AMERICAN PSYCHOLOGICAL ASSOCIATION

ESSENTIAL SCIENCE CONVERSATIONS:

ARTIFICIAL INTELLIGENCE AND PSYCHOLOGICAL RESEARCH: CAN AI REPLACE HUMAN PARTICIPANTS?

(APRIL 12, 2024)

TRANSCRIPT

Shandol Hoover: Hello, everyone, and welcome. Thank you for joining us today. I'm Shandol Hoover, APA Director of Science and Special Projects and Implementation. This program is part of an APA series called Essential Science Conversations where panelists and audience members can engage in an open dialogue about emerging topics and psychological science.

Before we get started with today's session, I want to share a few quick announcements. First, we hope you'll visit apa.org/science to learn how APA helps psychological scientists and join over 50,000 other subscribers by subscribing to a free *Science Spotlight* newsletter to get firsthand insight into funding, news, events, resources and more. We also invite you to subscribe to our Free Editor's Choice newsletter to get articles directly to your inbox.

Second, thanks to many of you who submitted questions for today's program when you registered. You can also ask a question as this program is taking place in real-time. There's a Q&A feature on the dashboard. Please enter your questions there. We'll be monitoring those questions throughout the program. Finally, this program is being recorded. In about two weeks, everyone who registered will receive an email with a link to the recording, and a transcript in about two weeks. Now, I'm excited to turn things over to Dr. Mitch Prinstein, Chief of Science.

Dr. Mitch Prinstein: Hi, everybody. Thank you so much for coming. Today, we are excited to host a lively conversation about the role of AI and research, including potentially for AI to replace human participants in psychological research. What are the possibilities and what are the concerns here? What are the considerations that we need to be thinking about or the unintended consequences of AI in the science world? Our researchers evaluate studies that use AI tools. I'm very pleased to introduce a distinguished panel for today's discussion.

We have Dr. Mohammad Atari, who's an assistant professor of psychology at UMass Amherst. We have Dr. Kurt Gray, a professor in psychology and neuroscience at UNC Chapel Hill. We have Dr. Jerri Lynn Hogg, a media psychologist, researcher and global speaker on AI, Dr. Rose Sokol, who's a publisher for APA journals and books, and we have Dr. Sang Eun Woo, a professor at Purdue University. All of them are amazing.

Thank you all so much for being here. This is great. I'm very excited that we're talking about this topic. Everyone is talking about AI these days, but it's going so fast that it's hard to keep up with what we're even referring to when we talk about AI and all of the different ways that this could be influencing research in so many aspects of what it is that we do.

Let me start with Mohammad, if I may. The idea of AI being used in research or much less for replacing human research participants. Walk me through this. Is this a great idea? Is this something that's even possible? What can you tell us about this?

Dr. Mohammad Atari: Yes, thank you very much for that question. First, thanks to the organizers. I'm really happy to be here today, and it's wonderful that we have so many people attending, which is a testament to the importance of the topic and the importance of having discussions like this. I think it is a really important question to ask, "Can AI replace human participants?"

I think there is a question to ask before this, and that is, "Should AI replace human participants?" We should first talk about the risks and opportunities in different kinds of AI, replacing human participants. After we have the answer to that question that others have posed. Dr. Molly Crockett, for example, has focused on that particular question among other researchers. I think first, it is important to have a discussion about that question, "Should AI replace human participants in psychological research?"

Then after we have an understanding of that question, we can proceed to asking can AI replace human participants. In both of these questions, we have the term human. One of the things that I'm really interested in is cultural psychological diversity around the globe. I just want to make sure that when we talk about human participants, we keep in mind that there are about 8 billion humans around the globe, and there is a lot of cultural, psychological diversity around the globe.

What humans or which human populations exactly do we want to study is AI a suitable replacement for a particular population? These are just some of the questions that I want to pose. I know that I did not exactly answer your question. I just posed more and more questions, but I think in this discussion, it is important to talk about the feasibility of AI replacing human participants and also really paying attention to the human part, which humans are we talking about here?

Dr. Prinstein: That's great. Thank you so much for raising that. Jerri Lynn, what do you see as some of the biggest opportunities or even limitations to thinking about AI within our research work?

Dr. Jerri Lynn Hogg: Thank you. Thank you, Mitch, for that question, and thank you all for being here today. Actually, I think that's a brilliant question that you also posed, Mohammad. That gets to some of the potential and opportunities in AI. For example, there's a company out called Real Chemistry, which is actually looking at rare diseases and looking at how because they're rare diseases, there's not a lot of information, but if we can access that through AI and all that information and potentially recreate participants, if you want to call them that, we might be able to better understand what's happening there.

