Abstract

IBM’s Watson achieved celebrity on the quiz show Jeopardy! but it is now weighing in on matters of life and death as an expert system in the healthcare setting. Increasing automation in areas such as oncological research and diagnostics promises significant reductions in morbidity and mortality, but several barriers stand in the way of such technology’s widespread implementation and acceptance. As Watson achieves greater accuracy and is increasingly relied upon by doctors, one key question will become: how should liability be assigned where medical errors can be traced to an “artificial” intelligence (AI)? Using philosophical and legal concepts, this article classifies Watson’s current capabilities and examines how humans make “good” decisions in order to better characterize Watson’s role and potential faults. The authors conclude that Watson warrants a unique legal status akin to personhood and is analogous to a medical resident, in that both require “oversight” by an attending physician. We therefore argue that liability for wrongful diagnoses by medical AI should attach on a medical malpractice basis rather than through a products liability or vicarious liability scheme, and can thus be addressed contractually between the AI provider and medical institutions rather than constituting an open question for legislatures.

Keywords: Artificial Intelligence, Expert System, Liability, Medical Malpractice, Watson

Hey Watson – Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine

Jason Chung, Amanda Zink

Legal scholarship pertaining to artificial intelligence (AI), especially in light of its increasing sophistication and decision-making prowess, is still nascent. Courts have traditionally deemed it impossible for machines to have legal liability as they are not legal persons. Indeed, U.S. courts have previously outlined that “robots cannot be sued.” However, some modern commentators argue that a rethinking of legal systems is necessary

to deal with AI in light of its growing capabilities.

Controversy about how to treat AI from a legal standpoint reflects deeper disagreements about how humans should interact with such technologies, with debate about AI’s potential tending to gravitate towards the extremes. Decades of popular culture depictions reflect this dichotomy. Isaac Asimov envisions a world of peaceful co-existence with sentient robots, with the First Law of Robotics dictating that robots may not injure a human or allow one to come to harm. At the other end of the spectrum, we have dystopias such as Terminator’s Skynet or The Matrix, wherein sentient machines overtake humanity as the world’s dominant species.

AI has not yet achieved sentience and presently exists only in the form of useful tools rather than as a race capable of truly interacting with humans—let alone challenging us for global dominance. However, this does not mean that AI is incapable of interpretation or giving advice. AI employing cognitive learning is regularly used to parse through data and to offer options to human operators.

IBM Watson for Oncology (hereinafter “Watson”) represents the most compelling usage of such technology. Clinicians and analysts train Watson, a so-called cognitive computing system, to “interpret cancer patients’ clinical information and identify individualized, evidence-based treatment options.”² But does this indicate a machine’s ability to reason? And, if so, should machines be held liable for exercising poor judgment?

This paper highlights Watson’s duties and abilities to argue that existing legal regimes are sufficient to deal with AI’s (including Watson’s) “humanity” at its current stage of technological advancement, and to adequately attribute blame and liability for its errors.

To arrive at this conclusion, this article will conduct analyses of the following considerations: what are the various categorizations of extant and projected “AI,” and where do Watson’s current capabilities fall along this spectrum; what does philosophy have to say about how humans make “good” decisions, and what does this tell us about whether AI can make good decisions; and how do these analyses influence how Watson’s “mistakes” should be addressed under the law, including whether the technology should be classified under a human- or machine-centric theory of recovery.

Specifically, we posit that based upon Watson’s current AI capabilities as an expert

---

Hey Watson – Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine

system used to aid clinical diagnosis and decision-making, Watson can be analogized to a medical student, with an appropriate level of resultant legal responsibility for medical errors and medical malpractice coverage. By adopting an approach that assesses Watson’s true influence on treatment outcomes and its ensuing level of appropriate responsibility under the law, legal regimes should enjoy a flexible and workable framework within which to regulate existing AI while a broader debate about the rights and role of AI in human society takes place.

I. What is Artificial Intelligence Anyway?

Artificial intelligence is commonly understood as intelligence displayed by machines, the ability and development of machines to perform tasks that normally require human intelligence, or some such variation. AI applications are already commonplace in our daily activities, such as when we ask Siri if it’s raining, or request that Alexa play some bluegrass music. These types of AI applications, focused on the execution of a single task, are referred to alternatively as “weak” or “narrow” AI.

In contrast, many people associate AI with Artificial General Intelligence (AGI), or “strong” AI, which does not yet exist. This refers to the achievement of machine sentience or consciousness; AGI computers would possess or display intelligence equivalent to that of humans in every respect. Various tests have been developed that would purportedly detect the existence of true AGI, such as the Turing Test – in which a human would be unable to distinguish the responses of an AGI and a human to various posed questions. Ray Kurzweil, Google’s Director of Engineering and a noted futurist with a track record for accurate predictions (he claims an 86% success rate out of his 147 predictions since the 1990s), has consistently predicted 2029 as the year an AI will first pass a valid Turing Test. AGI would be capable of recursive self-improvement, leading to the rapid emergence of Artificial Superintelligence (ASI), the limits of which are unknown – but which many predict will result in the so-called “technological singularity,” perhaps entailing the obsolescence of human beings.

---

For now, we have expert systems, which have existed and evolved since the 1970s: computer programs that utilize AI methodologies to solve problems within a specialized subject area. Such domains may include anything from logic problems and games like chess to financial investing, legal research, and of course, medical decision-making. Expert systems that extend beyond binary logical inquiries (yes/no, true/false) rely on “fuzzy logic,” which must process linguistic terms in conditions of imprecise knowledge. These philosophical categorizations of AI inform the discussion herein.

II. What is Watson for Oncology?

IBM’s AI software or “expert system,” Watson (named after IBM’s first CEO, industrialist Thomas J. Watson, and not Sherlock Holmes’ sidekick, as is popularly misattributed), arrived on the popular radar when it beat two all-time champions on Jeopardy! in 2011. The following year, IBM announced Watson’s first practical collaboration with the Cleveland Clinic, in hopes that the system’s ability to synthesize huge amounts of data and produce evidence-based hypotheses could aid clinicians and students in more accurately diagnosing and treating patients.

