Are We Ready for Artificial Ethics: A.I. and the Future of Ethical Decision Making

Shannon E. French

Kiju Lee

Margaret Kibben

Susannah Rose

Are We Ready for Artificial Ethics
AI and the Future of Ethical Decision Making

Shannon E. French
Inamori Professor in Ethics and Director,
Inamori International Center for Ethics and Excellence,
Case Western Reserve University,
Kiju Lee
Nord Distinguished Assistant Professor in Mechanical and
Aerospace Engineering,
Case Western Reserve University,
Margaret Grun Kibben
Rear Admiral (retired), US Navy
Former US Navy Chief of Chaplains
Doctor of Ministry, Princeton Theological Seminary
and
Susannah Rose
Associate Chief Experience Officer and Director of Research,
Office of Patient Experience,
Cleveland Clinic

SHANNON FRENCH: Ladies and gentlemen, it is my pleasure to welcome you to this evening's event. My name is Shannon French. I'm the director of the Inamori International Center for Ethics and Excellence here at Case Western Reserve University. We are delighted to have some distinguished panelists with us here this evening to discuss topics that concern us all as we look to the future and how the technology that we live with may possibly come to define us more or less than we want.

Before we begin, I'd like to tell you a bit about each of our panelists. Then, I will be directing some questions to them myself, the moderator's prerogative. Then, we will open the floor and any of you who have questions can chime in. We have folks that can run over with a microphone so we can hear you all. We are filming the event behind you there. When you stand up, if you are comfortable doing so, you're welcome to state your name and your affiliation. If you're not, that's also fine.

24

To tell you about our panelists, I'm going to start furthest from me here with my Case Western Reserve colleague. This is Kiju Lee. Professor Kiju Lee is the Nord Distinguished Assistant Professor in Mechanical and Aerospace Engineering here at Case Western Reserve University. Her research area is in distributed intelligence and robotics, including sensor network systems, swarm robots, novel robotic mechanism design, and human–robot interaction.

She directs the Distributed Intelligence and Robotics Laboratory, which facilitates research and hands-on educational activities for both our graduate and our undergraduate students. Her research has been funded by NSF, National Academies Keck Futures Initiative, Clinical and Translational Science Collaborative in Cleveland, and VA APT Center in Cleveland. Professor Lee earned her MS and PhD in mechanical engineering from Johns Hopkins University in Baltimore, Maryland, and her BSE in electrical and electronics engineering in Seoul, Korea. Thank you for joining us.

Our next speaker is Dr. Susannah Rose. Susannah Rose joined the professional staff at the Cleveland Clinic in 2011. She is currently the associate chief experience officer and director of research in the Office of Patient Experience. She's an assistant professor at the Cleveland Clinic's Lerner College of Medicine and again, here at Case Western in the Department of Bioethics. I'll also note that she is a member of the Inamori Center's advisory board. She is also an instructor at Harvard University's Chan School of Public Health, where she teaches health ethics.

She received her PhD from Harvard University's Health Policy Program with a concentration in ethics in 2010. Prior to her doctoral studies, she earned an MS in bioethics from Union College Albany Medical Center in 2006 and an MS in social work from Columbia University in 1998. Dr. Rose worked as a clinical social worker and researcher at Memorial Sloan Kettering Cancer Center in New York City. Her publications have appeared in highly ranked peer reviewed journals such as *JAMA*, *Internal Medicine*, *Journal of General Internal Medicine*, *The Journal of Clinical Oncology*, *The New England Journal of Medicine, PLOS ONE*, and her book chapters on health policy ethics and end of life care have been published by Oxford University Press.

Her current research focuses on the ethics of AI and the ethics of transparency; patient experience and testing; the impact of innovations in technology, communication, and end of life care on health outcomes and patient satisfaction. These areas of research have been generously funded by multiple sources, including Harvard University's Safra Center for Ethics,

NIH's Clinical and Translational Science Collaborative at the Cleveland Clinic, Case Western Reserve University, and the Greenwall Foundation. Thank you for joining us.

Last but certainly not the least, I will introduce to you Admiral Kibben. Admiral Margaret Kibben, in her final active duty assignment in the navy, was the US Navy's twenty-sixth chief chaplain, serving as the director of religious ministry for the Department of the Navy and advising the Secretary of the Navy, the chief of naval operations and the commandant of the marine corps, as well as the commandant of the coast guard on all matters pertaining to religion in their respective services.

Prior to that assignment, Rear Admiral Kibben served as the eighteenth chaplain of the marine corps. A native of Washington, Pennsylvania, Margaret Kibben earned a bachelor of arts degree from Goucher College in Towson, Maryland; her master of divinity and doctor of ministry degrees from Princeton Theological Seminary, Princeton, New Jersey; as well as a master's degree in national security and strategic studies from the Naval War College in Newport, Rhode Island.

Dr. Kibben served as a senior fellow at the United States Institute of Peace in Washington, DC. Dr. Kibben has over twenty years of senior executive experience leading and managing a 2,500 member global workforce serving as a strategic level adviser to the Department of Defense, ensuring constitutional protections for religious diversity and providing guidance to leadership regarding the spiritual health of the force and the religious impact on military operations and diplomatic relations.

Her background reflects a proven history of visionary expertise in ethical leader development, religious accommodation, and spiritual advisement. Presently, Dr. Kibben serves on the board of trustees for the Princeton Theological Seminary, our board of advisers for the Inamori Center, and the Marine Corps Scholarship Foundation board. She is a member of the Veterans and Military Advisory Committee for the Washington National Academy. Please join me in welcoming Margaret.

Now that you know who our speakers are, you can see why they were chosen to address these topics here tonight. I hope as we open our conversation that you will let your own questions percolate in your mind so that when I open it to the floor, you will be able to jump in with the questions that you have. We're going to start off in the middle here with Susannah, if I may. There's a lot of enthusiasm for incorporating AI in the fields of medicine and health, however, we know there have been some public missteps as well.

For example, there was IBM's Watson for Oncology, which was rejected by doctors for giving bad or even dangerous treatment recommendations—an issue, incidentally, that IBM blamed on the doctors that trained the system. As a bioethicist, where do you see positive promise for integrating AI into the health and medical area and what raises red flags for you?

SUSANNAH ROSE: Thank you, Shannon. Thank you for having me here. I absolutely love coming that big one mile down the street. It seems like a huge journey, but yet it's so close. What I think is fascinating about the IBM Watson Oncology example—and I'm going to just briefly discuss what that is so we are all on the same page—the IBM Watson for Oncology was really a very highly publicized use of artificial intelligence, or AI, in the healthcare domain.

Specifically, it was a partnership between Memorial Sloan Kettering Cancer Center in New York City, where, you may remember, I spent about a decade. It was housed essentially to use the expertise of the oncologists and the surgeons at Memorial Sloan Kettering, which is arguably one of the largest and best cancer hospitals in the United States. They used that knowledge to then train an artificial intelligence using machine learning to integrate what they knew about patients. They put in patient data. They also put in hypothetical data, which I'll get to in a minute. They also put in research findings and research data into this machine to try to determine what would be the best treatment for particular patients who have a certain disease makeup.

