McGill Law Journal
Revue de droit de McGill

Governing A.I.'s Professional Advice

Claudia E. Haupt

Résumé de l'article

Comment régler le conseil professionnel fournir par l'intelligence artificielle
(IA)? Les relations traditionnelles entre client et professionnel ou entre
médecin et patient sont gouvernées par un cadre de règles juridiques qui
assure les conseils fiables, complets et précis pour le client ou patient — afin
qu'ils puissent prendre avec confiance des décisions de vie importantes. Mais
un cadre semblable n'existe pas quand il s'agit des conseils donnés par IA. Cet
article suggère que la première étape pour régler ce nouveau type de conseil
professionnel consisterait de se tourner vers le cadre qui existe déjà pour
régler les conseils professionnels. Mettant en relief la relation
client-professionnel, ce cadre conventionnel souligne les points d'accès
réglementaires où la loi peut atteindre l'objectif de garantir de bons conseils,
qu'ils soient rendus par des êtres humains ou par l'IA.

GOVERNING A.I.'S PROFESSIONAL ADVICE

Claudia E. Haupt[*]

How should we govern professional advice given by artificial intelligence (AI)? The traditional professional-client or doctor-patient relationship is governed by a specific set of legal rules that constitute the legal framework of professional advice-giving. The goal of this legal framework is to ensure the client or patient receives reliable, comprehensive, and accurate advice in order to make important life decisions. But such a regime does not exist when AI gives professional advice. This article suggests that the first step in regulating professional AI should be to turn to the existing framework that regulates professional advice-giving. In focusing on the professional-client relationship, it foregrounds the regulatory access points at which the law can achieve the goal of ensuring good advice, whether rendered by humans or AI.

Comment régler le conseil professionnel fournir par l'intelligence artificielle (IA)? Les relations traditionnelles entre client et professionnel ou entre médecin et patient sont gouvernées par un cadre de règles juridiques qui assure les conseils fiables, complets et précis pour le client ou patient — afin qu'ils puissent prendre avec confiance des décisions de vie importantes. Mais un cadre semblable n'existe pas quand il s'agit des conseils donnés par IA. Cet article suggère que la première étape pour régler ce nouveau type de conseil professionnel consisterait de se tourner vers le cadre qui existe déjà pour régler les conseils professionnels. Mettant en relief la relation client-professionnel, ce cadre conventionnel souligne les points d'accès réglementaires où la loi peut atteindre l'objectif de garantir de bons conseils, qu'ils soient rendus par des êtres humains ou par l'IA.

Introduction

In the early days of the internet, two competing regulatory approaches emerged. The first approach—derived in part from the ethics of the communities who were early internet users[1]—was based on libertarian idealism, internet exceptionalism, and perhaps even internet utopianism.[2] Cyberspace was imagined as a sphere of human interaction separate and distinct from others.[3] The second approach, by contrast, was more skeptical of this new and unregulated environment, and more willing to explore how regulatory mechanisms could be imposed on a space that was perhaps not entirely unlike other forms of social interaction[4] (the famous "law of the horse" debate is but one instantiation of the early critique of internet exceptionalism[5]). Today, the question no longer is whether the internet should be regulated at all, but to what extent it should be regulated.[6]

We are now at a similar regulatory crossroads in the world of artificial intelligence (AI).[7] On the one hand, some suggest that AI must remain

---

[1] See generally Walter Isaacson, *The Innovators: How a Group of Hackers, Geniuses, and Geeks Created the Digital Revolution* (Simon & Schuster: New York, 2014) at 383–400 (describing early communities of internet users).

[2] See e.g. David R Johnson & David Post, "Law and Borders – The Rise of Law in Cyberspace" (1996) 48:5 Stan L Rev 1367; David G Post, "Governing Cyberspace" (1996) 43:1 Wayne L Rev 155; Joel R Reidenberg, "Governing Networks and Rule-Making in Cyberspace" (1996) 45 Emory LJ 911.

[3] See e.g. Lawrence Lessig, "The Zones of Cyberspace" (1996) 48:5 Stan L Rev 1403 at 1404.

[4] See Jack Goldsmith & Tim Wu, *Who Controls the Internet? Illusions of a Borderless World* (New York: Oxford University Press, 2006).

[5] See Frank H Easterbrook, "Cyberspace and the Law of the Horse" (1996) U Chicago Legal F 207; Lawrence Lessig, "The Law of the Horse: What Cyberlaw Might Teach" (1999) 113:2 Harv L Rev 501.

[6] See e.g. Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media* (New Haven: Yale University Press, 2018) at 10, 13; Jonathan Zittrain, *The Future of the Internet—And How to Stop It* (New Haven: Yale University Press, 2008). See also Rebecca Crootof & BJ Ard, "Structuring Techlaw", 34 Harv JL & Tech [forthcoming in 2021].