IBM wanted to create a “supercharged Siri for business” and it essentially has. Watson now uses the same “Deep QA” software it used to achieve quiz show glory to tackle some of the healthcare system’s most complex problems—including properly diagnosing and treating various forms of cancer. Deep QA, in the simplest terms, is the software architecture that “analyzes, reasons about, and answers the content fed into Watson.” Watson’s abilities in the oncology space are ever expanding. In 2014, IBM announced that physicians could start using Watson to connect genomic and medical data to help drive more personalized treatments, and Business Insider has predicted that Watson will eventually allow oncologists to “upload the DNA fingerprint of a patient’s tumor, which indicates which genes are

---


mutated [and] sift through thousands of mutations and try to identify which is driving the tumor, and therefore what a drug must target.”

Expert systems—with Watson on the cutting edge—comprise “the most visible effects of research on artificial intelligence,” containing a knowledge base and inference tools that allow the user to both ask questions in natural language and receive a reply in the same language. Nonetheless, expert systems solve problems only in “a narrow and well-defined area.”

While Watson is therefore a far cry from AGI or the replacement of doctors with cyborgs, IBM frequently (though not always) shies away from specifically calling Watson “AI,” favoring alternate nomenclatures such as “augmented intelligence,” or as in recent TV commercials, “the platform for cognitive business.” Despite Watson representing one of the most sophisticated uses of narrow AI to date, this is perhaps unsurprising given the above-discussed confusion surrounding the definitional tiers of AI. Indeed, IBM’s fine distinction between “cognitive computing” and “artificial intelligence” appears more designed to placate those wary of the decline of human centrality and agency in the healthcare decision-making process rather than an accurate representation of the intensive research and data gathering and interpretation work carried out by Watson—work which is key to outlining the options for treatment of human subjects. Already, Watson is being used to score and rank medical literature and summarize patient records, key tasks in helping to contextualize and justify treatment options. Watson currently draws on more data than any reasonable human can be expected to consult—300 medical journals, 200 textbooks and nearly 15 million pages of text in order to present treatment alternatives, drug options and instructions for administration. Rising concordance rates suggest that Watson is on the cusp of being viewed as a viable, reliable option for vetting and applying such vast quantities of information. In June 2017,


7 Furmankiewicz et al., “E-Health,” 553.

8 Ibid.

data presented at a meeting of the American Society of Clinical Oncology suggested Watson was highly likely to reach the same treatment conclusions as human doctors. At India’s Manipal Comprehensive Cancer Center, there was agreement between Watson and doctors on the appropriate course of treatment for 96.4% of 112 cases of lung cancer. For other forms of cancer, concordance rates ranged from 81% to 92.7%. Anecdotal cases have revealed stunning successes as well. University of Tokyo doctors, for example, reported that Watson saved a 60-year old woman’s life by identifying her rare form of leukemia: “The analytical machine took just 10 minutes to compare the patient’s genetic changes with a database of 20 million cancer research papers, delivering an accurate diagnosis and leading to proper treatment that had proven elusive.”

Results have not all been this favorable—for instance, concordance occurred in a mere 49% of 185 gastric cancer cases in South Korea. However, this discrepancy likely occurred because results tuned for (and trained by) Watson’s doctors at Sloan Kettering did not translate to treatment approaches in Korea. However, with additional input from Korean doctors, Watson should be able to adapt to information fed to it from Korean doctors and provide options more in line with regional orthodoxy. This exemplifies one of the complex, “fuzzy” ways in which Watson can and must learn to effectively problem-solve in a given population.

Watson’s abilities also go beyond research and analysis. It is already being used by some clinicians to provide “a natural language interface for the delivery of general and patient specific information” in the form of giving patients information and gaining feedback. Hence, even such traditionally human duties as patient interviews and education are being taught to Watson—much as a medical resident learns and takes over from his or her attending doctor.

The distinction between Watson’s capabilities as a cognitive learning network, expert

---


system, or AI, etc. are more theoretical than practically significant, particularly from a legal standpoint. What is key is that Watson currently can investigate, learn, adapt and communicate with us. Until now, no other non-human being has been capable of as much. But does this mean AI can make intelligent and responsible decisions?

III. How do Humans Make “Good” Decisions?

Watson is an incredible achievement of neural networking and heuristic learning. In terms of sifting through information, it can parse through unfathomable amounts of data, rank its quality, and integrate new information into its adaptive programming. It can do so at a rate that no human can ever hope to match.

Yet, Watson is not sentient and, more crucially, cannot yet internalize or express moral values. This alienation of morality from the AI’s cognitive abilities lies at the heart of our difficulty in classifying Watson or attributing “human” intelligence to it. Traditionally, we have viewed morality as being central to the enjoyment of full rights and legal responsibilities of natural persons. Mankind has long attributed significant weight to emotional intelligence and intuition as necessary components of overall intelligence and the ability to make good decisions. This explains why doctors are judged not only on their raw analytical ability but also on their “bedside manner” and ability to weigh and explain difficult treatment options related to the welfare of the patient.

This may explain why AI in healthcare is viewed with some trepidation by the general public. Watson, for all its ingenuity, cannot yet engage in such abstract thought and many seem to feel this means AI should have a narrow role and application as part of human medical systems. A common suspicion is that AI’s inability to intuit situations may lead to overly cold and harsh decisions that devalue human life.

Indeed, most Western concepts of morality naturally assume the primacy of human experience and decision-making. As it relates to the very nature of consciousness, René Descartes coined the phrase “je pense, donc je suis” to usher in the era of epistemological study. Through a first-person view, Descartes advanced the notion that interpreting

information culled from human senses and either accepting or doubting them was the key to rational thought and existence.\textsuperscript{15} According to Descartes, the ability to apply methodological reasoning arose not as part of our corporeal existence but through an “essence” communicated through awareness and interpretation of the external world.\textsuperscript{16}

To Descartes, this was a uniquely human ability. The animals with which we share the world were described as the “animal machine” (”\textit{bête-machine}”), or automata without self-consciousness which could be exploited for human use.\textsuperscript{17} Similarly, regarding the possibility of “thinking machines,” Descartes argues further in his \textit{Discourse on Method} that:

\begin{quote}
…they could never use words, or put together other signs, as we do in order to declare our thought to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words which correspond to bodily actions causing a change in its organs… But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is said in its presence, as the dullest of men do.\textsuperscript{18}
\end{quote}

While Descartes was arguably the progenitor of attempting to outline a systematic method of rationality, even he could never conceive of a world in which machines could communicate with human beings and/or provide appropriate situational responses. After all, Descartes believed that thinking machines would never be able to “reason” through “all the contingencies of life.”\textsuperscript{19} According to Descartes, both the complexity of human language and decision-making were beyond mechanization.\textsuperscript{20}

Similarly, Immanuel Kant also believed that human beings were special and he believed that this gave humans an intrinsic worth and dignity. All other things, including animals


\textsuperscript{16} Ibid.