I see that that could be further generated out through psychological research. I don't believe we're there yet, and I think that there still has to be a lot of work done before we can be there but a great question that you brought up. I think also to that same line of thinking, which is when you asked me, Mitch, about the opportunities and then possible pitfalls there, is that traditionally research has been done.

We know the medical research often was done on White middle-aged males and then applied to everybody else. We know that's not necessarily true. We do have to be very mindful, and I believe some of your research is on this about the whole cultural diversity, so look forward to hearing more about it. I want to go back real quick to make this example a little bit more poignant.

I don't know if anybody remembers back in the day when the supercomputer Watson was on Jeopardy and beat the top contestant on Jeopardy. IBM at the time then put Watson through

medical school, went through all the tests of medical school and then side by side against physicians about how Watson could perform versus physicians.

Well, Watson had the opportunity to access this wealth of information, read all the current journal articles, which I know any of us that are in practice or researcher know how daunting of a task even to get a decent percentage of that is. When you put those up against each other, there were some really positive possibilities and potential there. There's some research out of Emory University where they put doctors, and I'm using medical research because right now there's a lot more out about medical research as opposed to psychological research or mental health and wellbeing.

I see them very closely interwoven, but they put in the healthcare system, and they saw that the same processes that physicians were using to make diagnosis AI was, and they came up with very comparable outcomes and output. They also in triage, and we do know there's some problems with biases here, and I hope that we can get better at that, but in triage, again, the ChatGPT was able to match or beat the physicians and the triage, and that idea that we can continue to improve and get more thoughtful approaches, but to that other part in that is we do have to be aware of the risk of bias in artificial intelligence.

Dr. Prinstein: No, thank you. When you talk about the ways in which artificial intelligence can consume information and give us logical decision-making answers, it makes me wonder, Kurt, when we're talking about humans though, we're not talking about beings that are driven by solely logic. There are far more complex ways that we think about human morality, judgment, decision-making. I'm sure AI can memorize and repeat or conclude based on factual knowledge, but how well can AI replicate human morality or emotion?

Dr. Kurt Gray: That's a great question. There's a broader idea of what domains are especially human that has been with us since the advent of Robots and AI, and that domain is shrinking ever more as AI becomes better and better.

It used to be like chess, never chess or never poetry or never go, and every time AI gets in there. Let's say, moral judgments are a fairly easy domain for AI to make sense of because it's fairly formalized. There's been lots of conversations, lots of written about lots of different moral judgments. We have a paper that shows that across 400 plus moral judgment scenarios, the correlation between artificial intelligence, like GPT and 3.5 Turbo and human judgments is like 0.95.

We're working on this project where we hope to show that there'd be some similarities, and then we'd figure out how to improve it. It's already at ceiling. We have some other studies ongoing that shows that it's very good even at explaining those judgments. One thing that I think is challenging is that it's kind of a black box. GBT has, I don't know, millions or even billions of floating points that it uses to make sense of the questions you ask it.

Our human minds, I think are fairly simple, in the sense that there's clear cognitive mechanisms that we use to make judgments, but no one knows what's happening at all inside of these models. I think the challenge is going to be trying to figure out, is it making the same kind of judgments, even if they end up the same. Is it the same roots? Is it a different route? Specifying which models GBT should use to do it to simulate the methods of human cognition, whether that's a morality or anything else.

Dr. Prinstein: Wow. That's fascinating. Sang, what are the ways that we can be using AI, especially based on what we're hearing now, to really improve the way we're doing research, to improve our models or even our measurement tools for how it is that we do research as scientists?

Dr. Sang Eun Woo: Yes. Thank you, Mitch, for the question. This is actually something that I have been thinking about quite a bit for the past few years now. The first time when I heard about all the things that AI and machine learning, big data can do, my immediate thought today is like, "Wow, what can we do with those tools to actually improve what we do in psychology?" I was immediately thinking about those type of technologies as a research tool more than anything because I think sometimes we look at what is happening out there, we immediately go to the applications in the workplace or in people's lives or entertainment, like Go and all of those things.

As a psychology researcher, I think about ways in which our constructs can be improved, or the measurement of the constructs can be improved, by relying on some of those tools that are already out there that actually systematically takes into account the measurement bias. When we think about the bias, for example, we think about how bias can be seeping into our human cognition, social cognition, which is quite honestly what social psychology has been always talking about.

Then psychometrically, when we actually quantify the bias in terms of the algorithmic approaches, we can actually pinpoint the ways in which bias can actually affect the results of our measurement. I think that's probably one of the biggest ways in which bias can be really at the center of our attention, utilizing this AI and machine learning techniques. In addition to the construct measurement, one of the biggest things that I think what AI can bring to our table, AI, machine learning, big data all combined together, is our ability to detect a phenomenon that is not really known to us yet.

