Spring 2019

Robodoc: Ethics of AI in Medicine

Halley Egnew
Western Washington University

RoboDoc: Exploring the Ethics of Medical Technology

Halley Egnew, WWU Honors Capstone Project, 2019

ABSTRACT

What do we do when the doctor of the future may not be human? In order to assess the full effect of trying to replace human caregivers with AI machines, we must investigate the types of ethics that these machines would work under—implicit, explicit, and full. The type of AI that movies present us with are fully ethical AI; they have a sense of self. The possible implementation of AI in medicine forces us to confront not just new technology, but also the definition of consciousness and free will, so I advise that for now we just stick to implicit and explicitly ethical agents in medicine.

My undergraduate experience has been very broad. I'm about to graduate with a BA in English Literature, but I also minored in Chemistry and am currently applying to medical schools with the hope of one day working in pediatrics. That means that I've spent time in the English department, but also in Chemistry and Biology, as well as Honors. When I began to think about the kind of Capstone project that would really encompass my undergraduate experience, the subject wasn't immediately straightforward. I've spent a lot of time with other pre-med and science students in my science classes, but I've also spent a fair amount of time considering value outside of the scientific, data-driven realm in my English and Honors department courses.

My perspective throughout my preparation for medical school has been that I want to be a physician who can treat my whole patient, not just their symptoms—someone able to listen to them with empathy and help them conceptualize the story of their illness and treatment. Science likes to categorize trues and falses, finding that there *is* a right answer, even if you can get to it via multiple pathways. English has taught me that meaning makes things right, and the types of stories that we tell ourselves can justify any answer, be it right or wrong or (most likely) in between. I'd like to take this meaning-driven attitude in to medicine.

As I began to brainstorm about a project that would match the range of my experiences, I also wanted to explore a question that would apply to my future career—something forward-thinking, that could help me continue to grapple with real-world medical issues and prepare me for situations I might face when I do become a physician. I wanted to look forward, and ponder the role of the "Doctor of the Future."

When the phrase "Doctor of the Future" first popped into my head, the image that came to mind was robotic. But according to research on medical outcomes[1,2,3], a huge part of medicine's efficacy (in increased longevity and patient satisfaction) comes from the human-to-

human connection and trust within the doctor-patient relationship. So far, this empathetically healing relationship is not possible with robots. Picturing our future physicians as robots (or as robot-reliant) changes the nature of caregiver-patient relationships that are possible. Considering that in my future medical career I will likely work in teams alongside robotic elements, it is important to me to consider the ethical implications of technologizing medicine. This lead me to my research question: *What ethical considerations do we need to explore if medicine becomes increasingly technologized?*

To answer this question, I will survey the existing literature on the morality of medical technology, use case studies from moral philosophy, and solidify the ethical questions around medical AI that I will continue to encounter and explore throughout my medical career. I've arranged this paper into three sections, of which I've broken down subsections. Firstly, I look at the defining concepts that go hand-in-hand with robotics and AI, investigating the different ways that robots and machines act ethically. Secondly, I give examples of how AI currently makes moral decisions, and how those decisions may change in the future. Finally, I think through the implications of those examples, through responsibility, autonomy, and equity. I conclude that, personally, no matter how advanced the robot-patient relationship becomes, I believe human-human connection still promotes a fuller type of healing that can't be technologized.

I. Conceptual Background

What Is AI?

Before I really start picking apart the functions and implications of robot and AI technology, I want to clarify the difference between the two. Many of my sources under the category of machine ethics address AI specifically, but the fact that the field is called "machine"

ethics and not "AI" ethics made me wonder: is machine a synonym for AI? Because the distinction between the two wasn't immediately clear to me, I want to make sure that I understand and use these terms with clarity—differentiating between the robots used in medical settings as tools and extensions of providers and the development of AI programming to make those tools more independent.

So, what exactly is the difference between a robot and an AI? According to Dr. John McCarty, Stanford professor and inventor of the term "artificial intelligence," AI is the "science and engineering of making intelligent machines,[4]" defining intelligence as "the computational part of the ability to achieve goals in the world." Note that this draws a difference between computation/intelligence and function. The "intelligent" part is the AI, while the "machine" part is the physical robot. McCarty's definition of AI requires that it be embodied within a robot. In other words, AI is the decision-making part of a robot, that can emulate human decision-making in the specific aspect that it is programmed for.

If the machine as a whole is like a human body, the AI would be its brain and the robot would be the body of the machine. The robot is the technology that allows the programming to carry out its function. That can be the humanoid robotic body that allows a full AI program to carry out its "own" decisions, or the simpler metal box that carries out the non-AI programming of a microwave. In the context of machine ethics, a *machine* refers to both the "body" and "brain" of the device—it accommodates the AI or programming that would run the program, and the physical robot that would carry it out.