They were putting in a lot of variables which theoretically a human mind or a human doctor might not be able to balance in order to determine the best care for that patient. One of the concerns that was raised was that some of the treatment recommendations seemed maybe a little bit to the left–center and at some times may have even been dangerous. For example, suggesting a blood thinner for a patient with a blood disorder might have been dangerous.

From what I understand, there were not reported outcomes of harm, unless somebody else knows about this. It was more of a theoretical concern about the results that were coming. I think what's also important about this whole scenario is that there was very little research done before this was actually deployed and actually bought by many other healthcare systems. There wasn't a systematic structured research that was done on the AI incorporation in the healthcare system.

Perhaps some of these challenges might have been identified before the product was marketed. I think the final thing that is fascinating about this is also the use of data. I think we'll get into that a little bit later too, so I won't spend too much time on this. Some people criticize the training of this

because the people who were doing the training—I'm sure that my colleagues will verify this too, is that when you're training in a model—you have to have labeled data and you have to be able to determine the outcomes and then train the mechanisms to be able to assess based on those labeled data.

If you are making those determinations by a very small group of clinicians, even if they're the best, theoretically in the world, that might be a biased source of training of these data. For example, other countries, other scientists and other physicians actually said that they disagree with some of the recommendations that may have been made by some members of this great group of physicians.

You also might wonder about the data that were actually used. Sloan Kettering is a very specific hospital system and the type of data and the type of patients that go there might not be representative of not only the United States but certainly not of the world. We think about how these data are being used, are they representative and also how are the data being trained? I think these are all important issues for healthcare.

Finally, even though I sound a little pessimistic about the lessons that IBM Watson for Oncology gives us, I have gotten very interested in the use of AI and how it might benefit healthcare. We can talk about that a little bit later. I do see some really interesting benefits. First is the area of diagnostic, certain assistance, or what I like to call augmented intelligence. It's the combined use of an artificial intelligence and humans such as radiologists that work together to produce even better results than each of the systems individually.

I think what's also fascinating, not only in the diagnostic area, is looking at healthcare outcomes for large populations of people. In the United States, we see transitions in payer systems in health policy that are not only paid on how individuals fare but also on how groups fare, what we call population health. AI might be very helpful in combining many thousands of data sets and also thousands of variables to determine where we might go.

Lastly, allocation of healthcare systems. Even within healthcare systems and across healthcare systems, AI might be quite valuable in knowing which patients might be best served in certain environments more than perhaps an individual person or a group of people combining those data on their own. That's where I see the three areas where AI and healthcare might, at least initially, be most helpful.

FRENCH: Thank you. I think we will have to return both to the data question but also this language of augmented intelligence instead of just

artificial intelligence and what that can do to assist human beings and possibly shore up some of their natural weaknesses. That's intriguing. This one, I'm going to direct to Kiju, if I may. You have developed and you continue to design some amazing robots, including some that were inspired by origami shapes. Where do you see the most promise for positive ethical use of this kind of innovative technology? Do you worry about future uses of your technology that might not align with your personal ethics, such as for weaponry? As an engineer, are there things that you can do within the R & D process to make it less likely that your creations could be used one day for harm rather than good?

KIJU LEE: Sure. Thank you for the great questions. Yeah. I see a great positive impact that AI has brought to our society. First of all, it has been used for various applications, including assistive technologies that can help people with disabilities. I am talking about prosthetic devices that can help people walk or use their arms again or hands again. There are surgical robots that are now pretty widely used that are proven to reduce operation time and thus reduce the time of hospitalization and time for recovery.

That has tremendous benefits as well. Also, in particular, AI is becoming more important because of the personalization of technologies. In the old days, we were just automating certain tasks that people didn't want to do or were too dangerous to be done by humans. Now, it's becoming a part of our lives. You want to make the AI adapt to you rather than you adapting to the technologies. Because in the old days, you had to learn how to use computers, you had to learn the programing languages to do it. Now, my five-year-old daughter talks to Siri and plays the music that she wants.

The way that different generations interact with technology changes through the advancement of AI. I see a lot of positive impact, in particular for those things that we can actually change in a positive way. In particular, when you think about the aging population or people with disabilities, technology can do the work that people may not be able to provide. I see a great positive impact in our society.

FRENCH: That's wonderful. I'm sitting here or rather standing here with a bum ankle and possibly staring down surgery and so I was listening to that part going, "Good. This sounds good." I can also relate to the idea of the younger generations being digital natives and being so comfortable with all of this technology. I remember when my daughter was barely a toddler, when she wanted something bigger, she always pinched her fingers

outward. She thought she could make it go bigger, because it should. The technology adapts to us.

I am going to just press you on one point. You didn't speak about the concern about weaponry, but I know we've had some offline brief conversations. I like to worry people about that. Do you have any thoughts about whether your work could ever be taken that direction?

LEE: That's true. One of my research topics is in swarm robotics. I'm going to start a new research project probably within a month or so with DARPA Funding. We're going to actually work on these swarm algorithms that can work with military personnel for urban military missions on destroyed infrastructure with limited wireless communication and those things and can make sure those military personnel stay safe and then perform search and rescue type of purposes and also identify potential hazards.

AI plays a great role and robots are a special form of AI that has physical autonomy with software that can perform intelligent tasks together. In particular, swarm robotics has a great potential for those applications, but it depends on who uses it. It could be really good. It could be really bad. Whether I can embed any research and development process to implement certain features to guarantee the safety or ethical use, I cannot answer that. I do not have any way to incorporate that one.

That is up to our society to address as our policy makers and all those people who have to bring these topics up and discuss further. As an engineer, I do not have an answer yet.

FRENCH: That is a fair answer though. I think what that also brings up is the role of those of us who do work specifically in the field of ethics, the responsibility that we have to ask these questions and to get a seat at the table to ask these questions. I was also saying on the break before we started that I have been asked some of these questions about DARPA projects at various times because they always have an element called ELSI, which is the Ethical, Legal, and Social Implications of the work that they're doing. The opportunity to weigh in in that kind of ELSI side of the research is something that I and many others who work in these fields take very seriously to be that voice where you sometimes can't be . . .

LEE: Sure.

FRENCH: . . . able to take that role. Let's move over to Margaret now. This is something that I've also been following very closely, very aware of, given my interest in military ethics. There has been quite a push happening lately

for the DoD to design and deploy advanced AI in the military. In fact, some people have even talked about an AI arms race.

This could be amongst countries like our own, the US, and China and Russia for example. This causes concern to many. Any time you talk about an arms race, it brings up some bad memories of course. There's been much discussion about both the fact that AI itself may not simply be advanced enough let alone infallible as some people imagine. The issue of the importance of keeping human beings firmly in charge of life and death decisions. As a former military chaplain, what hopes or worries do you have about the future use of AI by the military?