When we look at all the data that's out there, all the patterns that are happening, I think with the help of AI tools, we can quantify those phenomenon in a way that is actually empirically testable. If we replicate that over and over, again, with the help of the big data and AI, what we can do is actually doing some of the inductive theory building. We're so used to this hypothetical deductive, top-down theory-driven approach to do our science. Sometimes I think we just need to open up our minds a little bit and let the data speak for itself.

A little bit of the balancing act, I think, is happening with the help of AI. People can actually look at those things that are happening in terms of the data land and do some of the bottom-up approach, more inductive approach to actually create the theory and the knowledge. Those two are the main big-picture questions that I think about. There are some more specific ways in which AI and machine learning tools can be helpful. There are some cool research that's out there in the I-O psychology world. I'm happy to share more of that later.

Dr. Prinstein: Can I just follow up on that? Anyone might have some thoughts on this, but I love these ideas, but I can't help but wonder, based on the comments before, how do we make sure what AI sees as bias is not actually rich, meaningful cultural differences that have been inadequately captured in the knowledge base, or what AI is writing for us isn't the world according to the people who programmed it, the people that have been disproportionately studied in science so far? What do we need to do to fix that before we go too far in potentially the wrong direction?

Dr. Hogg: I would suggest that we are going to continue to get more sophisticated with the artificial intelligence. We have to step back and remember it's a tool and that it's human-designed and human-driven. A lot of people have heard of prompts, prompts being the instructions that you give generative AI to do something. If we can get more detailed and specific into the directions,

whether you want to call them prompts or computer programming, coding, et cetera, we can ask AI to look for those kinds of things.

I think that's a real concern. That's why when we're bringing up the topic about whether we're ready to, in some incidents, replace AI for the actual participant, I'm like, "We're not there yet, but we can glean so much out of that. Then we can ask AI to look specifically towards those things." This is another reason why humans are always going to be really important in the process. That in many ways, AI is giving us more room, reducing our cognitive load for some of those kinds of tasks where we can apply it towards the research.

Dr. Atari: Yes, that's a great point. Maybe I can add a small point to that. In the typical machine learning pipeline, there are three phases in which bias can be baked into our models. The first one is the data, and then the second one is the annotators who annotate the data. The third one is our statistical models that make sense of these patterns. Our data, or like you said, our knowledge base already has a lot of biases that we need to talk about.

People have talked a lot about demographic biases, which are really important, like gender bias or racial bias and other things. One of the things that I want to point our attention to is the cultural bias and linguistic bias that we have in our existing data. More than 95 or 96% of our knowledge base in psychology is from a thin slice of human diversity which is weird people, Western-educated, industrialized, rich and democratic. Weird people make up a tiny slice of human diversity, but they are disproportionately represented in our knowledge base.

If we think that we can discover more about human psychology from this weird database, we are going to get weird patterns, weird theories, just looking at the existing database. I think it is important to also expand our knowledge from real humans, from non-weird cultures, study those populations, and then when we have a more inclusive database, we can definitely do more bottom-up exploratory data analysis for picking up interesting theories that we did not know before.

Dr. Gray: I think it is also possible, there is this huge corpus, as Mohammad pointed out, and it's very biased, but you could specify within that corpus to look at specific things, right? It's not that there's no written documentation or text from specific cultures or subcultures, but you could look within those, specify within GPT that it should be analyzing that.

We have a paper that just came out in JAMA Pediatrics about mirroring what adolescents think about messaging. Adolescents are part of weird culture, but it's still a kind of subsection-bending. You could imagine saying like, what about only 50-year-old Daisy anime fans? That's dumb, but you could imagine bidding in all sorts of different ways. If there's enough data out there, it might be able to do it. Some kind of hope.

Dr. Prinstein: I want to turn to Rose and ask a couple of questions about this brave new world for a second in a moment. In the thinking about transparency and clarity on how science is being done. Rose, the last time I went onto ChatGPT, I asked a couple of psychology questions, and the responses that I got back were what I would give an undergraduate term paper, kind of a B-. It was okay, but it wasn't really a very sophisticated review of what we knew in the literature on those topics.

It sounds like we are very close to the point where people could be writing manuscripts with AI, they could be collecting data with AI, and obviously, it's important that we know when that happens and that people aren't misleading the scientific community. Can you say anything about

just where are we when it comes to publishing psychological science in thinking through AI and the ramifications here?

Dr. Rose Sokol: Yes, thanks. That's a good very broad question because I'm hearing such echoes of transparency through everything the panelists are saying today. It's like if you're not transparent about the samples you work with or the populations you work with, that's going to lead to bias in the outcomes. You're not going to find those people that Kurt are talking about. You have to be transparent at every step in the research process so that the data that goes in is actually valuable data to get back out. We really advocate for people to be just very clear and disclose when they've used generative AI.