But how different are AI from normally programmed machines? Because of who's programming AI, it's difficult to say that it would ever have the capacity to arrive at conclusions that a human wouldn't. McCarty notes that "If doing a task requires only mechanisms that are

well understood today, computer programs can give very impressive performances on these tasks," pointing out that to be able to tell a robot how to go through the decision making process, a human computer scientist must have an idea of how such decision-making might occur. And when it comes to metacognition, it's easiest for humans to focus on themselves, which leads to programming of machine decision-making in a humanlike way. So even if AI were able to arrive at a larger number of conclusions than traditionally programmed machines, those decisions would likely be constrained by the fact that a human programmed AI in the first place—the same limitations that govern traditionally programmed machines.

Currently, many robotic medical devices exist to assist providers. The da Vinci surgical robot, released in 2000, is one of the most well known. This device is an implicitly ethical machine, as it acts to miniaturize a surgeon's movements in order to complete minimally invasive surgeries.[5] The da Vinci robot doesn't offer any suggestions or guidance on how the surgeon should proceed, whereas AI potentially would. AI could analyze the patient's body in more ways than a human can, seeing beneath skin to accomplish more precise results. Because AI could advise or even control how a surgery proceeds, and therefore would be making decisions that affect human life, it must have some kind of ethical code to abide by. Currently, examples of technology like a dialysis machine or a heart pump have control over human life, but don't have explicit morality—because they are programmed to carry out a singular function and don't make decisions about changing that function. When AI becomes surgical, and the surgeon potentially relies on it for guidance in decision-making, it must have morality.

Three types of Ethical Agency

In order to begin discussing the ethical implications of AI in medicine, I first have to consider what kind of ethics apply to it, or what kind of moral agent it could be. While

researching ethics of AI, I found an essay collection that was a wonderful place to start thinking about what I soon learned was called "machine ethics:" Issue 74 of the series "Intelligent Systems, Control and Automation: Science and Engineering," entitled "Machine Medical Ethics.[6]" Before diving in, I was expecting the field of machine morality to be exact, having no experience with the subject, but as I soon learned, there is no great consensus about what kind of AI is the best option to aid physicians, and each category of ethical agency has its own limitations.

In his paper "The Nature, Importance, and Difficulty of Machine Ethics," James H. Moor distinguishes three ways through which a machine could be considered ethical. Reliance on humans characterizes *implicit* ethics, the simplest type of moral agency that we can give a machine or AI. An **implicit ethical** agent makes no ethical decisions on its own, or really any decisions at all. The machine's actions are "constrained [by the programmer] to avoid unethical outcomes," [7]and moral decisions are made by its programming and operator. The programmer can limit the machine's possible activities, and the operator can ensure that the machine's action matches the situation. The machine is told both what to do and how to do it. Machines like this are very common—Moor's own examples of implicit ethical agents include an automatic teller machine (ATM), which "give[s] out or transfer the correct amount of money every time," though it lacks "a line of code telling the computer to be honest"[7]. An ATM's reliability and honesty (that is, their ethics) is programmed in.

The second type of moral agent is a little more independent. We're still not talking about a fully autonomous machine, but an **explicit ethical agent** can make some decisions on its own, according to its programmed ethical codes. The machine can run the situation through the ethical code to determine what the outcome should be, and then figure out a way to get there. Like the

implicit ethical agent, it will still be told what to do by its programming, but it can figure out how to do it in the most ethical way. In comparison to an ATM, an explicit agent's ethics exist as defined decision-making rules within the device's programming. Programmers would take an ethical code, such as utilitarianism or consequentialism, and program it directly in to the device, which can then make decisions by applying ethical principles to the situation, arriving at a decision without human intervention. It has partial autonomy because it can't "decide" the outcome, but can decide the actions it will take to get to that outcome.

The third type of moral agent is a **full ethical agent**. Full ethical agents can both decide what the right outcome is, and what actions to take to get there. A full ethical agent has the ability to justify their judgments, as well as to reason and to learn,[7] explaining why they acted as they did and supporting it with logical reasons like a human would. In this way, a full ethical agent would have to be autonomous—deciding and evaluating the validity of its decisions independently of human guidance beyond programming. Some people don't believe that a machine can be a full ethical agent, because the abilities to reason, learn, and weigh decisions are all skills that are other considered to be uniquely human. For example, an adult human is a full ethical agent, because we learn ethics from others and combine them to develop our own systems of ethics that make sense to us as individuals, based on our life experiences. The possibility of machines as full ethical agents begins to toe the line with humanity; if a machine can make *and* evaluate their decisions autonomously, they would potentially be very useful as a human replacement in understaffed fields such as medicine. The distinction between "mere machine" and the pop cultural idea of "AI" mirrors the distinction between "implicit ethical agent" and "full ethical agent."