[7] As Ryan Calo notes, "[t]here is no straightforward, consensus definition of artificial intelligence. AI is best understood as a set of techniques aimed at approximating some aspect of human or animal cognition using machines." See Ryan Calo, "Artificial Intelligence Policy: A Primer and Roadmap" (2017) 51:2 UC Davis L Rev 399 at 403–04. For the purposes of this discussion, I am primarily interested in machine learning. See generally David Lehr & Paul Ohm, "Playing with the Data: What Legal Scholars Should Learn About Machine Learning" (2017) 51:2 UC Davis L Rev 653 (explaining the basic concepts of machine learning for a legal audience).

largely free from regulatory interference.[8] On the other hand, various regulatory approaches are being explored.[9] This article picks up one slice of the AI universe—professional advice rendered by AI—to explore appropriate forms of regulation. I suggest that we first turn to the existing regulatory framework of professional advice-giving in regulating professional AI.[10] This targeted intervention is limited in its emphasis on professional-use AI. But its broader implication is the context-specific regulation of AI, with a focus on the underlying social relationships among humans. Keeping the social relationships among humans at the center of attention, I suggest, is the appropriate way to approach the larger pressing question of how to govern AI. This approach is vividly illustrated in the professional context.

Professionals have knowledge that their clients or patients lack but need in order to make important life decisions. Clients or patients who consult their doctors, lawyers, or accountants know that these professionals are regulated by a legal framework designed to ensure they give good advice. Only good advice—accurate, comprehensive, and in accordance with professional standards—enables clients or patients to make fully informed, autonomous choices about their own financial or physical wellbeing or other important matters. Before professionals may give advice, they typically must be licensed to practice. Bad professional advice is subject to malpractice liability, fiduciary duties exist between professional and client, and certain professional activities are subject to informed con-

---

8   See e.g. Andrew Burt, "Leave A.I. Alone", *The New York Times* (4 January 2018), online: <nytimes.com> [perma.cc/BDP4-PLGQ].

9   See e.g. Andrew Tutt, "An FDA for Algorithms" (2017) 69:1 Admin L Rev 83; Ignacio N Cofone, "Servers and Waiters: What Matters in the Law of A.I." (2018) 21:2 Stan Tech L Rev 167; Matthew U Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies" (2016) 29:2 Harv JL & Tech 353 at 393. For a comparative perspective, see e.g. Peter Georg Picht & Gaspare Tazio Loderer, "Framing Algorithms – Competition Law and (Other) Regulatory Tools" (2018) Max Planck Institute for Innovation and Competition Research Paper No 18-24, online: <ssrn.com/abstract=3275198> (looking to financial regulation and data protection regulation in the European Union as to regulating AI). See also United States, National Science & Technology Council Committee on Technology, Executive Office of the President, *Preparing for the Future of Artificial Intelligence* (October 2016) (looking at regulatory approaches to AI from a United States perspective).

10   I have previously made parts of this argument elsewhere with a specific focus on medical advice, see Claudia E Haupt, "AI in the Doctor-Patient Relationship: Identifying Some Legal Questions We Should Be Asking" (19 June 2018), online: *Data & Society Points* <points.datasociety.net> [perma.cc/9HD5-Z9RL] [Haupt, "AI in the Doctor-Patient Relationship"]; Claudia E Haupt, "The Algorithm Will See You Now" (26 October 2018), online (blog): *Balkinization* <balkin.blogspot.com> [perma.cc/5X8A-TGCX]; Claudia E Haupt, "Artificial Professional Advice" (2019) 18:3 Yale Journal of Health Policy, Law, & Ethics 55 / (2019) 21:3 Yale Journal of Law & Technology 55 [Haupt, "Professional Advice"].

sent. These are the key elements of the regulatory framework that governs human professionals.[11] Inserting AI into this traditional professional advice-giving relationship, however, potentially raises new regulatory challenges.

Scholars note that "[h]istorically, humans have entered into symbiotic intimate relationships with those with more (or more perceived) knowledge or expertise."[12] AI may, in fact, just be another iteration of this phenomenon, which is based on a social relationship that has as its end the transfer of knowledge or expertise. In turning to experts, individuals seek personalized responses from generalized knowledge to address their individual situations. Thus, humans

> have made these arrangements with healers, shamans, priests, medicine men, mystics, quacks and sellers of snake oil, doctors and, most recently, healthcare institutions such as hospitals and insurers. Automation is poised to make the most persuasive case yet for such an intimate relationship with our minds and bodies.[13]

The expectation in the medical field and elsewhere is that "increasingly capable machines will, in due course, be capable of generating bodies of practical expertise that can resolve the sort of problems that used to be the sole province of human experts in the professions."[14] Introducing AI into the process of professional advice-giving can take many forms. However, at least initially, these changes occur in a traditional existing regulatory framework that governs professionals and their advice.[15]

---

[11]   See Haupt, "AI in the Doctor-Patient Relationship", *supra* note 10.

[12]   Nicolas P Terry, "Appification, AI, and Healthcare's New Iron Triangle" (2018) 20:2 J Health Care L & Pol'y 117 at 125. It might be debatable whether the possession of actual knowledge is in fact necessary. Discussing "professional mystique", Richard Posner—citing the lack of real therapeutic knowledge in medicine "in the Middle Ages in Italy, where medicine was a highly prestigious profession"—suggests that "[t]he key to classifying an occupation as a profession ... is not the actual possession of specialized, socially valuable knowledge; it is the *belief* that some group has such knowledge". But "[t]he fact that a profession cultivates professional mystique does not prove that it lacks real knowledge; modern medicine is a case in point." See Richard A Posner, "Professionalisms" (1998) 40:1 Ariz L Rev 1 at 2–4.