There are cases where people might use generative AI in a manuscript and be very clear and just let that come through the peer review process so that people can review it in that light. The same way that you would disclose anything. If you don't include the full sample in your results, what are your exclusion criteria? The same thing with disclosing your use of generative AI and at what point in the manuscript.

We do see problems with the kind of nascent versions that people have access to now. We see things like hallucinations, so fraudulent information. That's probably why you graded it a B-. It probably wasn't all accurate. We see citations are made up so you might get a citation back that doesn't actually exist in the world. I think we should think about bias in that sense too, because the more common the name might be the more common the output of that made-up citation. We're kind of perpetuating biases that already exist in the psychology literature if we rely on these systems. There's also a lot of benefits to these systems as well.

The same systems that are kind of authors have access to publishers are starting to have access to. Can we detect generative AI-created content? It's a mixed bag right now, but it's going to get better as the systems get better, too. The flip side is the concern for editors is how do I know that this was generated by a human? How do I know that these research participants exist in real life? These kinds of questions that publishing industry really has to think through, how we can use the same systems to just be very transparent in the peer review process as well.

Dr. Prinstein: Thank you, Rose. Your response makes me wonder if one day, and this might be two weeks from now or two years from now, if AI could write the intro section to a journal manuscript, should it? I don't know what to say. Should humans spend their time working on the other sections, and should that become standard practice? I know it's very hard to crystal ball this, but you're at the front lines of seeing exactly what's happening before anyone else and with greater scope than probably anyone else. Where do you see us headed?

Dr. Sokol: I think that's a really good question too about what is the point of APA style. You have a publication manual, and people say it's formatting. It's not about the formatting. Generative AI could be a great tool for helping you with formatting, the kind of basic things. How do you write a citation? We already have that. You can go on PsychNet and get the citation in APA style, you can go in Grammarly. Those things exist.

I think it's the more complicated things that you're talking about, like introductions that we can't replace with computers. Because so much goes into writing and communicating clearly, being very transparent about the whole process, which you're going to have more insight into but also writing in a bias-free way. The words that you use matter, the word choice matters, and how you write up your manuscript. I just don't think that generative AI is going to be a replacement for that aspect of how we write and communicate as psychologists.

Dr. Prinstein: A question open up to everybody. I want to go back to this whole concept of AI replacing research participants. Just walk me through this. For anyone. Start at the beginning. How does one do that? Are you typing into ChatGPT? What would a hundred people respond to this question, or are you giving them a delayed discounting task and saying, "Create the variety of responses if this were a performance-based task or an IAT?" How does it work?

Dr. Hogg: Can I jump in and do a little thing like Mohammad just did? [laughs]

Dr. Prinstein: Please.

Dr. Hogg: That is asking the question is why and how do we know when we ask a self-assessment, or we put something into Mechanical Turk or something like that? How do we know that the participants are answering honestly or even spending the time to really absorb the information and responding back to what you're asking? I just wanted to put that out there, that our data currently is not infallible. What kinds of things would we want to be thinking about to try to make AI participants valuable to the research? I have not tried it yet, so I couldn't add more to it than that.

Dr. Woo: I have not used GPT in replacement of human subjects, but I know of several researchers who are trying that out at this moment. I'm really curious to hear from Kurt and Mohammad if you are doing something like that. I'm joining Mitch in asking you how; show us how to do it from start to finish.

Dr. Gray: We use the API, I'm sure Mohammad does as well, right? You basically set up an account within GPT, like a more researcher kind of account. It's not that hard. You provide a prompt where you say like, "What I'm going to do right now is ask you for--" I have the website here if you want to see it. I don't know if it's possible to post into the chat. You basically say like, "I'm going to ask you to provide some ratings. These are the ratings within this bands, and for each thing we give you provide the rating." Then it just gives you the rating.

There's some things you can set. There's something called temperature, which is just how much randomness is in the model. You think of temperature from a gas molecule, the higher temperature, the more the molecule bounces around. You set the temperature low, it'd be less randomness. That's at least how I understand it anyways. Again, correct me if I'm wrong, but it's not so hard. There's lots of different ways you can do it. You can just type it in GPT, but when it's a big set of data, it's easier to use the API.

Dr. Atari: Yes. In my lab, we have also done the same thing as Kurt was saying. You can use the API. There are also a number of open-source language models that are talked about less typically, like LLaMa 2 is one of the more reliable ones and higher quality ones. The pipeline is more or less the same. There's typically an API, and you feed different kinds of assessments. Of course, we are limited to the kinds of assessments that are language-based. You feed these language-based assessments, and you will get responses from GPT. There is a literature on prompt engineering, how to come up with the best prompt to make sure that model is responding in a readable way.