In the following pages, I don't expect to answer any large questions or come to a full conclusion. What I do hope is to explore the literature in the field of machine medical ethics, until I can clearly frame the questions created by the implementation of AI technology in the medical sphere. Should medical robots continue to be implicit ethical agents, or could AI development create medical robots to be explicit, or even full ethical agents?

II. Examples

Current Robots, Future AI

Though AI isn't actually used yet in surgery, just in diagnosis[8], more traditional, command following robotics along the lines of da Vinci have had a positive impact, such as when used to help guide screw placement in spine surgery. According to a 2018 review of studies comparing accuracy of screw placement when free-handed or robot-assisted found that robot-assisted placements outperformed freehand placements for accuracy, with less radiation used during surgery, but longer surgical duration than freehand placements[9]. Additionally, in a June 2018 study, remotely controlled robot-assisted retinal surgery [was] performed through a telemanipulation device[10]. Like the spinal screw insertions, the surgery took significantly longer than manual surgery (in this study, 4 min 55 s, vs. 1 min 20 s). Unlike the spinal screw surgeries, the retina operations did not have a significant difference in outcome results that favored either manual or robot-assisted surgery. However, I think it's fair to conclude from these examples that in surgeries where accuracy is more important than duration, increased robot assistance can improve outcomes. In the future, I would imagine that the diagnostics of AI could be combined with the robotic technology used in surgery in order to create a more streamlined surgical process. Increased integration of implicitly ethical machines like the ones described above would

increase accuracy, and development of fully ethical machines that encompass the machines already in use, but with higher decision-making capabilities could increase accuracy along with efficiency.

And perhaps, caregivers even welcome AI assistance, the same way that they have welcomed surgical robotics. Dr. Pardeep Kumar, a British physician interviewed by The Guardian said that with the use of robotic surgical assistance, he's "been able to carry out more operations, more quickly and successfully than I could have dreamed of."[11] Robotics also help him to see more patients, as surgical "strain on the neck, shoulders and back make it difficult" to keep performing surgery into his 50s, but robotics allow him to sit down during surgery—improving his stamina. Perhaps AI use could support the physician's decision-making in a similar way, allowing them to continue working for longer, even if their decision-making capability decreases along with their physical capability. This could help improve patient-physician relationship, as the patients could count on their physician to work a longer career, and stay with them through the course of their disease. A trusting caregiver/patient relationship results in more patient compliance and better health outcomes[12], so if caregivers are able to work longer, it's reasonable to assume that their patients will benefit.

AI also offers protections for physicians—in the case of Dr. Kumar, his physical limitations in surgery could have contributed to a sense of inefficacy and burnout, if it weren't for surgical technology enabling him to continue working with good outcomes. As Jennifer Bresnick points out[13], AI "doesn't need that fifteenth cup of coffee during the graveyard shift;" basically, it can work endless hours without suffering effects of fatigue—ensuring more equal quality of care for patients, no matter what time of day they happen to be admitted. Additionally, using AI to automate normal proceedings such as intake and taking vitals can give physicians

time to actually sit down with their patients for longer, benefitting patient care by helping doctors focus directly on the case. This also protects against burnout, because if caregivers had to do less paperwork and routine tasks, and had more time to connect with patients, they could take time to develop a relationship with the "person behind the symptoms,[14]" a factor that in a 2013 review paper "represented a crucial aspect of professional activity."

In the future, I imagine that AI assistance would make caregiver's jobs easier, allowing them to focus on human-human contact. AI could troubleshoot surgical plans, and make recommendations about treatment; or, constantly scan and compare the mid-surgery process to a data bank of previous surgeries, tracking that the procedure's proceedings. This could result in a kind of standardization of surgeries (more than they already are, of course), making procedures safer and decreasing patient suffering due to physician mistakes.

One practical dilemma of surgical AI implementation is the surgeon's autonomy. Surgeons have welcomed devices that enhance their physical abilities, but might balk at technology to support their brains. If AI were to evaluate the surgical plan, it might often agree with the surgeon's proposal, but sometimes it might disagree. In situations of disagreement, how do we define the line between AI recommendation and physician autonomy? The idea that we would defer to the machine's decision, and that human decision-making could be secondary to AI decision-making seems like a terrifying science-fiction; I imagine that though surgeons can (and based on the statistics—probably should) accept robotic assistance for certain procedures, they will want to protect their own autonomy in the operating room. Perhaps when AI and human surgeons disagree, the case could be referred to a surgical peer to tie-break. In order to begin using AI in surgery, we will have to define a hierarchy of decision-making.