MARGARET KIBBEN: Dr. French, thank you very much for the opportunity to speak here. Given the fact that this particular session was scheduled for two days, where there were weather challenges, I'm wondering if AI is fighting its own battle with Mother Nature. When you talk about the design and execution, the deployment of AI in any setting, as most of you know, it boils down to two things. Probably, you could extrapolate to many, but I believe it's just two. That's control and value.

Addressing these factors is only complicated when you start factoring in things like social norms, the intention of the user—the country in this case—with widely differing moral frameworks. What is the end game? What's the intention that people are setting out in the utilization of artificial intelligence? It's critical in the case for things like China and Russia to understand that they come with different agendas than perhaps we have. Those agendas have to be considered.

With respect to values, it boils down to this question: Whose values? That comes at an individual level as well as the global scale. We in the United States have measured our battlefield assessment, things like our crisis action, decision making, our overarching engagement strategy, all of those things are based on the principles of just war. That's the moral framework in which we have built all of our equipment to include our artificial intelligence.

Assuming we maintain that ethic in our own design and utilization of artificial intelligence in combat operations, can we count on our adversaries to hold the same value set that we do? On what moral framework does that country base its decisions? When you talk about things like rules of engagement, what we understand to have contributed to the development of those rules, the establishment of those rules: Has everybody else bought into the same thing?

It's at the very basic nature of any engagement from one country to the next but it's only extrapolated that much further when you start factoring in things like artificial intelligence. As for artificial intelligence specifically, as you all have indicated and Dr. French has put in a number of her articles, the GIGO (garbage in, garbage out) principle still applies. The garbage being, in this case, the moral code that is built into an intelligence system. Human fact–fallibility is extrapolated at the speed of Moore's Law.

When you're putting that into artificial intelligence, you don't know what you're going to get as you talk about who's placed in that construction. There's a second question that I think we also need to ask. That is, can morality be coded? Can you, in fact, program morality? There are some factors which I believe are clearly defined and you probably very well could put them into some sort of algorithm. I would argue that the thing that makes humans human is the soul. The intangible factor that allows humans to remain aware not just to the nuances of a situation, what's happening, the dynamism of that which is taking place around them, but also the moral implications of an action.

Things that enable us to preserve the virtues of love and hope and kindness. That's, I think, where we're going with respect to Dr. Lee's question: How do you know that when I intended it to go this way that when it's laid out there that unintended consequences aren't going to start reaping hazard into what we had really hoped for good? Regarding the second half of the question, the issue of control, in other words: What about our adversaries?

The issue of control is at the center of the concern. As a chaplain, pastor, I often wonder if humans really are in charge of life and death decisions. Theological debate aside—I realize this tends to be more of a science conversation—when we give up our responsibility to keep the safety on either the literal or the figurative trigger of any kind of weaponry, then we cede our agency. That is our responsibility to preserve those same values. There you have it. Those are my thoughts.

FRENCH: These are great thoughts. I love that question: Can morality be coded? That is a wonderful question. Of course, I always like it as a military ethicist when people invoke just war tradition. This makes me happy. I take your point as well about garbage in, garbage out and how do we see that in this context and how do we ensure also what might be going into these systems from our enemies, which is a problem we know has gone on since there has been conflict.

All of that is very worrisome, while it is also intriguing. I have a question now that I'm actually going to frame to all of you and then we'll let you

duke it out. No. You can then answer—all of you, one at a time. I think this is because it applies to all of your fields. It applies to everything that you work on. There's been a lot of research to suggest that human beings tend to defer to automated systems even when those systems are known to create errors and to be capable of great mistakes.

People using them have the training to know better, and yet they don't trust themselves over the machine. This is sometimes called "automation bias." In your various fields, do you see signs of automation bias and how concerned are you about it? Do you think this phenomenon, if it exists, might start to fade as people become more used to automated systems and no longer see them as mysterious black boxes that probably know more than we do?

I think for example of, in my own case, a very simple rudimentary system, my GPS. I count on it when I am going somewhere unfamiliar; I am wedded to every turn that it tells me. If I'm somewhere that I know well, I will laugh off my GPS and go, "That's a stupid way to go, I'm going to go this way." My own familiarity changes my approach. We have seen these very worrisome studies that suggest that people who even have experience will still defer to that automated voice, if you will. Which of you would like to tackle that first?

KIBBEN: I will. I like the Google Maps image because there is a dependency that we create on this idea. I watched a woman today, I am not kidding you, as we were driving in with the taxi. She was driving and made a left turn on a one way street and everybody was coming the opposite direction. The taxi driver says, "Doesn't everybody have GPS now?" Even if they do, my question is: Do they use it? Maybe they depend on it too much. That being said, I think that it's human nature.

This is where you're getting my perspective from my background, it's human nature. Aside from finding the chance for healthy competition, we are quick to defer to others who appear to be better, smarter, more important than ourselves. The reasons for that range anywhere from lack of self-esteem to laziness. This is part of the human condition. It's a natural inclination. It also applies in our response to automated systems as well.

What I think we missed in this acquiescence, as well as in any unhealthy competition, is that we lose sight of our individual talents, our individual skills, those which I would call our gifts which are unique to each person, but which when included and incorporated one with the other, contribute to the overall well-being of humanity. This is where I think Dr. Rose was going with her understanding of the augmentation, the individual augmentation.

You used one phrase, I have another phrase in my notes here I'm going to refer to. You talked about augmented intelligence, it's also intelligence augmentation. There are other theorists who are talking about this. If we look at intelligence as augmenting who we are uniquely as human beings, then I think we capture the whole picture. We don't lose sight of the fact that individually, we have very specific skills, very specific gifts.

ROSE: Yes. I'm going to agree with you and disagree with you. I'm not so sure about human nature, but I don't think I want to comment on that. I do disagree with you about the background of why this automation bias exists. I don't think it's low self-esteem and I don't think it's laziness, particularly in a lot of areas in healthcare. In healthcare, we see a lot of Epic and Allscripts and a whole bunch of other what we call electronic medical records that many of our clinicians use. Because of a lot of the data that are coming at most of our clinicians, we have developed what we call clinical decision support tools, which might suggest a certain prescription or something for a certain patient based on information that's provided on their medical record and on known treatments for that particular disease.

What we see though, with quite a few studies, is that clinicians will do exactly as you said. I'm not so sure it's due to laziness or self-esteem, but because the recommendation will be made they will often follow it. There have been a lot of studies in trying to deviate from this default system that humans tend to play into. We set alerts to go off, we have a red sign, a whole bunch of things we've been trying to accomplish as well as reducing workload. Because if people are in a hurry, they also tend to be more likely to make these errors and go with what the systems suggests.

What I suggest is that we do deal with this in healthcare. We don't have a different human nature than the rest of the world. I'm not convinced though that I've seen yet how we can entirely get beyond this. What we might want to look at though is the counterfactual of life without these support tools. If you're looking at a purely outcomes utilitarian analysis, in general, having these decision support tools does lead to better outcomes and recommendations compared to those who do not. Even that we do have, and this is a rough estimation, there are differences with different types of recommendation tools. We don't know how to completely overcome this yet.