[13]   Terry, *supra* note 12 at 125.

[14]   Richard Susskind & Daniel Susskind, *The Future of the Professions: How Technology Will Transform the Work of Human Experts* (Oxford: Oxford University Press, 2015) at 226. For a thoughtful rebuttal, see Frank Pasquale, "Automating the Professions: Utopian Pipe Dream or Dystopian Nightmare?" (15 March 2016), online: *Los Angeles Review of Books* <lareviewofbooks.org> [perma.cc/7R68-6NXH]. See also Steven J Frank, "Tort Adjudication and the Emergence of Artificial Intelligence Software" (1987) 21:3 Suffolk UL Rev 623 at 639–47 (discussing professionals and professional liability).

[15]   See Haupt, "Professional Advice", *supra* note 10.

As Jack Balkin recounts the story of the Golem of Prague—legend has it that a wise sixteenth-century Rabbi created a life-like creature out of clay to "deal with threats to the Jewish community"—he notes that the fact that "nothing goes wrong" in the process is because "the Golem is programmed and employed by the Maharal, a man of the greatest piety and learning. Only a truly righteous man, or a saint, you might say, is capable of using the Golem only for good."[16] Likewise, Balkin suggests, "[w]hen we talk about robots, or AI agents, or algorithms, we usually focus on whether they cause problems or threats. But in most cases, the problem isn't the robots; it's the humans."[17] The explanation he offers is fourfold: (1) it is humans who design, program, connect the algorithms, "and set them loose"; (2) it is humans who decide how, when, and for what purpose to use them; (3) humans select the data to program algorithms, a process that "contains the residue of earlier discriminations and injustices"; and (4) perhaps most importantly for present purposes, the "technologies mediate social relations between human beings and other human beings. Technology is embedded into—and often disguises—social relations."[18]

This insight shifts the lens from asking "what robots did or what AI agents did"[19] to what humans did by employing these technologies. The proper focus when assessing the role of algorithms, consequently, should be on the question of

> how the algorithms are engaged in reproducing and giving effect to particular social relations between human beings. These are social relations that produce and reproduce justice and injustice, power and powerlessness, superior status and subordination. The robots, AI agents, and algorithms are the devices through which these social relations are produced, and through which particular forms of power are processed and transformed.[20]

Similarly, I suggest we should first look to the regulatory framework governing the human professional-client relationship—a very specific kind of social relationship subject to a particular set of legal ramifications, with unique knowledge and power imbalances, and distinctive normative values—to determine how best to regulate professional AI.[21] And interest-

---

[16]   Jack M Balkin, "The Three Laws of Robotics in the Age of Big Data" (2017) 78:5 Ohio St LJ 1217 at 1222–23 [Balkin, "The Three Laws of Robotics"].

[17]   *Ibid* at 1223.

[18]   *Ibid.*

[19]   *Ibid*.

[20]   *Ibid.*

[21]   A parallel concern exists in contract law. See Andrea M Matwyshyn, "The Law of the Zebra" (2013) 28:1 Berkeley Tech LJ 155 at 158, noting that

ingly, even AI regulation skeptics seem to agree that sector-specific regulation may be appropriate.[22]

This article proceeds in two parts. Part I outlines the existing legal framework of professional advice-giving and investigates how well it maps onto concerns introduced by AI. In doing so, it will identify potential shortcomings of the current framework that make adjustments necessary. Part II will explore some adjustments that might be made to the existing framework to make it more responsive to the introduction of AI into professional advice-giving.

I.   The Framework of Professional Advice-Giving

This Part considers professional licensing, professional speech protection, professionals' fiduciary duties, professional malpractice liability, and professional ethics as core regulatory elements. The initial assessment of whether the existing regulatory regime of professional advice-giving is responsive to new questions raised by AI starts with mapping normative concerns for each of the regulatory access points to see how well they align with the introduction of AI. The focus ought to be on the interests underlying the professional-client relationship, a social relationship based on knowledge asymmetries, expertise, loyalty, and trust.[23]

As this discussion will illustrate, the existing regime is, for the most part, responsive to concerns raised by AI when viewed from a normative perspective. To the extent the existing framework is not responsive, this Part will identify some of its shortcomings.

---

courts are derailing traditional contract law approaches with an overzealous focus on the role of technology in disputes. Instead of asking whether a technology-specific 'law of the horse' should be crafted to fill gaps in existing law in technology contexts, courts now ask whether technology-specific approaches should usurp the traditional space of contract law ... Instead of using contract law in its traditional form to resolve disputes, and supplementing it with technology-exceptionalist approaches only where true novelty exists, some courts now reach aggressively for technology exceptionalist approaches as a first cut.

[22]   See e.g. Burt, *supra* note 8 ("[t]his is not, of course, to suggest that artificial intelligence should never be regulated. But if the past is any guide, treating it as a collection of separate technologies, in separate sectors, is destined to be the most effective way to control the benefits it creates — and the dangers it poses").

