Irène Buvat and Ken Herrmann Talk with Alexander Stremitzer, Kevin Tobia, and Aileen Nielsen

Irène Buvat[1], Ken Herrmann[2], Alexander Stremitzer[3], Kevin Tobia[4], and Aileen Nielsen[3]

[1]Institut Curie, Orsay, France; [2]Department of Nuclear Medicine, Universitätsklinikum Essen, Essen, Germany; [3]ETH Zurich Center for Law and Economics, Zurich, Switzerland; and [4]Georgetown Law, Washington, DC

Irène Buvat, PhD, Centre National de la Recherche Scientifique (CNRS) director of research and head of the Inserm Laboratory of Translational Imaging in Oncology at the Institut Curie (Orsay, France), and Ken Herrmann, MD, MBA, professor and chair, Department of Nuclear Medicine at the Universitätsklinikum Essen (Germany), talked with Alexander Stremitzer, JD, PhD, professor, ETH Zurich Center for Law and Economics (Switzerland); Kevin Tobia, JD, PhD, assistant professor, Georgetown Law (Washington, DC); and Aileen Nielsen, JD, ETH Zurich Center for Law and Economics, whose article "When does physician use of AI [artificial intelligence] increase liability?" appears in the January issue of *The Journal of Nuclear Medicine* (2021;62[1]).

Dr. Buvat: You have recently published a very scientific paper on a quite unusual topic for our journal. Could you please introduce yourselves?

Mr. Stremitzer: I am a lawyer and economist. I used to work at the University of California at Los Angeles until 2 years ago. I am now at ETH Zurich running an interdisciplinary group of lawyers and social scientists, psychologists, philosophers, and even a physicist, working on legal matters and running experiments on legal institutions.

Mr. Tobia: My background is in law and philosophy, as well as experimental cognitive science. I am also an affiliate and former member of Alex's group at the Center for Law and Economics at ETH Zurich.

Ms. Nielsen: I am a member of Alex's group at ETH Zurich. I have done previous studies in physics and anthropology, as well as being a U.S.-trained lawyer. I run experimental studies looking at the interplay of law and technology.

Dr. Buvat: Could you please summarize your study and its main findings?

Mr. Stremitzer: As you know AI precision tools often make personalized treatment recommendations, such as recommending the dosage of a drug based on a patient's file—an example we used in the article. Medical malpractice liability requires a deviation from the reasonable care standard, which is traditionally met if the physician exercised standard care. This legal standard, however, raises the possibility of exposing the physician to liability when following nonstandard advice offered by AI. This could undermine the very promise of precision medicine, which is to offer personalized and therefore nonstandard advice. We recognized that this was definitely a theoretical possibility from the point of view of legal

doctrine. But we also saw a countervailing force, namely that AI advice may increasingly be viewed as the new standard.

How the legal system weighs these points is partly an empirical question. Because medical AI is a relatively new development, we cannot observe a lot of court decisions. Therefore, we decided to exploit the fact that determinations about liability (at least in the United States) are made by jurors, who are laypeople. We randomly assigned a representative sample of U.S. adults to different scenarios to study how they judge the reasonableness of physicians' treatment decisions in different scenarios. In those scenarios, AI recommended either standard or nonstandard care and the physician either accepted or rejected that advice. All scenarios, however, shared a common result in that the patients turned out to have been wrongly treated.

We found that laypeople want doctors to follow both the standard treatment and the AI advice. In short, we found that physicians who accept standard advice were exposed to considerably less liability if something went wrong than physicians who rejected standard advice. The really interesting point was that whether physicians accepted or rejected nonstandard advice, they were almost equally likely to be held liable when something went wrong. The physician may even be in a slightly better position when accepting nonstandard advice from AI. This result suggests that laypeople, acting as jurors, are far more open to AI recommendations than one would have thought.

Dr. Buvat: Was that what you expected?

Mr. Tobia: We went into the study with an open mind, with the idea of attempting to adjudicate among several different models of how laypeople would make these kinds of liability judgments. A prior article in *JAMA* by Price, Gerke, and Cohen (2019;322[18]:1765–1766) reported on a related doctrinal analysis, and that analysis formed part of
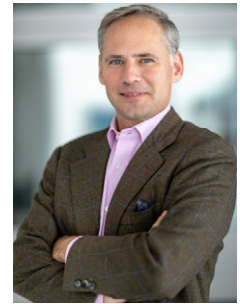
Alexander Stremitzer, JD, PhD

Kevin Tobia, JD, PhD

Aileen Nielsen, JD

our background expectation that the current state of the law would favor physicians following standard care over nonstandard AI recommendations. That article also suggested that the current favoring of standard care might change quickly, so that perhaps in the future physicians could incur liability for rejecting correct AI nonstandard recommendations.

So, an unexpected aspect of our results is that they suggest that we are actually a bit further down that line than expected. Another surprising aspect is that other research has found that laypeople are averse to algorithms in certain contexts. Here we found that when the AI recommendation was for nonstandard care, accepting that advice was evaluated overall as more reasonable than rejecting it.