FRENCH: One thing that I have explored in the military context, but I don't know if it would work as well in the medical context, is to focus on systems that simply don't give a single answer. In other words, they only

ever give at least two choices. Do you think there is some possibility there that you could avoid that automation bias if you still force the individual to make a decision?

ROSE: That's absolutely possible in a sense because I think that that forces somebody to make some kind of decision. The problem is that we sometimes, given the data that go in, there might be the possibility of several different decisions in certain scenarios. In healthcare especially if you're plugging in a lot of data, there might be some higher level certainly.

With these systems, we not only develop a probability or essentially a coefficient, we often have a confidence interval, right, that's also playing into this. We not only need to develop our systems to say, "Okay. This person is a seventy-five percent chance and a seventy-four percent chance in this." Those might be reasonable things to choose against. What is meaningful difference between the two, this is where we start to get into some really interesting gray areas. Because if you give a choice between two different options and one is ten percent probability better with the confidence interval, you see where I'm heading?

FRENCH: Yeah, absolutely.

ROSE: To give almost a false choice between two things that might not actually be so equivalent also has some ethical problems.

FRENCH: Absolutely intriguing. Kiju, would you like to jump in here?

LEE: Yeah, I actually wanted to give a very specific example of automation bias and then how the training can work. I teach a class called Computers and Mechanical Engineering. That course basically teaches that you have a mathematical model of physical systems or real world problems modeled into mathematical format, and our goal is to basically teach students how to use computers to solve it. By doing so, at the beginning, when you use Excel or a calculator, you barely doubt your solution, right? That is just one of the automation biases. Whenever you calculate, you feel like the number is a little bit too high or too small but since it is calculated, it must be true.

We do the same for using the computer, which can do a little bit more complex calculations than a small size calculator. Basically, you program things and then your computer. . . . There are mathematical tools where you can put those equations in and then it gives you a numerical solution. Numerical solutions are approximated solutions, not analytical ones, so often they are not an exact solution. Because computers can only use finite

numbers but there exist infinite numbers that we do handle with those kinds of things. The students at the beginning do not understand; they use the built-in functions in the program so they can solve it. They do not understand they could be wrong. Through the classes basically, we try to educate the students to learn what's behind those built-in functions.

How is it calculated and what is the outcome? Then, what are the limitations of that one? Then, the students understand the limitations of those built-in functions of that software. Then, they can actually make their own judgment that this math is not appropriate to solve this kind of problem and they apply different numeral techniques.

These are some examples of very simple automation bias of calculation. By knowing the system and then understanding the limitations and then the actual "what's behind that black box." Then, that allows the engineers or anybody to make a better decision and then basically identify what's wrong and then what could be correct. That is one specific example of how you can train people to know better about the limitations of the technology so that you can actually make better use of it.

FRENCH: Yeah, I'd love that idea. As a philosopher, I'm drawn to the idea of being able to get inside that black box in various ways but also to ask questions. I think as we think about the future development of any kind of artificial intelligence, the ability to ask it why it did something and get a response back so you can start to understand. If it tells you, you should not go that direction and you could say, "Why? Is there traffic? Is there this? Why are you trying to not have me go that way?" Then, it helps you make a more informed decision. You have to be taught to ask those kind of questions and to be that kind of a critical thinker. I love the idea of your students asking, "I used the system, why didn't it work? It's harder than that. This must be . . ." I bet you have that conversation quite a lot.

LEE: Yeah.

FRENCH: All right. I'm now going to a similar kind of topic—but slightly different, say a bit more about bias—and this one is to you, Susannah, because I know this happens quite a lot in your field, although, we may find in others as well. That in additional to the one we just discussed, automation bias, the other big concern that people have around this area is the concern around bias permeating algorithms, the things that drive the AI systems.

A fairly well-known example of this involves a beauty contest that was held, where AI was supposedly going to select the winners in some super

objective way, much better than human judges could. They therefore estab-lished a universal standard of beauty. That was the project and it was set up by this online company called Beauty AI, I believe. None of you, I think, since you laughed you will not be shocked to hear how this turned out.

The people in charge of the machine learning, in that case, it turned out had of course only shown the system images that they thought were beauty. Predictably, the system just amplified their inherent bias and only spat out the images that they thought were beautiful. There was nothing remotely objective about it. Bias is a big concern, we know, in bioethics and certainly in patient care. Can you tell us more about the kind of issues that AI raises in terms of fair treatment for patients?

ROSE: Yes, it's an excellent question. A little bit more on the story with Beauty AI, because it wasn't just the perception of beauty. What ended up happening was that certain racial groups were considered more attractive than others. We start to deepen the story about how this might pertain over to healthcare. We also see that it's not just beauty, but there are also other examples from other companies. I don't think this is one company issue, where we see certain racial groups not even designated as human, for example, designated as an animal as opposed to being human. Or, racial groups that are for facial recognition software where it is less likely to be accurate for women and for minority groups, particularly for women of color.

We start to see the same issues going on with not only with how the data play out but also how the model is actually being trained and developed. This has relevance to healthcare and healthcare disparities and also health disparities. Why I say this is that there are several different concerns we have with the data that might be perpetuating or even increasing healthcare disparities. On the flip side, we might be able to use the same models to actually diminish them. This is where we might be able to play and think about this strategically.

In the data process, we have to make sure that the data are unbiased, so as you might be aware, there are a lot of disparities in the United States cur-rently in our healthcare system. Not only do certain groups of people have worse health outcomes, they have worse access to healthcare. This creates several different problems. First of all, they might be underrepresented in the data sets. Secondly, they might have worse outcomes.

For training models that say a certain group of people have worse out-comes, for instance, based on organ transplant or chemotherapy or ICU admission. Yet, the model might say, "Oh well, that person actually is less

likely to benefit," therefore the model doesn't recommend this treatment. You might see this being perpetuated. You also have the possibility—what I brought up before with the IBM Watson Oncology example—that certain people who are training the system have certain beliefs of what is better and what is worse.

This brings in my perception from the patient experience world, that what we don't often have are these stakeholders who are potentially affected by this at the table in terms of looking at these data, how they are put in, what the data represent, and also the values that are being put on what is a good outcome and what is not. This is where I go back to this label data. Many times, we say cured/not cured, that might be pretty clear. However, if you notice in healthcare, for example, in oncology, we're having more of a system where you're living with the disease. Therefore, potentially subjective outcomes by patients are actually the endpoints that we hope to achieve.

This is where I have some worries in terms of how we're actually determining the better, worse, good, bad outcomes. We have to have stakeholder engagement in order to define those. Currently, I don't see that occurring. One of my big concerns in healthcare—and maybe this is not the case in defense—but in healthcare AI, robots, these things are fancy and awesome and healthcare administrators are either afraid of it because they think AI is going to take over doctors and there aren't going to be any more doctors. The robots are going to provide all the healthcare.