[23]   See generally Claudia E Haupt, "Professional Speech" (2016) 125:6 Yale LJ 1238 [Haupt, "Professional Speech"].

A. *Professional Licensing and Discipline*

Professionals usually need a license before they may advise clients. Traditionally, professional licensing regimes rest on the states' police powers.[24] At its core, professional licensing is intended to ensure competence.[25] Occupational licensing has grown over time, but it is increasingly questioned on several grounds, including its effect on wealth distribution.[26] However, giving professional advice requires knowledge, and properly designed licensing regimes remain a useful tool to signal competency.[27] The fundamental idea of imposing a licensing requirement to ensure competency is fully compatible with introducing AI into the professional-client relationship. This is relevant to the introduction to AI into the professional-client relationship, because AI may take over functions that, performed by a human, would require certification or licensing.[28]

We might, of course, consider whether licensing, certification or some other form of accreditation are appropriate. Here too, the social relationship among human actors should guide policy. Professional licensing scholarship rightly notes that the extent of potential harm ought to determine what is appropriate. Richard Posner thus ties "the professional's capacity to harm society" to the belief

> that entry into it should be controlled by the government: that not only should the title of "physician," "lawyer," etcetera be reserved for people who satisfy the profession's own criteria for entry to the profession, but no one should be allowed to perform the services per-

---

[24] *Cf Slaughter-House Cases*, 83 US 36, 16 Wall 36 (1872) (discussing extent of states' police powers).

[25] See Nick Robinson, "The Multiple Justifications of Occupational Licensing" (2018) 93:4 Wash L Rev 1903.

[26] See e.g. David E Bernstein, "The Due Process Right to Pursue a Lawful Occupation: A Brighter Future Ahead?" (2016) 126 Yale LJ Forum 287; Clark Neily, "Beating Rubber-Stamps into Gavels: A Fresh Look at Occupational Freedom" (2016) 126 Yale LJ Forum 304; Morris M Kleiner, "Reforming Occupational Licensing Policies" (March 2015), online: *Brookings Institution: Hamilton Project* <www.brookings.edu> [perma.cc/SBZ6-LFUV]; Dick M Carpenter II et al, "License to Work: A National Study of Burdens from Occupational Licensing" (May 2012), online: *Institute for Justice* <ij.org> [perma.cc/9P46-JC37]. For a related discussion of internet speech, see e.g. Stephen A Meli, "Do You Have a License to Say That? Occupational Licensing and Internet Speech" (2014) 21:3 Geo Mason L Rev 753.

[27] See Claudia E Haupt, "Licensing Knowledge" (2019) 72:2 Vand L Rev 501 at 522–24 [Haupt, "Licensing Knowledge"].

[28] See Calo, *supra* note 7 at 417; Scherer, *supra* note 9 at 354 (noting that AI is now "performing tasks that, until quite recently, could only be performed by a human with specialized knowledge, expensive training, or a government-issued license").

formed by the members of the profession without a license from the government.[29]

On this point, even those generally critical of professional licensing seem to agree.[30]

With respect to regulatory policy, Ryan Calo suggests that "where AI performs a task that, when done by a human, requires evidence of specialized skill or training," a licensing or certification requirement of some sort might be considered.[31] However, "[i]n some contexts, society has seemed comfortable thus far dispensing with the formal requirement of certification when technology can be shown to be capable through supervised use."[32] This dispensation, however, may be only acceptable on a temporary basis. Where AI does not replace human professionals, but complements them and is supervised by them, these professionals would be initially licensed as such. What that means for their qualifications regarding supervision of AI, however, is a different question. A properly licensed but technology-illiterate professional, for example, will lack the qualification necessary to supervise the AI.[33] As long as AI is used to complement, rather than replace a human professional advice-giver, the licensing question remains secondary. However, as soon as the advice is rendered by AI without supervision, the issue becomes imminent.

Such unsupervised applications may occur where AI provides skills that humans in the same environment lack.[34] In this context, the system itself ought to be subject to advance licensing. A similar situation may arise anytime that professional advice is given outside of the human professional-client relationship which is grounded in fiduciary and other legal duties.[35] Calo contends that "in an environment rich in AI," it is an

---

[29]   Posner, *supra* note 12 at 2.

[30]   See e.g. Kevin Dayaratna, Paul J Larkin, Jr & John O'Shea, "Reforming American Medical Licensure" (2019) 42:1 Harv JL & Pub Pol'y 253 at 276. But see Shirley Svorny, "End State Licensing of Physicians" (7 August 2015), online: *Cato Institute* <www.cato.org> [perma.cc/5BVZ-QNJJ] (arguing against physician licensing).

[31]   *Supra* note 7 at 419.

[32]   *Ibid*.

[33]   In the context of the professional regulation of lawyers, the American Bar Association's Comment 8 to Model Rule 1.1 states: "To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, *including the benefits and risks associated with relevant technology*, engage in continuing study and education and comply with all continuing legal education requirements to which the lawyer is subject" (emphasis added): American Bar Association, *Model Rules of Professional Conduct* (Chicago: America Bar Association, 2018).

[34]   See Calo, *supra* note 7 at 419.

