory of professionalism: Method and substance. *International review of sociology, 9*(1), 117-129.

Garth, B. G. (1983). Rethinking the Legal Profession's Approach to Collective Self-Improvement: Competence and the Consumer Perspective. *Wis. L. Rev.*, 639.

Gerhardt, U. (1987). Parsons, role theory, and health interaction. In G. Scambler (Ed.), *Sociological theory and medical sociology.* (pp. 110-133). London: Tavistock.

Greenwood, E. (1957). Attributes of a profession. *Social work*, 45-55.

Gunz, H. P., & Gunz, S. P. (2002). The lawyer's response to organizational professional conflict: an empirical study of the ethical decision making of in-house counsel. *American Business Law Journal, 39*(2), 241-288.

Harper, S. J. (2013). *The lawyer bubble : a profession in crisis*. New York: Basic Books.

Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983). Building expert system.
Hickson, D., & Thomas, M. (1969). Professionalization in Britain: A preliminary measurement. *Sociology, 3*(1), 37-53.
Hovy, D. (2014). *How well can we learn interpretable entity types from text?* Paper presented at the ACL.
Hsu, C.-C., & Ho, C.-S. (1998). *A hybrid case-based medical diagnosis system.* Paper presented at the Tools with Artificial Intelligence, 1998. Proceedings. Tenth IEEE International Conference on.
Johnson, T. (1972). IMPERIALISM AND THE PROFESSIONS: Notes on the Development of Professional Occupations in Britain's Colonies and the New States. *The Sociological Review, 20*(S1), 281-309.
Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *Am Psychol, 64*(6), 515-526. doi:10.1037/a0016755
Kearney, R. (1998). *Poetics of imagining: modern to post-modern*. New York: Fordham University Press.
Keogh, B. (2013). *Review into the quality of care and treatment provided by 14 hospital trusts in England: overview report*: NHS.
Kolodner, J. L. (1992). An introduction to case-based reasoning. *Artificial Intelligence Review, 6*(1), 3-34.
Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine, 23*(1), 89-109.
Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable Algorithms. *University of Pennsylvania Law Review, 165*.
Lagu, T., Goff, S. L., Hannon, N. S., Shatz, A., & Lindenauer, P. K. (2013). 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. *Joint Commission Journal on Quality and Patient Safety, 39*(1), 7-15.
Larson, M. S., & Larson, M. S. (1979). *The rise of professionalism: A sociological analysis* (Vol. 233): Univ of California Press.
Lawrence, N. (2016). Don't panic: deep learning will be mostly harmless. Retrieved from http://inverseprobability.com/2016/11/29/new-directions-in-kernels-and-gaussian-processes
Lombarts, K. M. J. M. H., Plochg, T., Thompson, C. A., Arah, O. A., & on behalf of the, D. P. C. (2014). Measuring Professionalism in Medicine and Nursing: Results of a European Survey. *PLoS ONE, 9*(5), e97069. doi:10.1371/journal.pone.0097069
Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: an argument for AI in education*. Retrieved from London:
Maslow, A. (1966). The Psychology of Science: A Reconnaissance. *South Bend, Indiana: Gateway*.
Maxwell, T., & Schafer, B. (2010). Natural language processing and query expansion in legal information retrieval: Challenges and a response. *International Review of Law, Computers & Technology, 24*(1), 63-72.
Moorhead, R. (2010). Lawyer specialization–managing the professional paradox. *Law & Policy, 32*(2), 226-259.
Moorhead, R. (2014a). Precarious Professionalism: Some Empirical and Behavioural Perspectives on Lawyers. *Current legal problems, 67*(1), 447-481. doi:10.1093/clp/cuu004

Moorhead, R. (2014b). Precarious Professionalism: Some Empirical and Behavioural Perspectives on Lawyers. *Current legal problems*, cuu004.
Moorhead, R., Sherr, A., & Paterson, A. (2003). Contesting professionalism: legal aid and nonlawyers in England and Wales. *Law & Society Review, 37*(4), 765-808.
Moorhead, R., Sherr, A., Webley, L., Rogers, S., Sherr, L., Paterson, A., & Domberger, S. (2001). Quality and cost: final report on the contracting of Civil, Non-Family Advice and Assistance Pilot.
Muzio, D., Brock, D. M., & Suddaby, R. (2013). Professions and Institutional Change: Towards an Institutionalist Sociology of the Professions. *Journal of Management Studies, 50*(5), 699-721. doi:10.1111/joms.12030
Neuberger, D. (2016, 15 June) *The Lord Slynn Memorial Lecture 2016: Ethics and advocacy in the twenty-first century*. supremecourt.uk.
O'Fallon, M. J., & Butterfield, K. D. (2005). A review of the empirical ethical decision-making literature: 1996-2003. *Journal of business ethics, 59*, 375-413.
Pal, S. K., Dillon, T. S., & Yeung, D. S. (2012). *Soft computing in case based reasoning*: Springer Science & Business Media.
Parsons, T. (2012). *The social system*. New Orleans: Quid Pro, LLC.
Paterson, A., & Sherr, A. (1992). Quality, Clients and Legal Aid. *New Law Journal, 142*, 783.
Sandel, M. J. (2012). *What money can't buy: the moral limits of markets*: Macmillan.
Sangiovanni, A. (2017). *Humanity without dignity: moral equality, respect and human rights*. Cambridge, MA: Harvard University Press.
Saunders, S. A. M. C., & Wilson, P. A. (1933). *The professions*.
Sherr, A., Moorhead, R., & Paterson, A. (1994). Lawyers-the quality agenda vol. 1 Assessing and developing competence and quality in legal aid; the report of the Birmingham Franchising Pilot.
Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., . . . Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature, 529*(7587), 484-489. doi:10.1038/nature16961
Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human-Computer Studies, 52*(4), 701-717. doi:10.1006/ijhc.1999.0349
Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991-1006.
Sommerlad, H. (1995). Managerialism and the legal profession: a new professional paradigm. *International Journal of the Legal Profession, 2*(2-3), 159-185.
Susskind, R., & Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts*: Oxford University Press, USA.
Susskind, R. E. (1998). *The future of law: facing the challenges of information technology*: Oxford University Press.
Williams, B. (1973). The idea of equality *Problems of the self: philosophical papers 1956-1972*. Cambridge: Cambridge University Press.
Williams, B. (2000). Naturalism and Genealogy. In E. Harcourt (Ed.), *Morality, Reflection and Ideology* (pp. 148-161). Oxford: Oxford University Press.