"AUTOPILOT": SELF-DRIVING CARS

I mentioned implicitly ethical machines like dialysis machines and heart pumps that are charged with human life, but what about "automatic" programs, like autopilot? They are marketed as independent decision makers and in direct control of human life. I'll use the example of the AI of self-driving cars to further explore the implications of AI on autonomy. Self driving cars exemplify the relationship between AI guidance and human oversight—a good case study for medical robotics. Self-driving is also much farther along in public implementation than medical robotics, so real-world examples are readily available. Of course, driving isn't exactly the same as medicine, but the cars' AI system still substitutes for human decision-making, and is still responsible for human lives, causing philosophers to explore many of the same moral questions.

In searching for philosophy papers about self-driving, I found a lot of discussion of ethics that was similar to what I'd already read when researching medical machine ethics. This makes sense, because with literal life-and-death repercussions, driver AIs take on similar effects of risk to that of medical AIs. Professor Sven Nyholm proposes the need for "ethics settings" (I, 2), when it comes to self-drivers. He asks, should people be able to change their car's setting to prioritize their lives in a crash? I wonder, how much programming control do we need to worry about giving individual drivers—should we prioritize the greater good, and can we trust them to do it at the potential cost of their own life? Currently, cars with AI-driven technology have the ability to maintain speed, maintain follow distance, steer, park itself without a person in the car, and stay on route to guide the person to their destination. But what about when they crash?

Sven Nyholm highlights three notable crash scenarios[15]. In February 2016, one of Google's self-driving cars crashed with a city bus, for which Google took "partial

responsibility." In May of the same year, a Tesla in "autopilot" mode crashed into another car that was not detected by the autopilot's sensors. Tesla did not take any responsibility for the crash. Then, in March 2018, a pedestrian was tragically struck and killed in Tempe, Arizona by a self-driving Uber test car[14]. Just months ago on March 1, 2019, a Tesla car in autopilot crashed into an undetected truck in a situation "nearly identical to those of the first publicly reported, deadly Autopilot crash, in May 2016."[16] This marks the fourth documented death related with autopilot features, and second associated with Tesla. Clearly, the consequences of self-driving cars, even without their widespread implementation, are no longer just theoretical. And so far, the cars' manufacturers are split on who takes legal responsibility.

Currently, when it comes to moral responsibility, "automatic" functions of cars, like distance-monitoring cruise control and self-parking are really seen as extensions of human control. However, even Tesla itself seems split on the full responsibility of its automatic driving—on its "Autopilot" web page, a video of an autopilot drive filmed from inside the car is shown after a statement that "the person in the driver's seat is only there for legal reasons. He is not doing anything. The car is driving itself."[17] However, later down the page, Tesla says that "current Autopilot features require active driver supervision and do not make the vehicle autonomous." These two statements seem in tension with one another. Clearly, Tesla wants to take responsibility for the novelty and genius of the autopilot technology, without having to take responsibility for its failure. Contrast this attitude with Audi and Volvo's claims that when their self-driving technology is released, they would take full responsibility for crashes that might occur[24].This implies a high level of confidence on their part that crashes won't actually happen. But, as we've already seen with Tesla, Google, and Uber, they inevitably will. The interesting question here isn't *if* the crashes will occur, but *who* will take responsibility.

We can assume the same thing with medical AI—risks are inherent just from its undertaking, and mistakes will be made. Certainly, as Nyholm proposes the machines will need "ethics settings" of some sort. If the machines are explicit ethical agents, perhaps each hospital or clinic will have its own ethics prioritization, either by choosing one of a few pre-programmed options, or by hiring its own programmer to design an ethics program that fits their needs. I imagine that medically assistive AIs will be treated like implicit moral agents, even if they aren't, because people aren't used to dealing with explicit or full moral agents in technology. We really aren't sure yet if machines with AI can become full ethical agents—and because explicit ethical agents are both feasible and well defined, I think that it's more realistic to develop medical AI as explicit ethical agents, not to try and push the envelope of medical technology and AI consciousness at the same time. Explicit ethical agents won't be able to do as much, but they can be controlled. And in medicine, control is reliability, which is good outcomes.

The question with my proposal would be, however, who exactly decides what ethics are programmed in to medical robots? Do they prioritize the patient's life over everything? Do they prioritize "sicker" patients to work on first like some sort of triage? Do they let human caregivers override their programming, or do they insist on the most efficient way of treatment? How do the patient's wishes tie in with a possible loss of autonomy for the surgeon?