[35]   See *ibid*.

open question whether traditional approach of professional education followed by entrance exams such as boards or bars are useful.[36] For now, this means that the licensing of supervising professionals would suffice under current conditions in which AI is usually not used as a freestanding advice-giver.

Licensed professionals are also subject to professional discipline.[37] As long as there is human supervision, it seems that not much changes. But whereas discipline may prompt a human actor (under threat of sanctions) to adhere to the professional standard, it is unlikely that AI (particularly machine learning AI) will likewise modify its behavior. Disciplinary action might be contemplated in relation to the programmers, but the more independently the AI operates and the more that divergence from the professional standard is a function of machine learning, the less the idea of a post-licensing disciplinary system becomes responsive to how the AI functions. Thus, while a licensing regime is fundamentally responsive to AI, the system of professional discipline in its current form appears less suitable.

B. Professional Speech

Scholars debate whether traditional First Amendment theory and doctrine applies to AI.[38] Assessing the traditional justifications of First Amendment protection—democratic self-government, autonomy, and the marketplace of ideas—some of them contend that all support protection for "strong AI speakers."[39] The answer must to some degree depend on the social context in which the speech occurs. This makes professional speech applicable to the professional advice-giving context.

Professionals operate under a variety of legal constraints that do not apply to other speakers. Most importantly, "bad professional advice—that is, advice inconsistent with the range of knowledge accepted by the rele-

---

[36] *Ibid.*

[37] See e.g. Nadia N Sawicki, "Character, Competence, and the Principles of Medical Discipline" (2010) 13:2 J Health Care L & Pol'y 285.

[38] See e.g. Toni M Massaro & Helen Norton, "Siri-Ously? Free Speech Rights and Artificial Intelligence" (2016) 110:5 Nw UL Rev 1169; Toni M Massaro, Helen Norton & Margot E Kaminski, "SIRI-OUSLY 2.0: What Artificial Intelligence Reveals About the First Amendment" (2017) 101:6 Minn L Rev 2481; Stuart Minor Benjamin, "Algorithms and Speech" (2013) 161:6 U Pa L Rev 1445; Tim Wu, "Machine Speech" (2013) 161:6 U Pa L Rev 1495; Richard K L Collins & David M Skover, *Robotica: Speech Rights and Artificial Intelligence* (New York: Cambridge University Press, 2018).

[39] Massaro & Norton, *supra* note 38 at 1176. Note, however, that Massaro and Norton "refer to ... as-yet-hypothetical machines that *actually* think as 'strong AIs,' as opposed to 'weak AI' machines that 'act *as if* they were intelligent'" (*ibid* at 1176 n 7)".

vant knowledge community—is subject to malpractice liability, and the First Amendment provides no defense."[40] The doctrine of content neutrality, moreover, is incompatible with professional speech.[41] Finally, the doctrine of prior restraint does not prohibit professional licensing requirements.[42]

What does that mean for the professional speech of AI? In terms of First Amendment protection, the same framework should apply that governs the speech of human professionals within the professional-client relationship. The AI's speech must be accurate according to the standards of the respective professional knowledge community.[43] State regulation should not alter the content of what is otherwise considered accurate advice because of the non-human nature of the speaker. Moreover, "the speech-conduct distinction could conceivably provide a reason to deny First Amendment protection to much of what computers produce."[44] This is true for professional speech by humans too. Consider only the sometimes-blurry line between medical speech and the practice of medicine.

Returning to the question of harm, only speech that is accurate under the professional standard is protected by the First Amendment; bad professional advice, conversely, may be sanctioned by way of malpractice liability, and the First Amendment provides no defense against liability.[45] The focus here is on the harm that may result from bad advice. Like human speakers, non-human speakers may be capable of producing speech that results in harm.[46] Thus, the normative interests in avoiding harm to the listener by providing accurate and comprehensive advice are the same, no matter the identity of the speaker.

C. Professionals' Fiduciary Duties

The law imposes fiduciary duties on professionals to address the knowledge asymmetry between professional and client.[47] These fiduciary

---

[40] Claudia E Haupt, "Unprofessional Advice" (2017) 19:3 U Pa J Const L 671 at 675 [Haupt "Unprofessional Advice"].

[41] See Claudia E Haupt, "Professional Speech and the Content-Neutrality Trap" (2016) 127 Yale LJ Forum 150.

[42] See Haupt, "Licensing Knowledge", *supra* note 28 at 553–55.

[43] *Cf* Haupt, "Professional Speech", *supra* note 23.

[44] Massaro & Norton, *supra* note 38 at 1187.

[45] See Haupt, "Unprofessional Advice", *supra* note 40.

[46] See e.g. Woodrow Hartzog, "Unfair and Deceptive Robots" (2015) 74:4 Md L Rev 785 at 790–97.