\textsuperscript{18} Descartes, \textit{Discourse}, 1993.

\textsuperscript{19} Ibid.

and machines, are therefore but “things”—a means to an end for humans. Kantian notions of morality are therefore reserved for humans as “rational agents” capable of making their own decisions, setting goals and applying reason to conduct. It is this capacity to reason that allows humans to make deontological moral decisions that respect the Categorical Imperative (CI). According to Kant, the CI is a rule that states “Act only according to that maxim by which you can at the same time will that it should become a universal law.”

Much like Descartes, Kant views humans as the only beings capable of comprehending, establishing and respecting CIs. The very concept of morality and responsible decision-making, according to Kant, is predicated on the conceit that the treatment of persons is an “end and never a means only.” In practical terms, that means difficult decisions can only arise from humans by humans, as no other entity would have the necessary intrinsic qualities to filter human experience and values. That may be why Kant viewed human beings as “above all price.”

From a broader societal perspective, utilitarian philosophers such as Jeremy Bentham, John Stuart Mill, and David Hume advanced the notion that raising overall human happiness should be the primary goal of human decision-making. Their consequences-based approach and devotion to the “common good” is again a human-centric one, and one that appears too nuanced for AI to understand. After all, even though AI excels at pure analytical calculations, how would it measure and compare intangible goods so that they can be weighed and pitted against one another?

These examples are not meant to be comprehensive, but to demonstrate that morality is viewed as the key to principled decision-making among humans. Judging from the diversity and breadth of philosophical texts on morality, balancing individualistic and societal goals is an exceedingly difficult process involving a metaphysical arithmetic that is highly unsettled and characteristically ill-suited to rigid categorization and valuation.

For a decision to be viewed as good and principled, however, it will have to be consistent with at least some vision of human morality. And this proves to be the greatest difficulty for most when thinking about whether we should accept AI as a component of healthcare


22 Ibid.

23 Ibid.
system—how can a machine devoid of intrinsic human sensations, perceptions, and intuition be trusted to make the “right” decisions? Even more crucially, how can AI be trusted to choose the correct option in cases where there are competing moral views and interests?

To be fair, this issue is not specific to “robots” and perplexes human decision makers as well. Even a cursory review of Western moral philosophy shows a wide disparity in what people believe makes humans “special” and how humans should interpret their universe and interact with others. This demonstrates the ambiguity of the supposed morality that underpins modern society. After all, even the weightiest human expression of values, our system of laws expressed through judicial systems, generally allows for some degree of variability in penalties and sentencing. Even when we have succeeded in establishing laws consistent with certain maxims (e.g. against murder), cases which offend these most basic maxims can involve lengthy debates about appropriate actions and mitigating circumstances.

For example, the interplay between reality and maxims is underscored by contemporary debates on euthanasia, where the social value of human life is weighed against the right of the individual to determine their own fate. There is no simple answer to this question, and to even begin to view this as a methodological or calculable matter among humans, let alone how to program such issues into AI language, engenders debate on the foundations of how to tackle such issues.

Further, it is unclear how humans can “teach” AI morality when we have failed to address glaring holes in our own extant theories of morality. For instance, with regard to utilitarianism, the challenge of dealing with imprecise measurables in calculating the common good remains a challenge—not only for AI but also for humans. As critiqued by political philosopher John Rawls, utilitarianism conflates “all systems of desires” into a singular conception of desirable social outcomes.24 Thus, the separateness of persons is sacrificed in favor of the utilitarian conceit that there is an impartial truth which maximizes human happiness.

This is problematic because ascertaining how to achieve the greatest net balance of satisfaction is not only difficult but can also obscure questions regarding the “source or quality” of such impetuses. After all, racism may make the majority happier overall but deprive a minority of their safety and liberty. Rawls notes that such a trade-off satisfies the

central tenets of utilitarianism but still represents a lack of justice.²⁵ Of course, this begs the question of what justice entails and who gets to define the concept. As one can imagine, this is a difficult conceptual conversation for humans, let alone for a programmer to code for the benefit of AI systems.

It is clear then that morality plays a key role in defining “appropriate” decisions among humans. As moral agents, we are not only called upon to make decisions but also pressured to be able to justify actions to both ourselves and others. Whether the influence is philosophical, religious, spiritual, or simply “intuitive,” human decision-making is assumed to be superior, paradoxically, because of our innate desire and ability to measure and compare incalculable moral considerations.

IV. Can AI Make “Good” Decisions?

Morality may lie at the center of human decision-making but is it a necessity for AI such as Watson, particularly as it relates to current usage in medicine? While the answer to that question may become complex as AI capabilities evolve, with respect to contemporary technological usage and implications, the answer is no.

There are many prerequisite steps before humans can apply moral principles to a decision, and not all decisions made in our day-to-day lives require the same level of complexity. For instance, the decision to drink water when thirsty is much more elemental than whether to prolong the lives of terminally ill patients in pain.

This reflects the fact that there is a hierarchy to thought and learning. The work of Benjamin Bloom and his colleagues in 1956 outlined learning theory that classified functional learning into three domains—knowledge (thinking) skills, psychomotor (physical) skills and affective (behavioral) skills—with graduated categories of learning from simplest (concrete) to most complex (abstract) in each domain.²⁶

Current AI technologies such as Watson exist to mimic human behavior in the knowledge (or cognitive) domain. For Watson to reach par with the decision-making capability of

²⁵ Ibid., 29-32.
adult humans, it stands to reason that it would be able to learn in a manner consistent with Bloom’s knowledge taxonomy. To learn on par with humans, AI would have to demonstrate achievement of each of the following steps\textsuperscript{27,28}:

(1) **Knowledge**: The ability to recall.
(2) **Comprehension**: The ability to understand and interpret a surface-level issue.
(3) **Application**: The ability to apply abstractions, general principles and methods to concrete problems.
(4) **Analysis**: The ability to understand the existence of and relationship between different aspects of a complex idea or problem and to separate hypothesis from fact.
(5) **Synthesis**: The ability to entertain and merge multiple ideas and concepts to form a new complex, integrated and meaningful idea.
(6) **Evaluation**: The ability to critically examine and judge ideas or methods using external information or methods.