Sometimes for some questions, some of the problems we have had is-- Some of these models sometimes say, "As an AI language model, I cannot answer this question," or things like that. There are ways to get around that. It requires a little bit of experimentation. That's a typical pipeline. There is one big problem that also Kurt talked about, which is the black-box nature of these models. We put something into these models, and we get something out. We don't exactly know what's going on inside these models, which is really important for us because psychologists are typically theory first, mechanism first. Researchers, we really care about mechanisms.

We want to see what's going on in our brain and in our minds. We might not have access to those mechanisms using GPT, LLaMa , Gemini and other language models. You can always inquiry, you can always ask these models, like "Why did you choose this particular option," for example. They can give you explanations, but I think there's a debate in the field whether that's actually mechanism, or it's just a post hoc justification or something that the model learned from existing data out there on the internet. These are good debates that are very recent, and we are still having those discussions in the field.

Dr. Gray: It's also interesting to look at where GPT goes wrong. Just looking at our one data set about moral judgments, there are a couple things where it didn't get, and I think that's instructive to know how humans in our heart of hearts differ from GPT. There was one question about whether it's okay for a coach to get excited and celebrate a goal of this other team, and people are like, "No," and GPT is like, "Yes." Because GPT is like, "Yay. Good for everyone."

In a sense, GPT is maybe less parochial and less groupy than people are. It suggests that maybe GPT is the better angel of our nature in some ways than perhaps people. People are actually more immoral than perhaps GPT, or at least less universal, which gets to the point that Mohammad was making. Morality can be very group-centric, and GPT might be a very broad ethical rules understanding of morality.

Dr. Prinstein: Sang, what would be some of the areas where you think AI would be a great replacement for human participants, and what are some of the areas where you think it would be very bad? [chuckles]

Dr. Woo: I think Mohammad and Kurt convinced me that that's actually a really good area to use AI as a replacement. It makes sense because you are mimicking, or the algorithms are mimicking the human judgment and decision processes that are already well documented and well codified. I think it gets a little bit trickier or a lot trickier if you're looking at an interpersonal dynamics between two people or multiple people.

Because I don't think the technology is quite there yet in terms of the theory of mind type of AI. We don't know what's going on in the algorithmic world personified as a human. The more dynamic and interactive it gets in the nature of the phenomenon that we are trying to understand, the more trickier, I would say.

Dr. Gray: I had the exact same intuition as Sang, and someone else replicated the Milgram study with GPT. It didn't really shock people obviously, but it got GPT to act as a human to shock someone to death. I totally agree, Sang, that these folks are like, "You can do anything," so maybe we should use GPT for the really unethical, terrible things we all want to run but can't.

[laughter]

Dr. Woo: Yes. I actually saw that article, too. It gave me chills.

Dr. Sokol: Kurt, that was one of my broader questions, too. Mohammad, Kurt, you're doing this research, and you're comparing how the machine does against the human. Would you trust the results of a replication? Would that help with the replicability problems in psychology to replicate research using these systems?

Dr. Gray: I would flip it around and say that I might not trust something that didn't replicate on GPT. Because if it didn't, then why didn't it?

Dr. Prinstein: Really? This is so interesting. First of all, let me say that there are a thousand comments in with people asking about some of the researchers that are doing some of this work. We'll put together a list of some readings and suggested readings from the panelists on the website accompanying the recording of this presentation. I wondered, folks, if you have a shout-out right now for people who are very anxious and excited of any authors or platforms or papers that you think would be especially good for people to be looking at right away, for people that are using AI and research, either for replication or to address bias? Anyone you want to give a shout out to right now?

Dr. Atari: I would probably recommend a recent paper published in Nature by Molly Crockett and colleagues. They talk about epistemic and some of the ethical issues associated with AI. There is also a group of researchers who publish the paper in Nature Reviews Psychology. I don't remember the first author's name, but Jamie Pennebaker is on the team. They have some interesting insights about the use of large language models in psychological research, broadly speaking. These are some of my favorite papers. There is more. There are a lot of interesting papers coming out every day. Some by colleagues on this panel, so I will just shout out to people who are not in this panel.

Dr. Gray: Igor Grossmann has a paper in science, too, on *The General Idea*, so that could be another one.

Dr. Hogg: It's ever so slightly different, but I want to talk about Dr. Kerri Lemoie, out of MIT. She's doing a lot of work in digital credentialing. Now, the zeitgeist for it really had to do with educational pieces, but I think that one of the critical things that we need in AI that we have alluded to but haven't said directly, and it's really some of the underlying parts of biases and things, is trust.