Robots with Empathy

One of the largest roadblocks so far in implementation of AI technology into medicine is the essential character of the empathetic doctor-patient relationship. Anna Paiva, in her survey "Empathy in Virtual Agents and Robots," mentions how the exponential increase in modern technology use has led to fears of "dehumanization of our modern way of living,"[18]especially

when it comes to areas that most rely on empathetic social connections—such as politics, or, she argues, medicine.

Why can't humans form empathetic relationships with robots? D.J. Gunkel explains that in traditional thought, "technology, no matter how sophisticated its design or operations, is considered to be nothing more than a tool or instrument of human endeavor"[19]. Consequently, robots "are not legitimate moral subjects that we need to care about"[18]. This fits technology and machinery in the traditional sense—after all, who ponders the responsibility that they have towards a stapler. However, I would counter Gunkel's point with the observation that as robots become AI directed and begin to have personalities, humans could have the ability to form more meaningful connections with them.

The Department of Computer Science at USC writes about a "Socially Assistive Robot" (SAR) that they're currently developing to work in physical rehabilitation. They describe empathy as essential to the healing relationship, and discuss their methods for its generation. "Though machines cannot *feel* empathy," they add, "they can *express* it" through different mirroring techniques (italics added). They highlight the importance of the robot's personality and ability to adapt to each user's needs; positing that "personality in a robot [is] an inherent component of the assistive context."[20] They also describe an effective SAR as having the ability to "adapt the robot's behavior in order to better model the user's personality and needs." In describing this idea of an adaptive, empathetic, robot with a personality, they're basically describing what I would consider as AI, using the definitions previously discussed in this paper.

Tapus and Mataric found that patients were significantly more pleased with robots that they found to have similar personalities to them, and that a robot can take input on voice pitch level, rang, and tempo, as well as heart rate, sweating, pupil dilation, and how close the patient

stands to the robot, to help determine the personality and social preferences of the patient. This allowed the robot to adapt their personality approach to match the patient (see graphic below), resulting in acts of expressed empathy by the robot that "encourages the patient to adhere to the treatment regime and helps to building doctor-patient trust." However, medicine is a two way street, and the AI working within medicine must be able to garner empathy from patients in order to build a truly trusting relationship, not just mimic it to them. Paiva posits that humans *can* and *do* extend empathy to machines, especially if the robot is humanoid[21], expresses empathy towards them,[22] and is physically there.[23] To me, this suggests that satisfying patient relationships AI are possible, and that they must be specially cultivated in medical robots if we wish them to take on a caregiver-like role in the medical field.

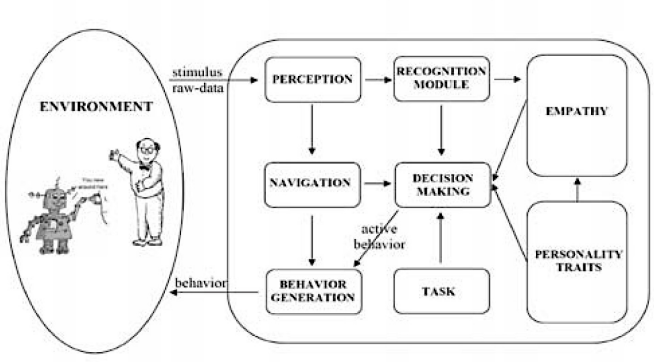

Figure 1: HRI Information Processing using the Personality Model of the User and the Empathy level

From Tapus and Mataric. HRI: Human Robot Interaction.

III. Implications

Machine "Autonomy" and Responsibility

The previous sections all point towards an underlying question of autonomy—with "auto" pilot, and the line between robot and surgeon autonomy. In addition, full ethical agents

require a level of autonomy, but can that even be achieved— can machines ever really be independent? What does machine autonomy really entail, and how does it compare to human autonomy? Can robots become actually as self-sufficient as true autonomy demands? Depending on how we answer these questions—we will have different outcomes. We don't know what kind of questions we will have to ask if machines are fully independent, because right now we assume that they are not.

The "bright line" argument is summarized by James Moor as the idea that consciousness exists as a specific divide between humans and AI; that it is uncertain whether AI will be able to achieve the consciousness necessary to make fully autonomous decisions (i.e. become a full ethical agent). For Moor, and many other scholars, true independence and autonomy are uniquely human characteristics that are not replicable. They believe that "no machine can have [the] consciousness, intentionality, and free will"[7] essential to being a full ethical agent. Therefore, they argue, no matter how complex, AI simply *mimics* the morality of real humans. This argument interests me, because it incorporates a larger question—where do we draw the line between human and machine?