Then, the other side is a sort of, I think it's going to be great. It's going to be horrible on one side and it's going to be great on the other. It's going to solve all of our problems in healthcare. We're not going to have any predicted uncertainty. Obviously, we know that probably neither of these cases are accurate. Where I see in healthcare, we don't have a whole as the clinician and as an ethicist, I think there's a whole lot of education that needs to be delivered to the people that are leading in our systems because we deal with life and death just like in the defense industry. I also think that there's a technology bias that sometimes occurs, particularly in healthcare in the United States.

Our entire healthcare system rewards what we call innovation. Innovation is defined often by a new product, something that is maybe better, or may not be better, but it provides some type of perceived benefit. Because our GDP—we're almost pushing twenty percent or something—is related to healthcare, you can see where this idea that new products and new things that serve and have not yet been proven to be beneficial still might be looking very fancy and new. And so I do play, I think, a big role for

self-education and educating our leaders on/about how these things can actually implement and what they actually cannot do.

FRENCH: That definitely speaks to me, just first of all when you talk about the two extremes. we do have that in military ethics certainly. It's expressed generally as the robots or tech will save us all side versus robots or tech will kill us all, the Skynet worry at the other extreme. Those definitely are out there. This idea of innovation being rewarded always whether it's good or bad innovation or whether it's been thought through in these ways, that is extremely important and very helpful in this conversation. Because I think in all of these fields, what you find is this magnet toward the shiny as you were describing. Most of us wouldn't be here if we were anti-tech. We're not trying to put genies back in bottles, we're for innovation. It always has to have these boundaries and these questions and these nuanced reflections in order to make sure that it's used in an ethical manner. I thank you for bringing that point out.

Let me go in a similar thread there following this along and ask a question to you, Kiju. Do you think that AI systems and other advanced technology will ultimately produce entities that are superior to human beings? I'm curious why or why not. I would like to know if you think it's helpful for us to think of robots as mere tools, just the latest tools that humans have invented to make lives easier. Because we know from the beginning, humans have made tools to try to improve our lives. Or, does that fail to capture something unique about what robots could be in the future? We've all seen science fiction shows. Could they ever be less like tools and more like persons?

LEE: Sure. Robots are not mere tools. I call my robots my babies. One of my robots is named Woody and my daughter often asks me, "When does Woody come home?" Because I told her that when it's ready, I'm going to bring him home. Then, she says that she is waiting for her little brother. Maybe my opinion is a little bit biased. They are certainly not mere tools, but they are not humans either—whether they're going to be superior to humans, they're already superior to humans in many tasks. They can calculate things much faster. They can process information of certain types much more efficiently than humans. They are not as ubiquitous or adoptable or flexible as human beings. They are not creative.

I can create my own machines and new mechanisms that can generate something new, but current AI cannot. There are huge gaps, but we should

not actually compare human beings to robots. I would even consider it as kind of a part of our environment, not as humans. It's like in the old days humans started making farms and building streets and those things. They started making the changes in the environment. Now, we are entering new types of environment, new types of built environment that is more dynamic and they can handle information differently.

In order to live in a certain environment, you have a certain amount of information to be processed by humans now that is shared by the environment. You build streets because now instead of handling all those uncertainties on the street, you make the streets, so you change the actual environment to handle certain types of information for you so you build the cars that can be more efficient. In the same way that AI will do something for humans although there will be always…we cannot prevent people using AI for bad purposes, unethical use but also…it's our homework too to work together. I think robots will be a different kind of entity, not something that we need to compare with humans. It's something that is there, we built our environment in a more intelligent and smarter way and in smarter ways that can help us to live in a better life.

FRENCH: I find that very helpful, that imagery with the environment. On a more comical note though, I have to say I now want to meet Woody too, so you've sold it. I also am reminded of how there have been stories, I don't know if you encountered any of this, of troops actually holding funerals for robots that were in service with them that, for example, had been helping to detect IEDs and other threats to them. When they were finally blown up, they genuinely felt a sense of loss. They felt something that had been there like a friend keeping them safe was now gone. I named my car but…

ROSE: I want to say, I felt the same when my car died. My old blue Volvo.

FRENCH: We have a history of bonding with this but this idea of the environment and how we change it and how it changes us and how we interact with it, I love the dynamism of that too that that's not one way either. It's just going to flow both directions. I think that's very, very helpful imagery.

So many great ideas, I hope you're all probably like, "When is she going to stop talking? Let us ask questions." We're getting there. I have one more for Margaret and then one more for the whole panel and then it's up to you, so be ready. I will be releasing you in a moment for your questions. Margaret, this one is to you. We know that people sometimes make bad

decisions—shocking revelation—because they lose perspective, particularly under great stress. Could we actually use AI systems to serve as ethical assistants, like having an automated conscience reminding us to take the ethical dimensions of our actions into account like a robo chaplain, as it were? I'm becoming a robo ethicist obviously.

KIBBEN: I have to tell you, I have a visceral reaction, negative, to the phrase robo chaplain.

FRENCH: Good to know, good to know.

KIBBEN: You could nuance it and call it a chap bot and it still wouldn't get any better. Partly because it robs the value of what not just chaplains or clergy in general bring to a person, but what they also bring to a community and to a situation. To take it away from the generic picture, let's talk about the possibilities to what you're after. That is that there are some situations, certain situations, even life threatening, which can be predicted. There are certain human reactions which can also be predicted. With enough input, artificial intelligence could be programed to advise on these.

One can even use artificial intelligence to develop training modules for situation response, for critical thinking opportunities, and then to evaluate that individual response or the unit response to those situations. I think you could definitely use artificial intelligence in that regard, in this scenario. You could also use artificial intelligence to simulate those scenarios more realistically. In other words, to enhance a training exercise and to enhance deployment workups to enable individuals to acclimate and adapt to a volatile environment. In other words, AI is creating all the pieces that create an environment in a benign situation to get us, the individual, to have a sense of all the kinetic elements they're a part of in that particular situation.

In fact, we're already using that. The element of this is that it enables an individual to create the muscle memory for more objective ethical behavior, to be able to set aside their own visceral responses, and to be able to respond in a more patterned idea. You used the phrase in one of your articles—and Lisa Lindsay is in here too, she co-authored this—here was a phrase you talked about: Could AI do "moral deskilling?"

FRENCH: Yup.

KIBBEN: I actually think you could use AI to skill in this regard. In other words, you're creating the environment using AI to simulate what's going on in some of these environments, with the controlled environment, with

the operative word there being controlled. If you were to do that, one would want to leave. . . . This is how we've been using it in military environments, one would leave room for human interaction and input, right, thus allowing the unprogrammed but just as right response. We'll assume that AI is providing only the right response. If in fact, in this training scenario, in this exercise you find that humans are responding differently than the way AI expected them to, you could actually evaluate that response as well.

To use AI in a dynamic life–threatening situation is fraught with limitations. The limitations fall along the lines of assuming your question implies that AI could advise in the moment. The elements of ethical behavior necessarily include the programmed value bias that we've been talking about already. If the situation that you find yourself in is live action, and you have AI standing beside you to advise you, one would want to be aware of what that bias is. The bias is limited in as much as it has no idea how dynamic the situation is. Humans are involved in the situation, and humans are complicated, and so is human reaction, and so is the climate in which you find yourself.