It's essential for us to be able to trust our AI systems, and that's something I think that we're continuing to build into it. The goal is to better understand, and in this particular case, when we're talking about psychological well-being and psychological research. The credentials, if I could take it a step back, the idea is, for example, let's say you're going into a bar, and they're insisting on seeing your driver's license. What do they need? What information do they need from you?

They only need a few key pieces. They don't need to know your street address or a whole variety of other things, so that you could just give those couple pieces of information that helps protect your privacy, as well as being able to have resources to say that the credentials for this information are real. She's using it, and the group that she's working with is using a lot in education where you could then verifiable that you had a degree or et cetera. I think that it can move into the AI space when we have more and more data that has been credentialed.

Dr. Prinstein: Great points. There are a number of questions in the chat, and I want to make sure that we get a chance to get to those. I also wanted to ask, one of them pertains, there are a few about limitations that you've discussed already, but you've talked about being able to segment the knowledge base within AI to be able to give you some information. Of course, there are new constructs that we don't have a knowledge base on how humans interact with virtual reality or on social media where we don't have quite as much information.

Is it a leap too far to use existing knowledge that we have to intuit how humans would respond to new stimuli or a new context, or do you draw a line and say, "This is a place where we really can't rely on AI. There's just not enough knowledge yet." Similarly, how about if you're trying to understand the responses of somebody experiencing severe mental illness perhaps. Do we have that kind of level knowledge for AI to be able to simulate the responses of somebody who's coming from a clinical population, for instance? Any thoughts or reactions?

Dr. Hogg: I have worked with several researchers that are doing work in immersive reality that have created some applications in VR to simulate what it might feel in certain situations to have autism. I think that's getting there, and there is some research now out in this area, but I don't know that we're quite to the place where you're talking about it.

Dr. Prinstein: Okay.

Dr. Gray: I think Sang mentioned this, right? Using this inductively to make the best guess to a new population. I think it's probably better than many of our intuitions, even if it's not perfect.

Dr. Woo: Yes, I agree. I personally have no knowledge of any studies that are specifically on the mental health-related populations, the clinical populations or situations where the human-computer interactions are utilizing the AI as the human. I think that just seems so far out at this point. Mitch, I think you're right. Maybe it's just a matter of time and more intentional efforts going into that direction. I personally have not heard of any studies.

Dr. Atari: In theoretical computer science, there is this question that can AI potentially extrapolate to something that it does not have access to? AI currently has access to a particular cognitive domain, can it extrapolate to some cognitive aspects that it currently does not have direct access to? That's an open question. There are different theories about that, and it's not within the scope of this conversation to talk about that.

One of the things that I'm a little concerned about is telling these models that you are interested in a particular slice of the population or a particular demographic might be akin to, might be similar to asking an adult to tell you about mental health or development as if they are a teenager. Is that valid? Again, this field is very new, so we can have discussions about that. It's just a concern that I wanted to voice. It is unclear whether we are actually limiting these language models to act and respond like a teenager, or we are just masking these models to mimic adolescent behavior.

Dr. Gray: I think that's a great point. You can imagine there's lots of memoirs from people who suffer schizophrenia or bipolar disorder. I wonder if I could mind there. I guess at the end of the day, it's an empirical question, is it just parroting back stereotypes, or is it actually getting to the meat of the question? I guess that's what we can use data to figure out, right?

Dr. Prinstein: Another theme of-- Oh please.

Dr. Woo: Sorry, Mitch. so of course the issue is the data bias. Not just the bias in the algorithmic machine learning training bias or the coder bias even. It's a data bias because we don't have access to those type of data. Should we have access to those type of data in the first place? Going back to Mohammad's question, just because we can, should we?

Dr. Prinstein: That's a great question. I am suddenly flashing back to 20 years ago when people started using missing data procedures. Can we use the information we have from the data that are available to assume we know how someone else might have responded in that same situation? We've come a long way in being able to accept that, but it's still based on some fundamental assumptions that need to continually be questioned, which I think is maybe sounding somewhat parallel to what we're thinking now with AI.

There are a couple questions coming out about ethics. It sounds like Mohammad, Kurt, you've been doing research specifically using AI as a data collection source. People want to know what is the purpose of IRB if these are not human research participants, and what are the challenges or

questions that are involved in making that ethically appropriate. Rose, also some questions about, again, what are some of the ways that you've seen that being discussed, and how are reviewers responding to that? Are you noticing an uptick in people accepting those kinds of procedures? Any thoughts?

Dr. Gray: I'll say quickly that there's no IRB, and that's why it's amazing because you're just getting the queries back. You want to do the totally unethical study you can't do with undergrads/ Boom, GPT.