Obviously, no one can be yet sure of the limits of artificial intelligence. Perhaps morality and self determination *is* the "bright line" which machines, including AI, will not be able to cross; the asymptote of artificial development. Although, even with this point of view, it seems certain that if AI is to work in hospitals, then it will need some explicit moral instructions other than that implied in its programming. Implicit moral agents are already at work in hospitals—in computer programs that protect patient privacy and allow communication, and the machines that ensure constant dosage of medicine. But, explicit moral agents would be able to handle a larger,

more autonomous role in patient care, bringing about the changes that we actually associate with artificial intelligence—automatized patient intake, patient assessment via robotic

So far, autonomous robots haven't changed our everyday life very much. We might even have a misconception of what autonomy means, because of its overuse with autopilot, autocorrect, and autofill. We know that the last two don't always work perfectly; they're just working within their programming. They're "automatic," but not "autonomous." The prefix "auto" comes from the Greek word for "self," and "matic" means "machine" whereas "nomous" comes from the Greek word for "govern." So, comparing that "automatic" means "self-machine" and "autonomous" means "self governed," the two words already hold the difference between traditional programming and AI: traditional machines are governed by their programming, whereas AI governs itself. A better synonym for "autonomous" as it really applies to technology is "self-directed." If technology became autonomous, it would have the independence to make its own decisions, and consequently, it wouldn't always listen to humans.

Amitai and Oren Etzioni discuss the often-misconstrued meaning of the word "autonomous" when it comes to artificial intelligence. Most technological lay-people imagine that "automatic" machines to decide their actions by their own volition. However, this isn't actually the case, and it's not clear if machines will ever fully become autonomous, especially when philosophers can't even agree if humans are. The Etzionis point out that "not every scholar is willing to take it for granted that even human beings act autonomously"[24], as even humans are suspect to influence from laws, social norms, and "sufficient antecedent conditions"[24].

The difference between AIs and humans, in this sense, is that humans are widely accepted as having free will to make their own decisions, whereas, so far at least, technology can't *decide* to go against its programming. With this in mind, we must change the way that we

think about autonomous machines, such as auto-pilot. The program's name includes "auto," but the program itself is not *actually* making any of its own decisions; it's just feeding input through human-programmed algorithms to generate a human-desired outcome. And that's great, if you know what you want the outcome to be. For planes and cars, we wish to arrive efficiently and safely to our destination. A truly "autonomous" device might not always follow the rules prescribed by its human operators—for example, choosing not to respond when asked to complete a task. If we assume that machines attain a similar consciousness to humans, we must be open to the possibility that some of them will not follow the rules.

For medical devices, reliability is key to saving human lives. And yet, the desired outcome of medical AI decision-making depends on the patients' wishes, and the stakes can be much higher than nonmedical situations. One of the primary moral tenets of medicine is the patient's freedom of choice about their care. Even if the patient chooses a less-aggressive treatment that could put their life at risk, it's their choice, and caregivers must respect it. Therefore, in everyday contexts, it could be detrimental to use autonomous robotic "caregivers;" they might not do what the patient tells them! For reliability's sake, it seems that robots in the medical field might be more easily managed if they don't have autonomy, and instead use their superhuman abilities to carry out research and advising.

Fully autonomous surgical AI is a long way off yet, and it seems much more likely, at least for now, that any surgical AI would be overseeing and monitoring surgery progress rather than actually doing the surgery themselves. In a situation where human lives are at risk, we may ask for AI's advice, but I believe that healthcare will be more straightforward if the main actors are humans.

Equity of Machines

It's already clear that in the future, AIs will become a larger part of the medical field—they solve problems of access, quality of care, and wait times. For example, the Aravind Eye Hospital in Madurai, India is already piloting a diagnostic AI that will help them more efficiently detect abnormalities in retina scans[8], increasing ease of care for their high-density patient population. However, AI implementation, especially if it moves towards replacing practitioners themselves, raises issues in terms of cost, loss of human connection, inequality, and responsibility. Will the technology be affordable? Will it be unequally available to those in wealthy countries? Will AI physicians be covered by the same malpractice that human physicians are?

If AI becomes effective and ubiquitous, I imagine two extreme and opposite limit situations: full-tech and no-tech. A hospital that runs majorly off of AI controlled robotics could be a super precise and accurate high-tech privilege for wealthy hospitals in first world countries. Or, it could be that as robotics become more popular, human touch becomes the sought-after commodity, whereas hospitals full of tech are relegated to poorer areas of the world that can't afford human caregivers. Considering each of these situations, how do we ensure that access to medical technology becomes fairer?