Dr. Herrmann: My conclusion is to follow AI advice. If AI advice is standard, I am lucky because I am not going to be sued anyway. And if I am unlucky and it is not standard, I have a higher likelihood to be sued. But still less than if I reject it. Correct?

Mr. Stremitzer: The short answer is yes. This would contradict the conclusion from the doctrinal analysis in the 2019 *JAMA* article, which was much more skeptical about how courts would treat AI in medical applications.

Dr. Buvat: The population that you surveyed was from the United States. Do you think the results might be different in Europe or in Asia, where citizens might not have the same views on AI algorithms?

Mr. Stremitzer: That is certainly possible. We tried to get a representative sample of U.S. adults, because in the United States

Dr. Buvat: Can't the AI algorithms just be considered as a different type of medical expert? If so, what does that change regarding the liability of the physicians?

Mr. Stremitzer: If AI is taking more and more responsibility away from the doctor, it would be conceivable that, in the future, we could see increased liability for tool developers, similar to product liability for other products. Liability might be shared or might be shifted away from the physician to the developer or manufacturer of the tool, but this is pure speculation.

Mr. Tobia: We have another study that addresses this question a little bit. We take up this strategy that you suggested of thinking about AI as a colleague. We consider how people evaluate an AI as something the physician might rely on versus a human that a physician might rely on. That project is still ongoing, and we hope to share the results soon. But perhaps people will start to see AI as more like a colleague and less like a tool. More broadly, building on prior work, we found that one feature that predicts people's reasonableness judgments is their conception of what is common. If AI is seen as more common in practice, people will grow increasingly tolerant of physicians relying on AI advice.

Dr. Herrmann: The acceptance of AI is, based on your study, already quite groundbreaking. If you talk to people who provide these systems, they say they provide you only with a decision support system. But based on your results, I would say in layman's terms, it is not only accepted as a decision equal, it is actually accepted as a superior. So if you do not follow an AI, you are screwed! For me,

---

"[L]aypeople, acting as jurors, are far more open to AI recommendations than one would have thought."

---

the legal system relies on the judgment of laypeople. One major difference in Europe is that the legal system relies on professional judges. This difference might be even more important than intercultural differences. We could easily run the study on laypeople in different countries, but it would be even more interesting to run a similar study on European judges. These are, of course, much more difficult subjects to recruit. It could very well be that there is a difference in how laypeople, medical professionals, and the legal elite view those questions.

Dr. Buvat: Do you think that when physicians do not agree with what the AI suggests, this should be mentioned in the medical report?

Mr. Stremitzer: It might be that the AI advice and the physician's judgment differ. I think it would be hugely valuable if doctors document such differences, even when they end up following the AI advice. Physicians may even be humble and say to themselves: "Who am I to know better in this situation, given that the AI has access to so much more data?" But if physicians document differences, areas are likely to emerge where AI advice and doctors' judgments systematically deviate from each other. Exploring the sources of these divergences would probably be very interesting for AI precision tool development.

Dr. Buvat: Might the liability of physicians be different if they gave some explanations associated with final decisions to follow or not follow the AI recommendations?

Mr. Tobia: That would be an interesting further question to investigate with a similar empirical study, testing whether laypeople's liability judgments respond to the fact that more justification or explanation has been offered.

the doctor going to school, having 15 years of experience, this is actually completely counterintuitive!

Mr. Stremitzer: The doctor's judgment is still important. Remember, if there is no harm there is no liability. In other words, if a doctor has a strong reason to believe that the AI is wrong, it would still be rational to reject the AI recommendation, as this would lower the probability of harm. Admittedly, if a doctor has no strong prior beliefs about the right treatment, the AI should be followed to avoid liability, even when getting nonstandard advice. But this might not be a bad thing.

Dr. Buvat: Your discussion about standard care is extremely interesting, because we are in the era of precision medicine. Precision medicine means that each patient is unique and that you have to customize the best treatment for each patient, so the notion of standard care might become less relevant. Do you think physician liability will increase because there might be less and less standard care?

Ms. Nielsen: Whether precision medicine increases the risk of liability will depend on the population and how particular subpopulations perceive personalization. It is true that our experiment suggests that receiving standard advice and accepting that advice creates a relatively strong shield against liability—more so than in other cases. However, this might not be the case as treatments become more and more specialized and what constitutes standard care becomes less clear. Also, as Alex mentioned, it is not clear to what extent the responsibility will stay with the doctor. Increasing reliance on AI products may go in tandem with this drive toward personalization, in which case we can expect more liability to be

allocated in the form of product liability to app developers who are creating the baseline algorithms.

Dr. Buvat: *So this is good for physicians. They will not get sued any longer!*

Ms. Nielsen: Right, potentially. So that would be having your cake and eating it, too, so that you have less liability but potentially retain full authority and decision-making power.