Dr. Atari: There is no need for IRB for these kinds of analyses because they are not human subject studies. Institutions I think are still figuring this out because it's a new thing, like Rose was saying the how APA as an institution is trying to understand some of these aspects and try to use AI for new purposes. Universities are no exception I think. Institutions within universities are also trying to see how they can navigate the new technology.

Yes, there is no need for IRB, but in terms of other ethical risks, for example, gender biases or racial biases that we have in these models, this is something that has been an ongoing debate within the computer science community and LP community in particular. These models that are out right now like GPT or LLaMa or Gemini, they are de-bias. They are supposed to not have any biases towards a particular demographic.

There is also this other literature called-- I forgot what it's called, but a bunch of researchers are trying to break these systems out of their de-bias mode and try to show their biases. There's research showing their gender bias, their age-based bias and their white-oriented bias as well. There are these ethical risks. Maybe they are not exactly directly relevant to IRB, but these are things that we should be aware of.

Dr. Gray: To build off that point, too, these models are not just a readout of human text. They're made by programmers to potentially show bias in a particular way. Mohammad's pointed about culture, also politics. A lot of these programmers are more progressive. GPT won't say anti-woke things as much, but it will say more anti-conservative things. Even then, there's all sorts of examples of showing how people beat the guardrails on it.

Get it to share a recipe for making methamphetamine, show a picture of planes crashing into the Twin Towers. It doesn't normally do it, but you can entice it to do it. For instance, one person said that they threatened self-harm unless it gave them the recipe for meth or something, and then it did. It is just ways in which you can interact with it in a certain way and break the guardrails. I don't know what that means for our research, but it's certainly interesting to think about.

Dr. Sokol: Mitch, I think your point about the publishing perspective, usually, we're very practical in publishing, but this whole conversation is all about agency. From the authorship perspective, somebody asks you, can AI author a book? It can, but will we publish it? No, we won't publish something that isn't written by a human being who can take responsibility for coming back to us and saying we need to correct something that we published that's not right. All of those things that happen throughout the publication life.

We would say that an AI system can't have that responsibility as an author. That's the question I think too with the research, are we publishing research in this space? For sure. I think you have to, it's a new methodology, and people are going to be excited to read about that and learn about that, but at some point, the question is how far can we take the agency of these systems? I don't know the answer to that.

Dr. Prinstein: Sang.

Dr. Woo: Thank you. Can I ask a follow-up question, Rose, because I think this is very interesting because will AI or should AI writers replace human writers? Absolutely not, but I think what we're seeing is the augmentation. Even Grammarly and all these other AI-enabled writing assistants are in use. It's a matter of the degree. I guess my question, hopefully on behalf of the audience, and myself obviously, is how far can it go? How far can we stretch it? Because even when the whole thing is initially written by AI, still the human author will look through everything, give some final touches, will that be acceptable?

Dr. Sokol: The copyright office in the US said no. There were the AI-generated images, and the author tried to copyright them, and they said, "You can't copyright something that wasn't created by a human, essentially." I guess that's what I'm telling you APA is saying that we're not going to publish something that says written by Grammarly or written by chat GPT, but I have seen publishers publish books that say just that. It's different across the industry. I think people are exploring in different ways. For us, it's really about that human accountability.

Dr. Prinstein: Was there a discussion among editors or reviewers that you've heard of of people saying they're okay with this description of this measure was written by AI or anything? What are you seeing there?

Dr. Sokol: I don't think we've seen a lot of examples where people have disclosed that, but that's a good question that relates back to plagiarism detection. We use AI systems for plagiarism detection just as you would in an institution. When you get back, and it's all in the methods section, and it's this one methodology, and there's not too many ways to describe it, usually, an editor or reviewer will say that's okay. That's an okay use of reusing your own words from the past. If it's about the deeper ideas and the introduction or the results, the discussion, that's not going to be okay. I think it's that kind of fine-tuning too with where you use the ChatGPT or the different systems in the writing process.

Dr. Woo: I'm going to jump in here because personality measurement has been revolutionized by AI item writers. It's not happening across the board, but it is shown to have the acceptable level of reliability and validity. We don't have to get into all these iterative writing, generative item writing processes as traditionally endorsed by the traditional psychometric anymore.

That's where I see that gray area where AI is supporting and assisting the research process but not entirely taking over. Again, intellectual property, is that really an issue here? If it is, how do we make that as a policy-level structure or some sort of a safeguide-- What's the good word? Safeguard against abusing this in place of the human intellect.

Dr. Sokol: I don't think I'm even going to attempt to answer that if that was for me. That's a good question, and that's the whole point of the panel, right?

Dr. Prinstein: Yes.