FRENCH: We are chaos generators.

KIBBEN: Precisely. Even if AI were able to advise human behavior in the thick fog of war, even if AI could come up with the right answer with enough deep learning and all the elements that have been programmed into artificial intelligence based on whatever preset and pure ethical principles it's been programmed with, I'm most concerned that regardless of the outcome, good or bad, an individual may say, "AI made me do it."

FRENCH: Accountability.

KIBBEN: Accountability. The other part of it is that what is the individual left with? If the artificial intelligence adviser gives them advice that morally the individual is at dissonance with, what is that person left with initially at that moment, how does that paralyze them at the moment? As we have seen time and again throughout battlefields, whether you want to call it post-traumatic stress or battlefield fatigue, there is a moral implication, there is in fact moral injury that takes place when some of these things are executed.

FRENCH: I'm so glad you brought in this idea of moral injury, because it is so central to all of these conversations. It is often understood, in part, as being the result of feeling betrayed by an authority. If that authority is

a program authority, you would still feel this and feel that you perhaps should have known better than to listen to the AI in that situation. It might actually magnify that injury.

SHARONA HOFFMAN: Hi, I'm Sharona Hoffman. I'm a Professor of Law and Bioethics at the Law School here at Case Western. This is for Susannah. All of you talked a lot about AI assisting in doing things currently, surgery, diagnosis, and so on. AI is also used for predictive purposes and I think that's a little bit different and raises some other concerns such as discrimination, loss of opportunities, and so on. I'm just wondering if you have any comments on predictive artificial intelligence in the health context.

ROSE: The way I understand—I'm not a technologist—but the way I understand AI is that it's all predictive. The idea is that you don't necessarily know the outcome. It's designed to essentially take large amounts of data to be able to anticipate and provide options depending on what the tool is made for.

If we go back to healthcare, the whole idea here is that, for example, if we link up our Fitbits and everything else, our Apple watches and everything else, it's not even healthcare data, we think about social media, we can think about all of these data sources that might be able to provide, even food and security in certain neighborhoods. We can use these data that are not even necessarily attributable to an individual person but perhaps attributable to multiple different aspects.

The idea is that we could absolutely start to think about anticipating these things in the future. The time horizon actually depends on the question that we're asking. When I mentioned healthcare allocation, I didn't literally only mean right this very second. We also think about health care allocation in terms of future states. We also can look at how disease progresses and we might be able to anticipate that future needs for various populations might grow that we might not anticipate currently with our current capabilities.

As we start to think about this, I was focusing a little bit more on healthcare ethics in the immediate sense. My background is actually in justice in healthcare allocation. When we think about these things, it's not just in the moment—my colleague Hilary Mable is here who's a clinical bioethicist at Cleveland Clinic. I could see the algorithms and training being applicable to the clinical environment, but I also think it's quite fascinating to think of how we might consider issues of justice and planning in the future too and how we think of theories of justice. How we might be able to use these

data to actually achieve perhaps something that we see as more fair or a better system. Right now, we're not necessarily using all the data sources we possibly can to achieve that outcome.

JOHN MANIN: My name is John Manin, I'm an AI programmer. I think that technology typically has lots of different problems, lots of technology can be used for military applications. This is for anyone who wants to address it. What is an ethical concern that you think applies to AI specifically, different from other technologies. What would be the type of safeguard that you would like to see in place to deal with that?

FRENCH: Anyone want to tackle that?

LEE: Yeah. I would start. I think, depending on who the users are and then the ethical mind of that user actually determines the ethical use of the technology as well. At the level of AI, one of the questions related to this panel discussion was whether we could be able to actually implement some safety features that will prevent the unethical use of technology. At the current level, I don't think it is feasible, but I think it is possible in the future. When the AI starts advancing and then the actual system itself has built-in features so if some kind of odd behavior is detected, you can actually block the system or block the user or something like that.

I know that a very limited level of that kind of feature is already implemented in some online systems or online behaviors. It's not in all AI systems. For instance, if you tell them new algorithms and new AI systems can be used for different robotic systems, depending on who uses it, there's no way you can actually block them and use it for unethical uses. I think that was more of the question from my side as well, probably. I will give it to you.

FRENCH: Yeah. I will just add that unfortunately, we have some examples, we gave a few. There are many examples of AI just amplifying what is not so great about people. I think you can say that about most tech, about any tool. I can take a pencil or a fork and kill somebody with it. It's not the thing that has the moral value, it's how I use it. I think that's absolutely valid. I think with any new tech we want to say, as we have been, is there a way to make it a little safer or not? We will relate to it differently whether it is capable of being safeguarded or not. It doesn't mean we won't use it, but we need to be aware of what it is or is incapable of. Either of you wanted to add?

KIBBEN: I would like to add to that. I think part of it too is that aside from the machinery, if you will, and the unintended consequences of how

a machine can take something forward, I think the challenge that I find in our culture today is that we've lost the ability to communicate well with each other.

FRENCH: This panel being obvious evidence to the contrary, but nevertheless.

KIBBEN: Right. You can take it back to all we do . . . when you watch kids break up on texts, right? If we've lost the ability to communicate with each other as human beings, then we will also have lost the muscle memory of how to program communication into systems. If we've lost the ability to communicate what our intent is and what our values are, then we don't know what we're going to end up with, unintentionally, in the machinery.

Interestingly enough, MIT has started a study to see if we could start talking about technical morality in the middle school. So that when kids are learning programming language at the onset, they're beginning to talk about things—not just how do I add things together? How do I make this goal happen? The why behind it and the consequences that are behind it and the value set that contributes to that dialogue. That's what my thoughts are.

FRENCH: I find that very helpful too, the idea of focusing people's attention on who is affected by everything that you do. Again, can you modify it in any way to keep people safer?

HILARY MABEL: My name is Hilary Mabel, I'm a clinical ethicist at Cleveland Clinic. I want to actually push the safeguard question a little bit further. We talked a little bit about building safety or safeguards into the technology itself and even training or ethics education on how it should be used. I want to ask, we've been talking a lot about AI in warfare, which I think is a very stark example of the harms that can come from AI. There are probably a lot more everyday harms that we might think about: privacy issues, trickery or disrupting civil discourse, different things that AI could do that could be harmful, right?

I want to ask, what is your perspective from the panelists on regulation, for example, what should that look like? Who should be at the table in terms of setting the agenda on something like that?

FRENCH: Another great question. Who wants to go first?

ROSE: I would say that at least in healthcare, the way I understand the way AI is being deployed is that currently, it's not in a systematic way. There are a lot of AI—I'm looking at my AI technologist here—experts. In healthcare,

I went to HIMSS Conference, which is one of the largest conferences in the United States. It focuses on health and information technology. What I found interesting is that even the course of the previous year's conference is just a pop up of AI pretty much everywhere.