Bloom’s taxonomy is a practical one and still widely used in the educational realm. Its appeal in this context derives from the fact that it separates out the ability to gauge mastery over increasingly complex tasks and gives us a clear way to judge AI capabilities in a functional and outcome-driven, rather than theoretical, manner.

According to IBM, Watson mimics human problem-solving ability by respecting the following steps\textsuperscript{29}:

(1) Observe
(2) Interpret
(3) Evaluate
(4) Decide

Watson does this by employing Deep QA software for the purposes of:

\textsuperscript{27} Ibid., 3.


(1) storing and updating a “corpus of knowledge” in the form of medical texts,
(2) curating the content (with human intervention) by culling through relevant medical
texts and discarding those which are not relevant,
(3) ingesting the content by creating indices and other metadata that make working with
the data more efficient (including creating knowledge graphs),
(4) learning from human experts who facilitate machine learning by uploading question/
answer pairs for the purposes of exposing Watson to linguistic patterns,
(5) continuing the learning process indefinitely with periodic review by human experts,
(6) identifying parts of speech in a question or inquiry,
(7) generating hypotheses,
(8) searching for evidence to support or refute the hypotheses,
(9) scoring each hypothesis based on statistical modeling for each piece of weighted
evidence, and
(10) performing evidence scoring and rating.

In applying Bloom’s taxonomy to Watson, we see that it excels in certain domains—to
the point of far exceeding human capacity—but lags in others. Indeed, it has been argued
that current AI machines “have very limited learning abilities” equivalent to only the third
level of Bloom’s knowledge taxonomy. 30

With respect to Watson, this argument appears correct. Regarding the “knowledge” level
of Bloom’s taxonomy, we have noted that Watson can store and recall information from a
staggering number of medical texts. With regard to “comprehension” and “application,”
Watson’s ability to understand inquiries as well as search through relevant data in order
to organize information necessary to present logical treatment options is, at this point,
indisputable given high concordance rates.

But in marketing Watson, IBM overhypes its capabilities. Unlike humans, Watson still
does not and likely cannot perform a key function—to decide on conclusions or courses of
action. While Watson excels at understanding human questions, the question and answer-
based model of Watson’s programming means that it cannot independently “analyze,”
“synthesize” or “evaluate” medical issues, or independently conceive of and search for new
and relevant information without human stewardship or guidance.

For true analysis, synthesis or evaluation to take place, there must be the ability to choose between a range of competing options. While one may argue that Watson can do that to some extent, in that answers to hypotheses can be scored and ranked, Watson still cannot do anything with this information. Watson’s role is as a passive assistant that can only interpret and answer external inputs.\textsuperscript{31} It does not have its own capacity or desire to formulate the question itself or to independently test and definitively prove or invalidate hypotheses.

Given that Watson’s learning model is not self-directed or motivated, asking it to consider even more abstract thoughts such as those relating to moral questions is also beyond reach. From a knowledge standpoint, it is nearly impossible to consider, rank and conceive of moral questions in a vacuum—that is, without the ability to go beyond existing data sets and properly contextualize information with other concepts that exist in the broader world. This may be why it has been exceedingly difficult to get AI to “feel” or to engage in “changes in interest, attitudes, and values, and the development of appreciations and adequate adjustment.”\textsuperscript{32} These are required for AI to gain affective or behavioral intelligence wherein value systems are ranked.

While human senses may be translated into AI language by converting external stimuli into electronic impulses, Watson’s lack of higher-level thinking required for critical thought and lack of context and diversity of interaction/sources for building an intrinsic value ranking scheme means that it fails at a very early level in the affective challenge of establishing and promulgating a consistent framework for morality. Of course, AI can be programmed to emulate human values by respecting fixed rules but this means that should Watson reach a paradox or an unanticipated situation, its response would likely be unpredictable and, from a human perspective, perhaps incomprehensible.\textsuperscript{33}

AI in its current state is therefore ill-suited to replacing human judgments as humans do more than simply apply the most logical solution to a given problem. Morality often dictates a variety of optimal solutions based on a multitude of personal sets of values and circumstances. For example, this explains why cancer patients are offered different treatment


\textsuperscript{32} Parker and Jaeger, “Learning,” 8.

\textsuperscript{33} Ibid., 11.
choices and conflicts frequently arise.

But this does not mean that Watson needs morality to make “good” decisions. Its ability to parse through voluminous texts and present logical options need not be encumbered by human questions of philosophy or ethics. In the practice of medicine, Watson’s current level of “intelligence” makes it a fine virtual assistant, soliciting the appropriate inputs (cases and patient’s medical condition) to provide a range of appropriate options. In that sense, it can make good decisions.

At its present stage, Watson exists in the health care space as an entity more akin to the trivia-slaying Jeopardy! champion than a physician asked to make and carry out life and death decisions. For the time being, Watson is—and should be—asked to play games, not God.

V. Towards A Functional View of AI

Given the existing limitations of AI and Watson, why do technology industry leaders like Elon Musk characterize AI as a threat to be regulated?34

There is an element that Musk is putting the cart before the horse. Musk fears that human obsolescence is imminent with technologies such as a neural lace being meshed with human brains to produce cyborgs.35 And indeed, injectable neural lace technology has been developed by researchers.36

Aside from the fact that this concept is similar to the one Kurzweil believes will result in humans who approach perfection, evidence suggests that Musk’s fears are theoretical rather than real. In the case of neural lace technology, its current envisioned application is to replace contemporary invasive neural electronics in monitoring brain activity. The ability to add on superhuman capabilities such as RAM-assisted memory through the neural lace is, for the

---


time being at least, a fanciful—if troubling—concept.  