The Aravind Eye Hospital's implementation of diagnostic AI, adds an interesting data point to the two situations described above. Perhaps the Aravind Eye Hospital felt pressure to implement the AI diagnostic screening because of its patient load. Or, though this view is admittedly cynical, perhaps[25] Google felt more confident rolling out their diagnostic AI in India because India has less strict consumer protection laws. If other hospitals in high-density areas of the world like India and China begin to rely more heavily on AI because of the demand that their

patients create—are Western countries just using these countries as their own low-stakes testing ground? Does that narrative just perpetuate old stereotypes, and does AI implementation increase access to care and therefore serve equity? Do the patients have a right to an interaction and diagnosis by a human caregiver? Or do they simply have a right to care—no matter what form it takes?

Personally, I think it all depends on the quality of the AI. If we can be reliably sure about the AI's sensitivity, then it could help increase access to quality medical services. Moving forward, I think it's very important that we ensure responsible development of AI medical assistance. Likely, AI will be unequally distributed to areas that can afford the new technology and its upkeep, but consumer (i.e. patient) protection should be paramount in a hospital's decision to turn to AI. In order to prepare for this, I think that many hospitals should charge their ethics programs[26] with creating an addition to the hospital's mission that addresses their ethical standard for the adoption of AI into diagnostic and surgical programs.

Conclusions and Future Questions

When I began this paper, I had a curiosity. I suspected, but didn't realize just how multifaceted the topic of AI in medicine is. I think that part of my misconception was because a lot of the situations that I describe—robot caregivers, surgical AI—are a long way off yet. Right now, when people think of AI, they think of Sophia, and her wish to "destroy humans,"[27] or HAL 9000 telling Dave that she "can't do that."[28] People have fears that if robots become their doctors, they won't be listened to; they're afraid that their wishes might be steamrolled in light of the data-driven best outcome. So, do we let robots make decisions for us? I think, ideally, no. If robots are to be given roles in surgical and procedural medicine, then they should be under our

control, following precise orders. To me, a free-thinking robot (if it can even exist) shouldn't be given both mental and physical autonomy—it should be consulted as a second opinion, or a sounding board for procedures.

When they hear "AI," not many people picture a robot that listens to them. A robot that empathizes with them, or that wants a hug. But as I've discovered, developers like Adriana Tapus and Maja J Mataric at USC are working hard to ensure that robots and humans can have meaningful relationships. This seems like a good application for robots that can replace humans as in-home assistants or overnight vitals monitors. Still, I think that human-human relationships will feel more natural than robot-human ones, and that while machines can supplement our lives, they should not fundamentally change their key relationships.

I was surprised at a couple of my findings. I was surprised at how many times the concept of autonomy was discussed—I figured that as autonomous human beings, we would have a clear definition of what that meant already. But clearly, like the Etzionis note, there is a lot of talk, and no singular consensus on what autonomy really is.

After my research on the limits and intricacies of artificial intelligence, I find myself agreeing with James Moor. I believe that no machine can be a full ethical agent, because if a machine is given (or develops) a form of consciousness, to me, it crosses the "bright line." A conscious machine seems like an oxymoron. At that point, the machine could continue working normally, or it could stop working for humans, and start to make its own decisions. Like this infamous television physician, Dr. House, the machine might make decisions that ultimately could be most effective at keeping the patient alive, but at the cost of their autonomy.

I envision a more cooperative future, quite different from my "all-tech" and "no-tech" situations, in which physicians rely on AI to check their diagnoses or to cross check symptoms

and troubleshoot surgical plans. But, principally, it is the doctors themselves that end up making the final decision. Even with all of the work being done with SAR, I don't think that humans are ready to give up that personal interaction with their caregiver. But, talk to me again in 50 years.

Core Questions:

- *How much* responsibility should machines have?
- How do we ensure equitable access to technology? Should we prioritize *quality* or *efficiency* of patient care?
- How does the introduction of robot caregivers change the patient relationship? Does it free physicians to be more involved in care and less in bureaucracy? Will it therefore decrease physician burnout?
- In the future, will the mark of a good hospital be the preponderance of medical robots and AI, the conservation of human physicians, or a blend of the two?

WORKS CITED

[1] Kaplan, et al. "Assessing the Effects of Physician-Patient Interactions on the Outcomes of Chronic Disease" Medical Care, Vol. 27, No. 3, Supplement: Advances in Health Status Assessment: Conference Proceedings (Mar., 1989), pp. S110-S127

[2] Hojat, et al. "Physicians' Empathy and Clinical Outcomes for Diabetic Patients" Academic Medicine, Vol. 86, No. 3 / pp. 359–364. doi: 10.1097/ACM.0b013e3182086fe1

[3] Kelley, John M et al. "The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials." *PloS one* vol. 9,4 e94207. 9 Apr. 2014, doi:10.1371/journal.pone.0094207

[4] McCarthy, John. "What Is AI? / Basic Questions." Professor John McCarthy: Father of AI, Stanford University, jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html.