Dr. Sokol: It's like, "We don't know yet. It's so new."

Dr. Prinstein: I'd like to read you a comment in our listeners' Q&A and ask you to react to it. How can we possibly replicate the ambiguities and paradoxes of human experience with virtually every emotion and emotional investment? My main concern with AI is that it doesn't experience anxiety and therefore a more vulnerable and multifaceted experience of life. Also, it seems very dubious

that AI could ever perceive life with the awe, wonder, veneration and dread and thus stakes that humans have in our relationships to people in life. An algorithmic aggregation cannot possibly replicate that risk and unsettlement, which also intensifies the amazement and appreciation for our human experience.

Dr. Gray: I think that's a good point. I think my every day doesn't feel like that. My every day I feel a little more like a robot, mindlessly going about my days. I think, at least empirically speaking, AI is able to capture our judgments because there's just so much out there. Even that text, that's perfect, that text about what it's like to be a person is probably searchable now by GPT and will be included in the board by GPT. GPT will use that amazing description of what it's like to be human in making human-like judgments. It seems empirically to capture it okay.

Dr. Prinstein: We're about out of time. I wanted to ask you for final thoughts on a related, but slightly, I'm going to reverse your perspective a little bit and ask you, if you had the opportunity to meet with the leaders of AI, the people who are creating it, generating and guiding the future of AI, what do you think they need to know about psychological science to make AI better serve their purposes for which it is designed? Final thoughts on how you think psychological science could be used to improve AI.

Dr. Hogg: Oh, you took a tough one for the last one, didn't you?

Dr. Prinstein: I'm sorry. [laughs]

Dr. Hogg: I'll go. I don't know that I'll be brilliant with this. I do think that developers in general, and I'm going to generalize, technology developers in general, often think first, "Oh, that's cool. Let's see if we can do it." When they do it, "Hey, what else can we do with it," as opposed to thinking about the ramifications, the psychological impact. It's often hard, once it's out of the gate, to pull it back in.

It'd be great if we could be involved in the creation phase. I'm not sure we're past that, so that idea that how can they design it so it isn't a tool to be used ethically in powerful ways to better understand ourselves and to better support well-being.

Dr. Prinstein: Thank you.

Dr. Gray: I also think computer scientists should read our work. It's up to us to get it. When I work with computer scientists, they have no idea about how the mind works. They're generally not amazing at theory of mind. They don't have great intuitive models of other human behavior. I think by getting our work to computer scientists, it will help them make models that better approximate how we think and feel.

Dr. Prinstein: Give me another example. What do you want them to know about our work? Theory of mind? What's something else you would want a computer scientist to know? Anything

Dr. Gray: Cognition, the nature of categorization, how people feel emotion, maybe psychopathologies, the structure of psychopathology, the work that we all do or that you apply every day, just getting it out there. Even popular psychology books about how the brain or mind works, that's a good start.

Dr. Prinstein: Great.

Dr. Atari: Speaking with the leaders of AI, I would probably tell them that there are more humans than English-speaking and weird people. I would also give them a gift, and that gift is a book called *The WEIRDest People in the World* by Joe Henrich. I will ask them to read that, which is exactly what Kurt was talking about, so they know more about how the human mind works and how much cultural, psychological diversity we have around the globe.

Dr. Woo: I echo all those suggestions. If I may add one more, it would be the psychological measurement principles, how the psychological constructs are measured and assessed in a way that is actually reliable and valid that actually helps the theory. I think sometimes the measurement piece is really the first step towards creating a reliable tools for us to use to go further, to investigate the psychological phenomenon. If the measurement is not done in a way that makes sense theoretically, we're not talking about the same construct in the first place.

Dr. Sokol: I think I'd take advantage to first ask to respect copyright and confidentiality in the systems and the creations and training of the systems. Beyond that, Kurt was talking about infusing psych in the beginning, and I think infuse psychologists at each step at each iteration. Slow down. Let psychologists play in the system and point out these biases and point out these challenges. You need to know your assumptions going into a system, and psychologists can really help figure that out. What could use improving in the next iteration to just keep making it better instead of just keep making it faster.

Dr. Prinstein: Thank you all so much today for an amazing discussion and thanks for your participation. To everyone that's listening, I wanted to mention that there will be a one-minute survey after the broadcast for your feedback. We also would love to hear from everybody on topics for our Essential Science Conversation series or any other thoughts that you have, at science@apa.org.

Last, we invite you to subscribe to *Science Spotlight*, your source for the most relevant news, information for psychological scientists by psychological scientists. That's free for everyone. You don't have to be a psychologist, an APA member or even a human to subscribe to that newsletter. Thank you so much. We hope to see you at future events. Thanks for sharing your feedback. Take care.