Why this is relevant is that there seems to be a deployment of AI in whatever form it may be into various microsystems within the healthcare industry. What I see happening is there's not necessarily a thought–out process on this. There's certainly, to my knowledge, there's no standard that this is actually regulated by really anybody, right? Mostly, it's not even research. It doesn't even fall under IRB, which is where we think of human subjects research being evaluated.

There has been a call at every level for increased regulation. I also think that the organizations themselves hold some responsibility for this, if not a lot. Organizations themselves haven't really thought through how they're actually deploying many technologies, including AI and many other computing capabilities. Where I see this going is not . . . I don't think regulation is necessarily going to be an answer. I look to the folks that know more than I do on this, because it is never going to necessarily keep up.

What I do see at play that's really important is how we put those people at the table. You know what my answer is going to be in healthcare is that we have a very few . . . an underrepresentative very important stakeholder voice of the patients, of family, of clinicians that are not necessarily physicians. We also have the community, religious leaders, we have other folks that are . . . we have people that are at the table, journalists. There are so many other stakeholders that bring so much to this perspective. I don't know if this is actually specific to AI, I think this has to do with a lot of other technologies as well.

KIBBEN: I think you're spot on. I think from the control perspective, one needs to keep an eye on it. The challenge with regulation is that it then becomes law. Just because it's legal doesn't mean it's right. Frankly, just because it's right, doesn't mean it's kind. I think, now I'm back to the virtue piece of this. When you talk about things at the table, I think you have to have elements of all of those arenas, as was already articulated. You need to be able to talk about the people who can control the language, the people who understand what the rightness of the intent is and also the virtue of what is intended.

TONY JACKSON: Tony Jackson, philosophy. I want to pick up on the themes of automation bias and can we code morality. In the classic trolley

problem of the fat man on the bridge, it's very easy for anyone to articulate and my students can regularly do it. Why you should push the fat man, right, it's a matter of simple math. In most cases, students at least know that much. It's much harder to articulate why you shouldn't push the fat man, even though every civilized legal system says it's illegal and it seems to be a clear human rights violation.

We know that more empathic people are more inclined toward not pushing the fat man and more analytic thinking makes you go toward pushing the fat man. Of course, ethicists and the whole trolley problem came up because ethicists don't all agree.

FRENCH: Wait? We don't? [Laughs]

JACKSON: Yeah. No, I agree with you.

FRENCH: I agree.

JACKSON: Okay. My question is like a creeping effect I'm worried about with AI. Some types of morality look pretty easy to code. I mean, utilitarianism looks perfectly suited for coded morality, right? Down to logical principles, okay, well, good luck, values…

FRENCH: Virtue ethics that…

JACKSON: Right. Explaining what a virtue or a value is doesn't look like something, but maybe Dr. Lee will contradict this. I guess what I'm asking the panelists is, do you think there are particular types of ethical understanding that will forever evade AI? Therefore, is there a concern of a creep? I guess, I'm particularly interested in Dr. Rose's response because my sense is the reason the clinic has someone who is the head of patient experience is because they realize that the kind of utilitarian thinking that drives a lot of medicine leaves out something, which is something like human experience. I wonder if that's one thing that might get left out the AI equation.

ROSE: The way I'm understanding your argument/question is that you're asking whether there are certain types of ethical frameworks, for example, or theories that might be more amenable to programing? A, I don't think that's the case, because yes, you're right. On a first blush utilitarianism might be able to. You're a philosopher, right? These ideas about utilitarianism, you start digging in and things get quite complicated pretty quickly. If we're just looking at very simple calculations of saving life versus not saving life, perhaps.

I think that as I mentioned before, there's, I think, very nicely with the idea of value is that these endpoints are not necessarily so clean cut, which

you're right, it brings in the idea of the patient and the patient's perspectives and values. I don't know if one theory is going to be more amenable for programing than the other. I do not know the answer to your question in terms of can we actually have—I think my panelists might be more suited to answer this question—in terms of how we can actually start to imagine ethical decision making being programmed.

I'm not so sure that it's contrary or not impossible. We have a database that has—I don't even know how many cases now at Cleveland Clinic— over a decade of clinical cases that we have gone through and evaluated in our bioethics department. It would be fascinating to go through and look at these data in terms of the outcome, the recommendation on how this might be married up with AI technologies, right? I don't know. There could be some unearthing of consistency that we might not completely be aware of even.

I would actually be fascinated to see if that happens. In terms of technological capabilities, I don't think we're there. I think you said that that right now the programming of more ingrained ethical theory is certainly not the current standard of the technology.

LEE: Right. I think it's just the challenges in technology. As an engineer, the challenges are in the human's opinions or human's judgment. Ethics, what is ethical and what is unethical. Even we do not have a clear cut answer. Those things are the challenging components. When you think about what is good and bad. If for instance, one specific example, when we develop a sociable robot, you want to have the robot to understand human's emotions.

That is one of the very challenging questions. The robot, we train a tremendous amount of data that is labeled by humans saying that this data set is happy face, this data set is neutral face, this data set is sad. Then, we train it. Then, the system can automatically determine, okay, that the person interacting with the robot is happy face for a while. When the robot does something, it's sad and upset or something like that. Even that level of the human behavior is so complicated for our current status of the technologies to understand.

Then, when you think about ethics, it's much higher level complexity. It's not because we don't have the technologies. We have the technology or AI that can handle such a large amount of data, but the thing is who determines what is ethical? Something that is ethical in this country is not ethical in some other countries. The people's preference or individual

differences plays a role too. There are so many questions and it's more like a fluid and not a solid. That makes AI adopting those human notions into the system more difficult and complicated.

FRENCH: I would just add on that point that even internationally across cultural differences, ethicists may agree on some very core principles and I would argue that they often do and so do cultures themselves agree on some core principles. Then again, the expression of that and how that plays out in practice gets incredibly nuanced and complicated. You're absolutely right about that. Did you want to add from your . . .

KIBBEN: I think the only thing I would do is underscore something that Dr. Lee said earlier. The mechanical engineer indicated that the one thing that humans have that machines don't is creativity. I think one also finds creativity in ethical decision making and ethical response. We will never ever—and I can say this unequivocally—lose our creative and virtue inclination. I do not think that machinery can outpace us in that regard.

SAMI: Thank you for sharing your expertise today. I'm Sami, I'm an MPH student here at Case.

I'm also the president of the Global Ethical Leaders Society here at the Inamori Ethics Center.

I want to preface my question a little bit, Dr. Rose, you talked about the ability to assimilate or even replace the role of medical providers on an individual level. I wanted to ask more about healthcare decision making on an organizational level serving populations. What about hospital administrators, board members of health centers or even healthcare policy makers, the ability of AI to consider the needs and characteristics of a population pulling in all these data. Like Dr. Lee said, being able to have that superior capacity to make these complex calculations, incorporating the interest of stakeholders, considering health disparities like you said.