[47] See Haupt, "Professional Speech", *supra* note 23 at 1271.

duties also reflect that professional relationships are social relationships based on trust.[48] Fiduciary duties consist of the duty of loyalty and the duty of care. Thus, fiduciaries "must take care to act competently and diligently so as not to harm the interests of the principal, beneficiary, or client."[49] Moreover, they "must keep their clients' interests in mind and act in their clients' interests."[50]

How do fiduciary duties apply when AI is introduced into the professional relationship? To account for the algorithmic role, Balkin develops the framework of "information fiduciaries."[51] He acknowledges that they are not the same as the classic fiduciaries, nor do they have the same range of obligations. But in the professional realm, they do have the same obligations because they are part of the professional-client relationship where those obligations apply.[52] Balkin invokes the lawyer-client and the doctor-patient relationship as examples of fiduciary relationships.[53] Endorsing Balkin's information fiduciary theory, Frank Pasquale notes that "software-driven devices are increasingly taking on roles once reserved to professionals with clear fiduciary duties."[54] Thus, Pasquale asserts, "[a] manufacturer of a medical device offering diagnoses should be held to the same standards we would impose on the physician it is replacing."[55]

Where services are of a professional nature, built on expertise, the resulting fiduciary duties are those of professionals. But insisting on fiduciary duties in this configuration actually says more about the concept of professionals than it says about information fiduciaries. In other words, I would suggest that the fiduciary duties imposed on professional AI are simply those of professionals, not those extended by analogy to professionals via the concept of information fiduciaries.

---

[48]   See e.g. Mark A Hall, "Law, Medicine, and Trust" (2002) 55:2 Stanford L Rev 463 [Hall, "Law, Medicine, and Trust"].

[49]   *Ibid* at 1207–08.

[50]   *Ibid* at 1208.

[51]   Balkin, "The Three Laws of Robotics", *supra* note 16 at 1227–31; Jack M Balkin, "Information Fiduciaries and the First Amendment" (2016) 49:4 UC Davis L Rev 1183 [Balkin, "Information Fiduciaries"]; Jack M Balkin & Jonathan Zittrain, "A Grand Bargain to Make Tech Companies Trustworthy" (3 October 2016), online: *The Atlantic* <www.theatlantic.com> [perma.cc/V5LT-A95K]; Jack M Balkin, "Information Fiduciaries in the Digital Age" (5 March 2014), online: *Balkinization* <balkin.blogspot.com> [perma.cc/9WSL-Q57M].

[52]   See Balkin, "The Three Laws of Robotics", *supra* note 16 at 1230–31.

[53]   See Balkin, "Information Fiduciaries", *supra* note 51 at 1205.

[54]   Frank Pasquale, "Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society" (2017) 78:5 Ohio St LJ 1243 at 1244.

[55]   *Ibid*.

D. Professional Malpractice Liability

In the human professional context, the tort regime imposes liability on those professionals who fall below the standard dictated by custom. This approach has itself been criticized as hampering innovation. Scholars argue "that courts' reliance on customs and conventional technologies as the benchmark for assigning tort liability chills innovation and distorts its path. This reliance taxes innovators and subsidizes users and replicators of conventional technologies."[56] Initially then, the professional who departs from custom increases their liability risk.[57] Moreover, in light of the liability framework, "[i]nstead of focusing upon genuine technological breakthroughs, innovators will strive to produce incremental improvements on customary and conventional technologies."[58]

The professional malpractice standard is determined by the practice of the profession. But what is the appropriate standard for AI? Whereas some scholars maintain that technologies such as driverless cars must be "safer than humans,"[59] it is not clear that this liability standard easily translates into the professional context. For example, what happens to the standard of care when AI becomes "better" at diagnosis than human doctors?[60] Again, this question is particularly salient with respect to machine learning AI. But raising the question is not to suggest that the tort system is incapable of addressing AI. In fact, the questions raised echo traditional torts questions arising when the tort system is confronted with new technologies.[61]

One policy problem the use of AI raises is "who bears responsibility for the choices of machines."[62] This question gains traction as "AI systems do more than process information and assist officials to make decisions of consequence. Many systems ... exert direct and physical control over ob-

---

[56] Gideon Parchomovsky & Alex Stein, "Torts and Innovation" (2008) 107:2 Mich L Rev 285 at 286.

[57] *Cf ibid* at 288.

[58] *Ibid* at 289.

[59] Calo, *supra* note 7 at 417.

[60] See generally A Michael Froomkin, Ian Kerr & Joelle Pineau, "When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning" (2019) 61:1 Ariz L Rev 33.

[61] See e.g. *The T J Hooper*, 60 F (2d) 737 at 739–40 (2d Cir 1932). See also Rebecca Crootof, "The Internet of Torts: Expanding Civil Liability Standards to Address Corporate Remote Interference" (2019) 69:3 Duke LJ 583 at 641–46 ("[t]he history of tort law is regularly punctuated with instances where new technologies alter social relations between entities, spurring legal evolution" at 642).

[62] Calo, *supra* note 7 at 416.

jects in the human environment."[63] The move from tool to agent to actor traces the evolution of liability.[64] Surgery robots, for example, for purposes of tort liability are treated as agents.[65] Once we move toward more fully autonomous AI, however, the question becomes whether the product liability regime might be appropriate. The formulation for design defects differs between editions of the *Restatement*. Whereas the *Restatement (Second) of Torts* contemplates the "consumer expectations test,"[66] the *Restatement (Third) of Torts: Product Liability* employs the "risk-utility test."[67] Both initially seem capable of capturing the changing liability landscape from tool to autonomous system.