Dr. Buvat: *I was wondering whether you think physicians should be trained more on AI technologies to have a better understanding about the ways these things work and make the best use of them. Or is the alternative that as long as they are U.S. Food and Drug Administration (FDA)– or European Commission– approved, physicians will rely on them anyway?*

Ms. Nielsen: What we as lawyers can tell you is how FDA approval works and how that should affect physicians. What we know is that FDA approval establishes a bare legal minimum for sale in a medical market. It usually does not shield you from the possibility of tort liability for medical malpractice or for product liability. Clearly, physicians need to know more when they are considering liability implications, just as with a medication, a medical test, or a medical device. Mere knowledge of FDA approval is neither enough to operate responsibly nor enough to make liability decisions. But the extent of that education and the shape of that education very much need to be determined based on experimental and empirical studies to optimize outcomes when these tools are incorporated into practice.

Dr. Herrmann: *But shouldn't the layperson be trained to understand how the algorithms work, because they seem to trust them quite a bit?*

Mr. Tobia: That is a great question. We are studying laypeople in part as potential patients, but largely as potential jurors. In a real trial, there is also going to be other evidence beyond what we have in our study. There will be expert testimony about the device that is used, so this also highlights a limitation. There are a number of good reasons to have that education of laypeople in terms of how medical AI is and could be used. It would help in certain contexts to secure more meaningful informed consent and in coordinating expectations between patients and physicians. It is an open question whether a physician's or medical expert's view of AI is similar to what we found in the lay sample or different. Empirical work and also education can help clarify that question and also coordinate those sets of expectations.

Dr. Buvat: *As lawyers, do you see any danger in the spread of AI algorithms in the health-care system?*

Ms. Nielsen: Absolutely. Of course, it is our job as lawyers to see potential problems or dangers. Rubber stamping is a concern. We want to make sure that we keep human physicians engaged, especially as they will continue to have meaningful exposure to legal liability. We want to make sure that their decision-making authority continues to be proportionate to that. Thinking about the data science element, we also know that existing data-driven methods do not necessarily have all the strengths that human physicians have. We know that there is a danger of runaway feedback loops. We know there is a danger from biased datasets. So, we very much see a place for humans in the loop to avoid these dangers.

People might think that these problems are solved. I would point out that in October 2019 researchers reported in *Science* (2019;366[6464]:447–453) the case of a major U.S. insurance company rolling out an unintentionally racially biased algorithm because it had used medical spending as a proxy for seriousness of medical condition. Unfortunately, medical spending has a great deal of racial bias built into it. So even a very sophisticated national health insurance company unintentionally rolled out a biased product. It is very important to recognize that this continues to be a problem and that having humans in the loop may be helpful.

We can also wonder about the balance of innovation versus safety. Laypeople are perhaps enthusiastic but not sufficiently educated about these AI tools. We can worry that there might be some sort of public relations push, for example, by hospitals or competitive pressures to adapt these tools before they are fully vetted or appropriate. Likewise, regulators are just getting into the business of regulating AI products. The FDA, for example, said more than 10 years ago that they would regulate these products, but there have not been that many products through the pipeline. Unlike in medication or devices, where they have decades and decades of experience, this is clearly going to be a new learning experience, even for very sophisticated regulators, to think about how to balance innovation with patient safety.

Finally, we want to think about a balanced perspective and encourage that balanced perspective in many stakeholder populations. We want patients, physicians, hospital administrators, regulators, and app developers to be skeptical but not too skeptical. These are some of the dangers. Now we need to think about how to avoid these dangers as we promote the proper regulation and legal regime.

Dr. Buvat: *Can you tell us about the advantages that you see from the lawyer point of view?*

Ms. Nielsen: Many of the advantages from a legal perspective are, in fact, solving some of the dangers. Bias in medicine is certainly not new or the result of AI. We know that our physicians are unfortunately and unintentionally sometimes producing unintentional bias in the system. You can see AI as an opportunity to reduce that. Just as humans can complement machine learning weaknesses, machines can complement human weaknesses. Humans are not necessarily good at high-dimensional data. Humans are not necessarily good when they get tired. These are the sorts of things that can lead to legal liability. When a human is looking at a file, he or she may be overwhelmed because there are so many inputs or may be working overtime and so may not be physiologically in the best position to make decisions. They will nonetheless face potential legal liability for mistakes. With AI, humans now have a machine that may help them prevent some of these errors.

There are certainly opportunities for growth as far as the legal perspective on medical malpractice. But, ultimately, I would say we are a little bit out of our depth because we would also be looking to scientists and physicians to tell us where the opportunities for improvement are on the ground. Then we can think about how to make sure the legal system sets the appropriate incentives for adopting those when they mature and are appropriate for mainstream use.

Dr. Buvat and Dr. Herrmann: *Thank you all for your time. Your input in the passionate "AI for health care" debate is extremely valuable.*