In a lot of these decision making processes, you're looking at tons of pros and cons but may not be able to quantify all of these. For these board members that are making these decisions, could AI have an impact or are they already potentially using AI to make these organizational level management decisions. If so, how effective are they? I'm sorry, this is a second part question too, but Admiral Kibben, you also touched on the topic of accountability, we already use AI systems to make these kinds of large scale decisions who would be liable if this had a negative impact on a subset of a population? Sorry, it's long.

ROSE: I actually have no idea if hospital administrators are using AI. I suspect not right now. I think where healthcare generally is applying these technologies is generally in the diagnostic area right now and somewhat in the populations health area. It's possible, as I said before, that it could play a very important role, particularly when we look at allocation and decision making across an entire system or systems. There's actually a potential for this idea of augmented intelligence, which goes back to the question we talked about before. I think about my role as a bioethicist in a hospital and I'm not even sure necessarily if I'm consistent across my own self in situations.

I cannot get inside the heads of my colleagues. What I am curious about with AI and other types of situations like this, especially if we can quantify in some way, some of these issues that we currently don't quantify is perhaps with assistance with some kind of augmented intelligence that might help us be more consistent. Back to the idea of, well, Susannah, maybe you don't agree with Hilary, right? Maybe even these nudges, right, that potentially, if they're done in the right way might make us more consistent if we value consistency across similar situations.

You might see perhaps, in many ways, the structuring as being even helpful in those individual decisions. I am really a skeptic of AI in certain areas and I have seen head-to-head trials on a very, very few scenarios in healthcare. I should make that very clear. I'm a researcher and so until I start to see how the counterfactuals, if I start to see randomized trials, I want to actually see how administrators make decisions with and without. Because though we say that thirty percent of the time we're able to prevent a stroke by using this system, well, how does it compare to before? Would the patient actually know the symptoms? Were they actually better than somebody telling them to go to the ER?

We actually don't look at this. Human judgment might be better in some cases and it might be worse. We just don't know that because we haven't analyzed it, certainly at the hospital administrator level, unless somebody can correct me on that one.

FRENCH: Margaret, did you want to add comments on the account-ability issue?

KIBBEN: I really appreciate you raising the question about accountability because I've talked a little bit about the accountability in the programing piece of it, in the coding part of it, and that we want to make sure we're holding ourselves accountable to some sort of moral code. I also think

we need to make sure that we are maintaining our accountability to the execution and utilization of artificial intelligence. I think we launch it out there and say, "Right, shiny penny, it's going to work," there we go. I think Dr. Rose's point is well taken and really speaks clearly to this idea of accountability. How do we know and how do we make sure that we're tweaking it in such a way that it improves and doesn't harm?

FRENCH: That reminds me back to, again, Kiju's point earlier about how you can't control who is going to take once you create certain systems, people can take that potential. As someone, again, who works in military ethics, one of the things we often have to face is that even technology that was invented and does have some safeguards in one area leaks into other fields. Something you created in one way gets adopted elsewhere and that is almost an organic process that you can't really put the brakes on. I guess I said genie out of bottle. Right here, sir.

ANDY PODGURSKI: Andy Podgurski. I'm a computer science professor here at Case. I was happy to hear Dr. Rose use the word skepticism in relation to AI. I wonder if we would be better served as a society if we were a little more skeptical about it, because when we talk about AI as if it were close to accomplishing human intelligence, I think we mystify it. From the standpoint of most people, a better model for the way AI works today rather than the human mind is statistical pattern recognition. To understand its strengths and weaknesses, you're probably better off focusing on things like sample size.

FRENCH: We can't get the bias out of that either. We have had a lot of issues with bias.

PODGURSKI: Right. The same issues that statisticians worry about are huge issues in machine learning. Another thing to bear in mind is that my colleagues in artificial intelligence have been very good marketers over the years. Since I was a graduate student, I had to tip my hat for their marketing genius. For instance, the term deep learning was a renaming of something that's been around for thirty or more years. It really captured the imagination of the public. It's the same thing—sorry, forty years ago—certainly, there have been improvements in the technology and there have been impressive achievements. I don't mean to diminish it entirely but it's not really close to human intelligence.

FRENCH: Thank you. I appreciate that so much in particular because one thing that I've argued—and it sounds tongue in cheek that actually

I'm dead serious—is that rather than seeing it as an arms race and that we have to beat Russia and China to the development of AI, I've argued: let them do it first. The old saying, it's the second mouse that gets the cheese. If we let them rush to deploy systems that can't do what they've been sold that they can do, it becomes a weakness for them that we can exploit. From that point of view, the skepticism is not only healthy, it's strategic and ultimately ethical if we keep to just war tradition, hopefully. Kiju, would you like to comment any further on some of these questions of potential?

LEE: Sure. I actually totally agree that AI is not something that we have to compare to ourselves: It's just an entity. It's a technology that we want to use it in a good way and in a helpful way to do more good than harm. It's not like we are not in danger of losing our human beingness because of AI. AI will be a part of our lives, part of our environment. We want to just continue our environment to make us live in a better life. I totally agree with the question.

ROSE: I think some days I'm sold, some days I'm more skeptical. I think I'm vacillating between the two, right? I think that's how it is in healthcare as well with many technologies. Many of them have actually been around for quite a while, but then they're deployed and used in much wider ways and in ways that originally were not necessarily intended. That's where we get into some really fascinating areas in healthcare. I also, from what I understand and correct me if I'm wrong, it's not necessarily some of the new technology, the systems, but it's actually the access of data and data availability.

That's where I also see—and I think Hilary brought this up—in terms of data security and other issues. We didn't really touch on that here. I think that there's fascinating issues with regard to that, particularly because the computing capabilities are getting so much better and faster and bigger. Especially as we, no pun intended, bleed out to other areas. I guess I could hold for healthcare and sadly for war, where you have these other data sources that might...in healthcare we have HIPAA. We have other regulations that put some protections around data going out. There are actually fascinating ethical issues of data coming in, in terms of are you actually agreeing to let your Fitbit hook up with our EMR records?

Are we, in terms of healthcare systems, pulling in other data that might be helpful in terms of directing and suggesting care that might be really, really good for people. However, you start to think about where do those boundaries start to take place. As a society, we risk violating trust if we don't

have some idea around this. I think there's a whole set of issues, they're not necessarily specific to AI exactly. It's because of AI capabilities and the whole process of data acquisition that I think we have some unique challenges.

FRENCH: Excellent. Would you like to comment AI skeptic or fan?

KIBBEN: Yes.

FRENCH: Both.

KIBBEN: Yes. I think I'd like to widen the aperture even further. As a person of faith, I believe that we were given the gift of our brains as well as many other things that we have had the opportunity to use. We've also been given the gift of autonomy, our own autonomy, our own will, our own volition. With each of our gifts, our skills, our talents, the technologies, all of these things, our autonomy, we were also given responsibility. The responsibility that we have is to use them for the benefit of creation. Wherever we find ourselves in the mechanics of artificial intelligence at its creation, at its utilization, at its oversight, we need always to keep our eyes on our responsibility to the benefit of what it is we've been given and how we use it.

FRENCH: On that note, thank you, ladies and gentlemen. Let's thank our panel one more time.