But in another sense, Musk is right to worry. AI is being used to replace humans at an increasing number of junctures in the decision-making process in numerous industries. This means that there are fewer and fewer opportunities for humans to inject their judgment. 

Take Watson. It is already designed to comb through a “sea of knowledge” to find applicable treatments for specific patients and reduce the efforts of hours of work by teams of cancer doctors and oncology experts into a mere minutes-long search by AI. 

Overall, this is a positive development as it supports a better allocation of time by doctors. By reducing the amount of time spent poring through voluminous bleeding edge texts, human doctors will be able to focus on other priorities such as patient care and evaluation. Watson can also help alleviate the burden in areas where doctors are scarce—after all, the training of doctors is both lengthy and costly. 

But what happens when something goes wrong? Determining who may be to blame for medical errors is typically straightforward—a misdiagnosis would likely be the fault of the overseeing physician; a patient harmed by a medical device would be able to sue the manufacturer or operator. Given Watson’s role as a computer system that aids in the clinical diagnostic process, however, it is unclear how to assign liability for its faults. Currently, we have a hodgepodge of theories of recovery for injuries due to medical treatment—primarily medical malpractice, vicarious liability, and products liability— but Watson doesn’t fit neatly into any of these categories.

VI. Existing Liability Schemes Potentially Applicable to Varying Subsets of “Artificial Intelligence”

Far from a mere philosophical debate, whether or not we treat Watson as a person will

---


have real and lasting implications going forward. This is because theories of recovery occupy different categories of tort law. Broadly speaking, they can be divided into “human-centric” and “machine-centric” categories.

Regarding the human-centric categories, both negligence and vicarious liability as causes of action appear to require personhood. Negligence-based tort law, of which medical malpractice is a subset, applies where a breach of duty of care owed by one person to another occurs. Vicarious liability applies when “one individual can be held legally responsible for the acts of another.”

Medical malpractice applies where a physician is negligent in failing to meet the professional standards of medicine and, as a result, injures a patient who is entitled to recover damages. To succeed in a cause of action, a plaintiff must demonstrate (beyond a preponderance of the evidence) that (1) defendant had a duty of care to the plaintiff, (2) defendant failed to conform to the required standard of care either by his acts or failure to act, (3) the plaintiff sustained damages, and (4) the breach of the defendant’s duty was the proximate cause of those damages. The law of medical malpractice is thus intended to hold “physicians and other healthcare providers liable only for the subset of iatrogenic injury (adverse outcome of medical care) that results from professional negligence.”

Vicarious liability may also attach under the principles of agency, which hold an employer civilly liable “for injuries to the person or property of third persons occasioned by the tortious negligence of an employee which occurred within the scope of employment.” This is also referred to as the doctrine of “respondeat superior,” or “Let the master answer.” This imposition of no-fault liability is often justified as a type of “enterprise liability,” wherein the entity that profits from a subordinate’s activities is the more just risk-taker or loss-bearer. Such entities also typically have the “deeper pockets” from which an injured person can most properly seek compensation. Three elements are required for a valid agency relationship that may give rise to vicarious liability: (1) the parties must voluntarily consent to an arrangement in which the agent will act for the benefit, and under the direction

40 Ibid., 1063-1065.


and control, of the principal; (2) the agents must be acting on behalf of the principal, i.e., performing a task contemplated by the scope of employment; and (3) the principal must possess the general power to control and the right to direct the means and methods of the agent’s work.\textsuperscript{43}

On the other hand, machine-centric categories encompassed by \textit{products liability} attach to things and generally entail the strict liability of the manufacturer for defects. Under the umbrella of products liability, plaintiffs may sue for manufacturing defects, design defects, and failures to warn. Manufacturing defects may arise where a product “does not conform to the manufacturer’s specifications” and causes harm.\textsuperscript{44} Design defects may be found to have occurred where a product is in a “defective condition unreasonably dangerous to a user or consumer” or where “the foreseeable risks of harm could have been reduced by the adoption of a reasonable alternative design.”\textsuperscript{45} Finally, a failure to warn claim may arise based on a “manufacturer’s failure to provide adequate warnings to the consumer of dangers inherent in the product or to provide instructions for safe use of the product.”\textsuperscript{46}

\textbf{VII. Should Watson’s “Mistakes” Be Addressed under a Human- or Machine-Centric Theory of Recovery?}

How does one choose between these two camps? In order to do so, we must adopt a functional view of what Watson currently does and analogize its basket of duties as best as possible to either an equivalent human worker or a machine.

What, then, are Watson’s duties? As noted in earlier sections, they are to collect information from patients, analyze patient records, survey existing texts, and test hypotheses in order to make diagnostic and treatment recommendations. Its suggestions cannot be implemented without a treating physician accepting and acting upon them. Each of these

\textsuperscript{43} Kapp, “Legal Implications,” 295.


\textsuperscript{45} Ibid., 135.

duties has an equivalent as it relates to either a human or machine working at a medical institution – that of a medical student.

As noted by Brigham and Women’s Hospital, medical students perform each of these duties under strict supervision by a resident or attending physician. They perform interviews with patients to collect information about their histories, collect and track the progress of patients as well as their labs and test results, and review literature pertinent to the case to present differential diagnoses with leading and alternative hypotheses. Watson and medical students share another key similarity: they perform complex work as part of the direct patient care team but do not have the autonomy and decision-making authority of attending physicians or even residents.

When a patient claims to be injured as a result of the care of a medical student or resident, potential defendants in a suit for recovery include the “student, resident, clinical faculty members, medical school, and affiliated health care institution.” In the eyes of the law, medical students are essentially “no different from unlicensed laypersons [and] cannot diagnose, prescribe, or administer treatment except under the supervision and control of a licensed physician.” However, providing such supervision and control poses a great challenge for attending physicians, who frequently cannot be physically present or may write prescriptions or otherwise embark upon (or fail to take) courses of action based on erroneous information from the student. Nonetheless, the clinical faculty member may be held either directly or vicariously liable for the conduct of students under his or her purview. For example, a physician may be held directly liable for failing to properly consider the meritorious findings and interpretations of the student when acting as the patient’s attending physician. Conversely, the physician may be held directly liable for negligently delegating a task the physician knew or should have known the student was inadequately trained for or otherwise incapable of executing.