E.   *Professional Ethics*

While there is some movement toward "a professional ethics of AI," scholars warn that historically, such developments of ethics codes have been susceptible to challenges as restraining trade.[68] Moreover, ethics enforcement without a "hard enforcement mechanism" tends to be difficult.[69] But when AI is adopted in the professional context, existing professional ethics frameworks—such as the ethics of self-regulated professions—are already in place. Moreover, professional ethics provisions traditionally are accompanied by more or less robust enforcement mechanisms. Thus, rather than focusing on AI ethics, the dominant framework to consider is provided by the ethics of the professions.

<div align="center">***</div>

When AI is used in the context of professional advice-giving, it is embedded in the regulatory framework governing human professional advice. To be sure, this framework may itself need adjustments. Considering the role of AI within the specific social relationship of professional-client interactions may usefully highlight areas for improvement. Moreover, the specific nature of AI itself may require modifications to the regulatory

---

63   *Ibid* at 417.

64   See David C Vladeck, "Machines Without Principals: Liability Rules and Artificial Intelligence" (2014) 89:1 Wash L Rev 117 at 120–21.

65   See *ibid* at 121, citing *O'Brien v Intuitive Surgical Inc*, 2011 WL 3040479 (ND Ill); *Mracek v Bryn Mawr Hosp*, 610 F Supp (2d) 401 (ED Pa 2009), aff'd 363 F Appx 925 (3d Cir 2010).

66   §402A at 351 (1965).

67   §2(b) (1998).

68   Calo, *supra* note 7 at 408.

69   *Ibid*.

framework. Ultimately, any assessment of the regulatory framework should be guided by the values it seeks to protect.[70]

II. Regulation and Innovation: "It's the Humans"

Legal changes in response to developments in AI "will occur contextually, as the ways in which humans actually use new technologies shape the legal doctrines designed to govern them."[71] This Part analyzes how the existing framework can be adapted to better accommodate the changes likely to be brought about by the use of AI in the professional-client relationship. I will focus on two questions in particular. First, how can the existing regulatory access points be used to regulate this social relationship? And, second, who is best situated to regulate professional AI?

In answering these questions, the guiding principle ought to be that rather than regulating AI, the focus should be on regulating the human relationship into which AI is introduced. Focusing the regulatory response to AI in this way preserves the normative basis of these human relationships. This approach corresponds to Balkin's larger theoretical premise, referenced at the outset, that AI and other "technologies mediate social relations between human beings and other human beings. Technology is embedded into—and often disguises—social relations."[72] Consequently, the proper focus ought not to be on the AI, but rather on the humans.[73]

A. Filling Regulatory Gaps

First, the discussion in the previous part has illustrated that the current framework of professional advice-giving, based on human-to-human interactions, is responsive to most, but not all uses of AI in this relationship. To the extent it does not perfectly map on to the use of AI, this section offers some suggestions to supplement or modify the existing framework. This, to reiterate, does not mean exclusive regulation of AI by the existing framework of professional advice-giving; rather, it means that the existing framework should be our starting point.

As the previous discussion has illustrated, there are some areas in which the regulatory system should be modified to be more responsive to changes introduced by AI. The first gap identified concerns professional disci-

---

[70]   See Haupt, "Professional Advice", *supra* note 10 at 66.

[71]   Massaro & Norton, *supra* note 38 at 1171.

[72]   Balkin, "The Three Laws of Robotics", *supra* note 16 at 1223.

[73]   See *supra* notes 17–20 and accompanying text.

pline. To the extent human professionals are expected to respond to the threat of professional discipline by conforming their professional behavior to the professional standard, a similar reaction will not likely follow with AI.

However, the question is what underlying normative interests are served. First and foremost, professional discipline should track competence. But disciplining AI may be beside the point. More relevant ought to be steps to ensure that professionals themselves are competent to use professional AI. In the context of legal advice, "[s]everal states have adopted regulatory measures to ensure that lawyers keep up with technology and understand the technology their firms use."[74] However, these rules are criticized as too vague in their application to the use of AI.[75] At the same time, current rules of professional responsibility require "independent professional judgment,"[76] which will likely be incompatible if AI is performing part of the function of professional advice. Indeed, "when a lawyer relies on AI technology, he adopts the transmitted results."[77] Yet some commentators suggest that such reliance may be in violation of the professional duty.[78] These examples illustrate that updating the professional obligations will be necessary to accommodate professional-use AI. However, the regulatory move will not be to impose professional obligations on the AI agent itself, but on its human professional user.[79]

A related question concerns the extent to which the professional must be competent to use AI, or what types of competence are relevant. A central critique here concerns the lack of explanation and the black-box character of results.[80]

---

[74] Katherine Medianik, "Artificially Intelligent Lawyers: Updating the Model Rules of Professional Conduct in Accordance with the New Technological Era" (2018) 39:4 Cardozo L Rev 1497 at 1515.

[75] See *ibid* (noting that "there are currently no standards in place about what it means to be a prudent or competent lawyer in relation to AI usage" at 1516).

[76] American Bar Association, *supra* note 33, r 2.1.

[77] Medianik, *supra* note 74 at 1518.