[5] "How the Da Vinci Si Works." *Robotic Surgery Center*, NYU Langone Health, med.nyu.edu/robotic-surgery/physicians/what-robotic-surgery/how-da-vinci-si-works.

[6] Tzafestas, S G, et al., editors. Machine Medical Ethics. Vol. 74, Springer, 2016.

[7] Moor, James. "The Nature, Importance, and Difficulty of Machine Ethics." IEEE Intelligent Systems. 21:18-21. (2006). 10.1109/MIS.2006.80.

[8] Metz, Cade. "India Fights Diabetic Blindness With Help From A.I." The New York Times, The New York Times, 10 Mar. 2019, www.nytimes.com/2019/03/10/technology/artificial-intelligence-eye-hospital-india.html.

[9] Gao, S., Lv, Z. & Fang, H. "Robot-assisted and conventional freehand pedicle screw placement: a systematic review and meta-analysis of randomized controlled trials" Eur Spine J (2018) 27: 921. https://doi.org/10.1007/s00586-017-5333-y

[10] Edwards, TL et al. "First-in-human study of the safety and viability of intraocular robotic surgery" *Nature Biomedical Engineering* 2, 649–656 (2018). https://doi.org/10.1038/s41551-018-0248-4

[11] Piesing, Mark "Medical robotics: Would you trust a robot with a scalpel?" The Guardian. 10 Oct 2014.

[12] Stewart, M A. "Effective physician-patient communication and health outcomes: a review." *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* vol. 152,9 (1995): 1423-33.

[13] Bresnick, Jennifer. "Arguing the Pros and Cons of Artificial Intelligence in Healthcare" Health IT Analytics. xtelligent Healthcare Media. 17 Sept 2018.

[14]Zwack, Julika, and Jochen Schweitzer. "If Every Fifth Physician Is Affected by Burnout, What About the Other Four? Resilience Strategies of Experienced Physicians." Academic Medicine, vol. 88, no. 3, Mar. 2013, pp. 382–389., doi:10.1097/acm.0b013e318281696b.

[15] Nyholm, S. "The ethics of crashes with self‐driving cars: A roadmap, I." *Philosophy Compass*. 2018; 13:e12507. https://doi.org/10.1111/phc3.12507

[16] Davies, Alex. “Tesla's Latest Autopilot Death Looks Just Like a Prior Crash.” Wired, Conde Nast, 16 May 2019, www.wired.com/story/teslas-latest-autopilot-death-looks-like-prior-crash/.

[17] “Autopilot.” *Tesla, Inc*, www.tesla.com/autopilot.

[18] Paiva et al, “Empathy in Virtual Agents and Robots: A Survey” ACM Trans. Interact. Intell. Syst. 7:3, (September 2017), 40 pages. https://doi.org/10.1145/2912150

[19] Gunkel, David. “The Rights of Machines: Caring for Robotic Care-Givers.” Machine Medical Ethics, vol. 74, Springer, 2016, pp. 151–166.

[20] Tapus A, Mataric MJ “Socially Assistive Robots: The Link between Personality, Empathy, Physiological Signals, and Task Performance”. Association for the Advancement of Artificial Intelligence. 2007.

[21] Pavia et al, 11:19

[22] Pavia et al, 11:20

[23] Pavia et al, 11:21-22

[24] Etzioni, Amitai and Oren. “Incorporating Ethics into Artificial Intelligence." Journal of Ethics. 21 (4), (2017): 403–418.

[25] Shrivastava, SaurabhR & Shrivastava, PrateekS & Ramasamy, Jegadeesh. Scope of consumer protection act in medical profession in India. Journal of Clinical Sciences. 11:25. (2014). 10.4103/1595-9587.137250.

[26] Pearlman, Robert A. “Ethics Committees, Programs and Consultation.” *Ethics Committees and Ethics Consultation: Ethical Topic in Medicine*, University of Washington, depts.washington.edu/bioethx/topics/ethics.html.

[27] Weller, Chris. “The First 'Robot Citizen' in the World Once Said She Wants to 'Destroy Humans'.” *Inc.com*, Mansueto Ventures, 26 Oct. 2017, www.inc.com/business-insider/sophia-humanoid-first-robot-citizen-of-the-world-saudi-arabia-2017.html.

[28] “2001: A Space Odyssey (Film).” *Wikiquote*, Wikimedia Foundation, Inc., 5 June 2019, en.wikiquote.org/wiki/2001:_A_Space_Odyssey_(film).