47 “Medical Student Responsibilities,” Brigham and Women’s Hospital, last modified March 10, 2016, http://www.brighamandwomens.org/Departments_and_Services/neurology/clerkship/responsibility.aspx.


50 Kapp, “Legal Implications,” 293.

51 Ibid., 293.
We therefore posit that Watson is best analogized to a medical student, and as such, can be held liable under the existing scheme of medical malpractice. Not only should Watson be able to be held liable in its student-like capacity, it is easy to see how a physician may be held directly or vicariously liable for failing to properly consider Watson’s recommendations, especially given Watson’s accuracy in providing diagnostic and treatment options as evidenced by high concordance rates with licensed physicians.

Of course, this is not a perfect analogy given the current capability gap between natural persons and machines. For instance, as natural persons, medical students can perform procedures that Watson still cannot. However, Watson offsets this weakness with unique abilities that natural persons cannot hope to match—such as the ability to review many more cases and parse through much more relevant literature than any human.

Though imperfect, classifying Watson as akin to a medical student is superior to other human alternatives. For instance, it has elsewhere been suggested that Watson is most similar to a “consulting physician,” whom current jurisprudence does not deem to have a duty to the patient since such a physician does not actually “interact with” the patient. In other words, “physicians who merely give advice, as opposed to orders or directions, without ever examining or talking to the patients, do not establish a physician-patient relationship giving rise to a cognizable medical malpractice claim.”

But Watson exceeds the influential capacity of a consulting physician because it does indeed “examine” patients through access to the patient’s medical history, phenotype, and, increasingly, genomic data. Furthermore, Watson is already being advertised directly to patients, with IBM’s website stating “If you are a patient interested in Watson Genomics from Quest Diagnostics [which utilizes genomics sequencing capabilities to help oncologists identify personalized treatment regimens for their patients], speak with your oncologist to determine if this test may be right for you.” This speaks to the fact that a consulting physician regime is inadequate to describe Watson. Machine or not, Watson most definitely is, and is marketed as, a member of the team.

Likewise, the products liability regime is not a good fit, as Watson is definitely not a


53 Ibid.

typical medical device. It is not a mere monitoring device surgically embedded in patients or directly performing any treatments, such as a pacemaker or cyber-surgical tool. Instead, its duties are higher-level as classified under Bloom’s taxonomy, and require interpretation and analysis.

Several additional factors weigh against addressing medical errors made by Watson under a products liability scheme. First, hospitals and other health care providers have traditionally been immune from products liability claims since their primary function is to provide services rather than to sell goods. Second, patients would not be able to sue Watson’s designer/manufacturer directly even if it did qualify as a medical device because of the learned intermediary doctrine, which essentially holds physicians responsible for assessing the risks and benefits of a device for a given patient. Third, software has not generally been found to fall within the scope of products liability, so any such actions would likely be restricted to blatant hardware malfunctions (such as where “Watson catches fire and burns someone” or “shuts down while monitoring and managing life support [something it does not yet do], resulting in the patient’s death”).

Relatedly, if products liability law did expand to cover software, this would likely raise difficult issues of proof—“[t]rying to distinguish between hardware and software failures for such a complex system could be a formidable challenge for any court to assess.”

For these reasons, Watson should be viewed under the law as a legal person with duties and responsibilities equivalent to those of a medical student in the event of “errors.” While Watson shares its platform with other machines, its functions are markedly human.

Operating under the medical malpractice scheme will allow greater transparency (as it pertains to potential legal liability) between IBM and the medical institutions that choose to utilize Watson. This scheme is also likely to be appropriate for many years to come, even as Watson grows ever “smarter” and more accurate. After all, even medical residents who are licensed to practice medicine by the state and who presumably have much greater expertise and skill than a medical student are still considered to be in a trainee capacity, working under supervisory control. Notably, the progressive capabilities of trainees must be considered by


56 Ibid.

supervisors in determining which tasks and degrees of direct scrutiny are appropriate, and by extension, which level of deference is warranted. This ought also to apply to Watson—it is easy to imagine a day when a doctor who can only correctly identify a case of lung cancer 89% of the time can be held directly liable for ignoring a recommendation from Watson, which (hypothetically) has a 99% accuracy rate.

Ideally, Watson will be able to reduce the current level of errors involving medical students by employing better “judgment” and streamlining the inherently collaborative process between trainees and attending physicians. A 2007 investigation revealed that the most prevalent factors in cases of malpractice claims involving trainees were errors in judgment (72%), teamwork breakdowns (70%), and lack of technical competence (58%). Diagnostic errors overall have been estimated to outnumber other types by 2- to 4-fold, representing almost 40% of all ambulatory malpractice claims. Of these, 75% are attributable to “cognitive factors,” which include: anchoring bias (tendency to adhere to an initial impression), framing bias or faulty context generation (over-reliance on the specific manner in which a question is posed), availability of search (jumping to a conclusion based upon a recent experience), satisfaction of search (failing to consider additional possibilities after a probable answer is identified), and premature closure (accepting an answer prior to its verification).

In theory, at least, Watson should excel at minimizing mistakes in these domains – it is designed to avoid making them - leading to less liability for trainees and attending physicians alike, as well as better outcomes for patients. A meta-analysis from researchers at Johns Hopkins found that up to 40,500 people annually die in intensive care units due to missed diagnoses, a “surprising and alarming” figure, according to the study’s lead author. Encouraging the use and expansion of Watson’s capabilities by outlining a clear scheme of accountability that limits liability to that of a learning but promising student appears prudent at this time.

Of course, there may be instances where there is a clear programmatic error by Watson’s


59 Singh, “Medical Errors.”

human creators. In that case, products liability claims may apply, but this represents a question of fact to be argued and determined before the courts.

VIII. Intent and Scope of Classifying Watson as the Equivalent of a Medical Student

Some may misconstrue our suggested classification of Watson as a medical student to mean we believe that Watson should be treated as a human. To be clear, this is not the case. What we are arguing is that Watson should be classified as a legal person for the purposes of apportioning liability so that Watson’s activities can be insured at a level that rises to that of a medical student.