[78] See *ibid* ("[t]his willingness on the part of the lawyer to circumscribe his efforts and to compromise his thoroughness by offering clients legal advice attained from the blind reliance on technology is not in the best interests of the client and may be considered a violation of Model Rule 2.1 for failing to exercise independent professional judgment").

[79] For a proposal in the legal context, see Medianik, *supra* note 74, at 1524–29.

[80] See e.g. Frank Pasquale, *The Black Box Society: The Secret Algorithms that Control Money and Information* (Cambridge, Mass: Harvard University Press, 2015). In the medical context, see generally W Nicholson Price II, "Black-Box Medicine" (2015) 28:2 Harv J L & Tech 419 at 457–66; W Nicholson Price II, "Regulating Black-Box Medicine" (2017) 116:3 Mich L Rev 421 at 434–57.

B. The Locus of Regulatory Power

Moreover, there may be structural constraints to the usefulness of the existing framework. One potential limit of this approach is that it is largely state law that governs professionals, for example in licensing and in the tort law of professional malpractice. This raises several questions. First, while I favor a regulatory approach across different professions,[81] there might be a danger of fracturing the regulatory regime among states. But the existing framework governing professions has dealt with this issue of state-specific regulatory regimes in a variety of ways. It is not clear that accommodating AI poses challenges beyond state law's grasp. For example, in medical malpractice, we have seen a shift from the traditional locality rule that has given way over time to a national standard.[82] Similarly, with respect to licensing, we are seeing an increasingly national orientation of knowledge necessary to practice—consider, for instance, the role of the multistate bar exam—while maintaining state jurisdiction over licensing. With respect to state-based adjudication, moreover, California Supreme Court Justice Mariano-Florentino Cuéllar has noted that "AI is becoming an increasingly relevant development for the American system of incremental, common law adjudication."[83] As already noted earlier, adapting to new technological developments is not foreign to the common law: "Just as courts once had to translate common law concepts like chattel trespass to cyberspace, new legal disputes … will proliferate as reliance on AI becomes more common."[84]

Second, and related to the previous point, there is a concern that expertise needed to devise appropriate regulation is more readily available at the federal than the state level.[85] Several proposals for AI regulation beyond the professional advice-giving realm have addressed this problem by advocating for a federal agency solution.[86] Importantly, however, these discussions tend to focus on regulating AI development.[87] From this per-

---

[81] *Cf* Haupt, "Professional Speech", *supra* note 23 at 1246–47 (defending a professional speech approach across professions).

[82] See e.g. *Brune v Belinkoff*, 235 NE (2d) 793 at 798 (Mass Sup Jud Ct 1968).

[83] Mariano-Florentino Cuéllar, "A Common Law for the Age of Artificial Intelligence: Incremental Adjudication, Institutions, and Relational Non-Arbitrariness" (2019) 119:7 Colum L Rev 1773 at 1775.

[84] *Ibid* at 1776.

[85] See e.g. Frank Pasquale, "Data-Informed Duties in AI Development" (2019) 119:7 Colum L Rev 1917 (discussing the interplay of the common law and regulation).

[86] See e.g. Tutt, *supra* note 9 at 90; Scherer, *supra* note 9 at 393.

[87] See Scherer, *supra* note 9 at 377 ("examin[ing] the competencies of three separate institutions—national legislatures, administrative agencies, and the common law tort

spective, the higher the level of the regulator, the better.[88] Such an approach makes sense if the goal is to capture all of AI. However, a sector-specific approach better captures the social interactions that are already traditionally regulated at the respective level. In other words, when the regulation of human professionals is allocated to the states, the best anchor for non-human professionals will also be with the states. Indeed, to the extent that expertise is a concern, the focus thus far seems to have been expertise in the realm of technology and AI rather than expertise in the sector for which AI is employed.

Conclusion

Toni Massaro and Helen Norton posit that technology is "neither inherently good nor bad."[89] Therefore, "to declare it uniformly good or bad, useful or disruptive, presumptively protected from government regulation or presumptively subject to regulation, would be foolish. It should and will depend on context, and on what the new technology does to us and for us."[90] They also note that "new forms of communicative technology seem to have gained considerable dominion over us."[91] As consumers, "[w]e welcome their movie, restaurant, and book selections, not to mention their ability to guide airplanes and surgeons, keep us safer from domestic and foreign perils, help us avoid bad financial and health decisions, and foil sneaky consumer scams."[92]

In addressing the social relationships in which these technologies now interact with humans, however, we should not lose sight of the underlying normative interests. The specific case of the professional-client relationship—a special social relationship defined by the core features of knowledge and trust—illustrates the need for regulation of AI that protects the interests that make this social relationship distinctive.

---

system—particularly with respect to managing the public risks presented by AI" and suggesting that the focus on national legislatures is warranted "because of the diffuse and easily transportable nature of AI research. Because of these factors, most regional and local legislatures would be able to regulate only a small fraction of AI research.").

[88]   See *ibid* (arguing that "any substantive regulations adopted solely by a single sub-national political unit would not likely have a significant effect on the development and deployment of AI as a whole. Of course, national regulations suffer the same disadvantages when compared to international treaties.").

[89]   Massaro & Norton, *supra* note 38 at 1171.

[90]   *Ibid.*

[91]   *Ibid* at 1170.

[92]   *Ibid.*