This is important as the torts system is a crucial hallmark of the American legal process, allowing those directly and negatively affected to sue without intermediaries. Theoretically, this equalizes power between the people and corporations and keeps corporations accountable. Without clarification on the status of AI, cases promise to become longer and costlier for people as litigation focuses on who to pursue rather than the merits of the case. With increased usage of AI, these issues only promise to grow and the legislators and courts would be wise to adopt a consistent approach.

The establishment of Watson as a person equivalent in treatment to a medical student is a legal fiction that offers additional benefits. Indeed, we are aware that medical students are virtually never pursued for medical malpractice as a result of diagnostic error given the overriding responsibility of the attending physician. But the classification of Watson as a person offers courts additional leeway in determining Watson’s involvement in the decision-making process and promises to aid in the establishment of the truth when medical malpractice is alleged.

This is due to the additional information that would be available to claimants as part of the discovery process. As noted by New York attorney Eric Turkewitz, those named in a lawsuit have a higher evidentiary burden than non-party witnesses. This means that parties to a lawsuit must provide copies of statements and other evidence that may be otherwise
privileged during the process of conducting internal committee reviews. As such, if Watson is named as a defendant in a civil suit, those suing may be able to gain greater access to Watson’s logs to determine its exact involvement (or perhaps over-involvement) and history on the case.

Furthermore, as noted by injury lawyer Max Kennerly, adding students or residents to lawsuits helps with the attribution of responsibility for faults. He notes that many supervising physicians will deny legal responsibility for patient injuries as will the hospital itself. As such, a limbo often arises in which it is unclear who is even involved in the patient’s care, making the attribution of fault difficult. Allowing Watson to be a party to lawsuits should help clarify the degree of its involvement in patient care and establish who may have deviated from Watson’s recommended treatment options and for what reasons.

In short, analogizing Watson to a person would grant greater access to Watson’s determinations and enable them to understand Watson’s role in the alleged medical error. This also would aid triers of fact in making their determinations of fault. But keeping Watson’s degree of potential liability as limited as that of a medical student also ensures that hospitals will not be dis-incentivized from “employing” AI such as Watson as a helpful supplement to their medical teams by not subjecting it to the same duty of care owed by supervising or attending physicians.

Of course, legislators could choose to create an entirely new legal regime to deal with AI. Indeed, some experts have advocated such a wholesale change. But such an approach would be complex and create a raft of unpredictable new legal questions. For instance, legal experts might get bogged down on what constitutes the threshold between AI and a simple diagnostic machine. Questions such as whether AI has to have the ability to converse

---


63 See Allain, “From Jeopardy! to Jaundice.”

naturally or whether other forms of interaction are sufficient to demonstrate sufficient ‘intelligence’ threaten to dominate any new foray into creating a new legal liability regime specific to AI.

Similarly unrealistic is IBM’s position that because Watson will revolutionize healthcare, Congress should “get out of the way.”\textsuperscript{65} While Watson has yet to demonstrate such a dramatic impact on cancer care, a technology with the potential to affect the morbidity and mortality of millions via little-understood algorithmic decision-making processes certainly calls for some level of neutral, third-party oversight.

Meghan Dierks, director of clinical systems analysis at Beth Israel Deaconess Medical Center in Boston, is in the camp asserting that the implications of the new man/machine union to treat patients must be carefully considered. While IBM claims Watson is intended to “democratize” medicine and assist doctors with less specialized expertise in caring for cancer patients, “in advising those doctors, Watson cannot fully explain the rationale for its decisions. It can cite medical literature, but it cannot explain why it selected a particular treatment for a particular patient.”\textsuperscript{66} Dierks asserts that in such situations, where “dependency on the machine is higher, and the underlying rationale for the decision less clear, more oversight of the system might be warranted.”\textsuperscript{67}

While significant lobbying efforts by IBM have already helped wall off Watson from the types of FDA regulation applied to “higher risk technologies” thus far, Watson isn’t off the regulatory hook just yet. In the coming months the FDA will issue guidelines clarifying exactly which software products will be exempt under the new law, the 21\textsuperscript{st} Century Cures Act, invoking huge implications for Watson for Oncology and related products such as those that match patients to clinical trials and provide treatment recommendations based on genomic data.

The FDA’s commissioner, Scott Gottlieb, has expressed a desire to “streamline regulation of digital health technologies,” perhaps via a certification process that enables companies to “commercialize products without pre-market approval, or through a very limited review


\textsuperscript{66} Ibid.

\textsuperscript{67} Ibid.
process.” We believe our proposed liability scheme for Watson is in keeping with the dual goals of avoiding thwarting useful advances in patient care, and ensuring an appropriate “check” on a novel medical technology still in the nascent stages of implementation.

These comments from practitioners and regulators demonstrate that some level of oversight is desired and, as AI becomes more pervasive, the current legislative and regulatory vacuum will most certainly be addressed. Therefore, IBM and other AI manufacturers would be wise to embrace a regime that addresses what AI actually does and limits the application of standards of strict liability. The functional approach we have outlined above provides a practical fault-based regime that insulates manufacturers from the sometimes unpredictable consequences of self-learning machines. In this way, tortfeasors could be identified and classified by the capacity and the authority they have to commit (or share responsibility for) the tort, whether they be human or AI.

While it may seem a drastic step to grant legal personhood to AI, this would be perfectly in line with both reality and existing law. In the contemporary world, AI shares more common, if reduced, capabilities with humans than do other natural beings. For instance, wild animals are governed by the regime of strict liability because they are viewed as unable to be fully tamed. Their logic and impulses are alien to us and therefore their actions can often be too random and unpredictable to be compatible with our own. As such, they cannot enjoy personhood as they lack a relatable intelligence and organization.

By contrast, the ability and authority of AI can easily be characterized and analogized with humans. Much like students, Watson can, with periodic guidance, independently “learn” from texts and apply that knowledge to specific medical cases. It also communicates desirable treatment options in a manner intelligible to humans. As such, it achieves at least the third level of Bloom’s taxonomy and demonstrates quasi-human intelligence. The way we have programmed AI—in our language and image—means that we can already interface and assign tasks to computers in a way that often has a direct correlation with a human counterpart.

The legal concept of personhood—in contrast to the term’s colloquial usage—is also flexible enough to encompass AI and it would be easy to create a legal definition of...
personhood that fits both current and evolving levels of AI. Proceeding as we propose would have immediate legal and administrative benefits. Settling Watson’s legal status would grant all parties clarity on how to proceed in cases of error by Watson. The simple legal classification of Watson as a legal person would allow us to circumvent a key problem that we face now—the prospect of every case requiring a detailed existential analysis of what Watson represents and does.

Granting legal personhood does not mean that an uprising of intelligent AI machines is imminent. Legal personhood is merely a pre-existing legal fiction used to hold entities, not just natural persons, accountable. As noted by Michael Dorf, professor of law at Cornell, “Personhood is a legal status for which sentience is neither a necessary nor a sufficient condition. It’s not a necessary condition, because as a matter of law artificial entities like corporations can have personhood…” As such, there appears to be little harm in attributing legal personhood to Watson and much to gain. Watson’s existence doesn’t necessitate a rethink of our entire system of human laws—it just requires flexibility in what we consider a person.

**IX. The Difference Between Watson and Other AI (Such as Self-Driving Cars)**

The above does not necessarily mean that all forms of AI are the same. In fact, other forms of AI in which human decision-making is further removed from the equation present even more novel challenges than Watson.

For instance, the salient distinction between Watson and other common conceptions of AI such as self-driving cars is that self-driving cars are both making decisions (based upon programming by humans) and directly implementing them, thus capable of directly causing harm to individuals. While there may be instances in which Watson fails to live up to its own potential based on a flaw in “design” or specific programming (which may indeed give rise to a products liability claim and would be subject to factual determination), Watson still relies on a human intermediary to interact with (and thus, potentially harm, or fail to benefit)

---

the patient.

Hence, while AI may be used by both self-driving cars and in healthcare technology, the difficulties posed by AI usage in these different contexts vary wildly. As noted above, Watson’s detachment from administering treatment based on its diagnoses means that difficult moral and practical decisions involving patient care are left to human intermediaries, namely physicians. By contrast, self-driving cars will have to be programmed, and trusted, to make situational decisions that may involve life and death decisions.\(^{71}\) Ethical decisions, such as whether it is more acceptable to drive into the elderly rather than children, will have to be anticipated and programmed into self-driving cars, with AI entrusted to make such determinations in real time. In essence, the algorithm underpinning the self-driving car will dictate who deserves to live.\(^{72}\)

Clearly, the law will have to be flexible in order to accommodate the vastly different responsibilities and capacities of AI in much the same way we attribute differing standards to minors and adults. As technologies mature, laws to regulate AI must mature organically to fit the technology. While fleshing out the liability of AI distinct from or more advanced than Watson is beyond the scope of this paper, we offer some preliminary suggestions based upon the process by which we analyzed Watson’s appropriate liability. As guiding questions, we propose that lawmakers adopt the following four-part test before beginning to regulate and/or restrict AI:

(1) To what degree does the machine enjoy autonomy?
(2) To what degree does the machine interact with users/patients?
(3) To what degree does the machine provide reliable options?
(4) To what degree does the machine implement such options?

\section{Conclusion}

There is good reason to believe that we ought to incentivize the proliferation of


\(^{72}\) Ibid.
Watson in healthcare settings, particularly when it comes to expediting the diagnosis and management of cancer in Korea. Cancer is the leading cause of death in South Korea, and a severe shortage of South Korean healthcare workers is projected in the coming years. The Korea Institute for Health and Social Affairs estimates that the country will face a shortage of 110,000 nurses by 2020, rising to 158,000 by 2030.\textsuperscript{73} An additional 1,800 doctors and 7,000 pharmacists will also be needed by 2020, with requirements rising to 7,600 and 10,000 by 2030, respectively.\textsuperscript{74} The increasing automation of oncological research and precision medicine therefore has the potential to significantly reduce morbidity and mortality in the country.

Even given the position that we ought to encourage the adoption and development of AI, we must accept that there is a level of risk caused by unpredictability. Among other things, the potential of Watson for Oncology must be considered in the context of actual outcomes as more data becomes available in various settings and geographic locations. The University of Texas MD Anderson Cancer Center recently put its nearly 5-year endeavor with IBM to develop Watson Oncology Expert Advisor (OEA) on hold. While some characterize this as a loss of $62.1 million with “little to show for it,”\textsuperscript{75} this oversimplifies the reality. The audit by UT Texas that led to the project’s cessation highlighted several key issues: a change in focus from leukemia to lung cancer; an update to the MD Anderson Electronic Health Record (EHR) system on which Watson had initially been trained, leading to incompatibility, and; a failure by IBM to pilot the OEA at additional partner institutions (as was originally planned). However, IBM maintained that the project was a success and “likely could have been taken forward,”\textsuperscript{76} staff members noted that OEA’s lung cancer recommendation were in concordance with experts 90% of the time, and the audit report itself cautioned “Results stated herein should not be interpreted as an opinion on the scientific basis or functional

\begin{footnotesize}

\textsuperscript{74} Ibid.


\end{footnotesize}
The challenges raised by the MD Anderson project illustrate the reality that the integration of AI into human life, particularly in an industry as crucial and complex as healthcare, will present novel challenges. Success rates of individual endeavors and partnerships will vary, and may often depend on more than the capabilities of the technology itself.

What would be unfortunate is to let an overabundance of caution stymie efforts to harness the immense potential of AI. While there are certainly ethical dilemmas to address prior to rolling out certain forms of AI, such as self-driving cars, technologies like Watson provide a glimpse into how AI can supplement rather than replace human expertise. By combining the analytical and problem-solving capabilities of machines with the moral judgment of humans, we are on the verge of being able to tackle more medical cases and improve the lives of more people than ever before.

The law must be flexible to facilitate such a future. To invent sweeping new legal regimes that comprehensively spell out AI legal rights and responsibilities in the hopes of regulating all forms of AI would be a Sisyphean task given current levels of innovation. Instead, in arguing that Watson should be granted legal personhood, we are proposing a permissive regime that ensures accountability on the part of AI manufacturers and healthcare professionals through existing tort law frameworks.

Such an imprecise way of classifying and regulating AI may not be perfect, but that just makes it all the more human.

Received: November 24, 2017
Revised: February 13, 2018
Accepted: March 6, 2018