Transcript Excerpt
Higher Education & Generative AI: Evolving lessons from the field
New America Webinar
Thu, Apr 20, 2023
Learn more & watch the video

Question: *How should we interact with generative AI, as individuals, as educators and as a society?*

Meredith Broussard
In the classroom, we need to clarify for our students what AI is and isn't. That's where all of our conversations need to start.

AI is math. It's very complicated, beautiful math. But it is just that – math. Real AI is not what we see on screens, coming out of Hollywood. Real AI is not fantasies of artificial general intelligence. Today's generative AI systems are computer systems that work in specific ways. Generally, you feed a pile of data into a computer, the computer makes a model of the mathematical patterns in the data, and then you can use that model to make new decisions. It's complicated, but it is possible to understand the process. So we need to start with a shared understanding of the real and what's imaginary.

We also need to think about the current moment as part of a larger history of technology. We're in a hype cycle right now around AI, especially around generative AI, and people are saying things like "oh, it's going to change everything. It's going to make new jobs. It's going to eliminate jobs. It's going to change education forever." It is going to change a couple of things. But this is not the invention of fire. It's just another AI program.

The hype cycle does not have room for ambiguity, and ambiguity and bias and social problems are everywhere when it comes to AI. Social problems like structural discrimination, racism, sexism, ableism– all of these problems are reflected in the data that we're using to train AI systems. Then these problems get crystalized in code, and become very hard to see and hard to eliminate. Which is why we need more investment in the work of algorithmic accountability reporters. Algorithmic accountability journalism is a newer kind of journalism that is about holding algorithms and their makers accountable.

In terms of what people should be thinking about for syllabi, I would be remiss if I didn't suggest my own books, which explain AI in plain language and introduce some of the complex social issues we're discussing in the public interest technology space. Those titles are *More Than A Glitch: Confronting Race, Gender, and Ability Bias in Tech* and *Artificial Unintelligence: How Computers Misunderstand the World.* In addition, the first place that I always send people is to the resource list at the Center for Critical Race and Digital Studies.

I also really like *The Tech That Comes Next: How Changemakers, Philanthropists, and Technologists Can Build an Equitable World* by Amy Sample Ward and Afua Bruce. This PIT reading would pair nicely with a generative AI activity to get students thinking about larger ethical implications aside from the ubiquitous narrative of "this is new technology, isn't it so exciting?"

Todd Richmond
I take a little bit of a different viewpoint than Meredith. I do think this is a sea change, just like the internet was a sea change for humanity. And I think it's time for a fundamental rethink of what constitutes a human endeavor.

If education is supposed to prepare our students for the future, we need to understand what the future is going to look like, and what metrics for success are going to look like. How do we prepare them for that world? Six months ago, we were not having these conversations about generative AI, or at least not to the level that we are having them now. You can attribute some of that to the hype cycle, but you can also attribute it to the fact that the rate of change is astronomical.

Back in January, we did a quick and dirty task analysis with three of my graduate students to assess how generative AI could be used in their work. Our scorecard was that 97 percent of the tasks that our graduate students do can have some generative AI aspect. Some of them are using it to summarize dense academic articles they read for classes.

I think about *mechanical*, *operational*, and *conceptual* tasks. ChatGPT is pretty darn good at some mechanical and operational things. It can do calculus. The conceptual part, however, is a weakness. I describe generative AI s a food processor that has access to every vegetable. You ask it to make pico de gallo, it'll make pico de gallo. But you don't know who's recipe it pulled, and it doesn't know what pico de gallo tastes like – but it knows how other people have described it.

With this in mind, do students need to know higher math? My economics colleague would argue that yes, they need to know calculus. But again, is that just a mechanical skill they need, or one that helps them develop a conceptual skill? You can go back to the arguments around calculators in the classroom, but this is a much bigger fish to fry.

Finally, we have writing. Is writing, and by extension reading, going away? That's a question that we have to ask, especially given the fact that I have very smart grad students who are very skilled, who are using ChatGPT to do some of their reading. This is the really important philosophical question: what are human endeavors? We have seen that as technologies come in, if they are compelling and convenient, they will replace what came before it. So this is an opportunity to really think deeply about "what, how and why" humans do the things that they do.

Vanessa Parli
At Stanford's Institute for Human Centered Artificial Intelligence, we believe that interdisciplinary collaboration is essential in ensuring these technologies benefit all of us. That interdisciplinary mindset is reflected in our faculty leadership that comes from medicine, science, engineering, humanities, social sciences. It is also reflected in all of our programming. The annual AI Index, a report covering many topics summarizing the state of AI, is made up of a steering committee of experts from academia, industry and government.

While the report is over 300 pages, I will provide some report highlights to perhaps spark interest. The majority of AI systems are developed in the U.S., Canada, the EU and China. What might this mean for the rest of the world? Certain values, cultures, norms are embedded within these technologies, and then they're distributed across the world, where not everyone has the same culture, norms, etc. Do we want to be doing that? How can we build these systems so that norms can be adjusted or modified, depending on where you are in the world? That is one of the many reasons we need diverse perspectives participating in all phases of development of these technologies.

Computer Science PhD students are gradually becoming more diverse, and undergraduate students even more so, however I would personally say that this is not good enough, and we still have a long way to go. The portion of new women in AI PhD programs has remained at around 20 percent, and women are making up an increasingly greater portion of computer science faculty. But again, I personally don't think this is good enough.

Question: *What should an equitable and ethical approach to generative AI look like? And what should governance of this technology look like?*

Meredith Broussard
One of the things that gets lost in the hype around generative AI is the incredibly toxic nature of the training data used to create these systems. In this case, it's data scraped from the open web, which has a lot of really wonderful stuff and a lot of really toxic stuff.

A recent analysis by the Washington Post looks at what, specifically, are the sites that make up the Common Crawl, which is the data set that is used to feed ChatGPT and other big generative AI systems. They're all drinking from the same well. There aren't that many places to get massive data sets, and everybody's pretty much using the same stuff.

Generative AI systems that use Common Crawl are being fed with text from 4Chan, with data scraped from StormFront, and other sites that publish hate speech. There's a lot of toxic material in the training data, and a lot of copyrighted material. You really need to be careful about trusting these generative AI systems.

I think that ethical use of generative AI starts with emphasizing that this is just a tool. It has zero foundation in truth. It is given to hallucinations; when a generative writes something, there is no guarantee that it is generating information that is true. It makes up citations, for example. So even though it looks like it's working, it's not necessarily working the way a student needs it to in order to learn class material. People have a lot of trouble with that. If you use generative AI to summarize a complex scholarly article, great– that's a good way to get started on a challenging task. But you should still go back and read the original scholarly article, because there is no guarantee that the summary is accurate.

Todd Richmond
I think "entirely" is the key word. I don't think that everything that ChatGPT spits out is nonsense, because you can fact-check it. You can't trace back to see where the original sources were, but the graduate student who used it for his reading did a sanity check on it, and found that it was a pretty accurate distillation of the original source material.

One area where we need to think through equity and ethics is intellectual property and fair use. Someone recently created a fake Drake song using generative AI that went viral. Almost all of these generative AI companies are making the same claims that the search engine companies made, which is "well, the stuff on the Internet is out there. It's free to use, for anything we want to use it for." For search engines, it was one thing to index it. It's a completely different thing to use that information to build an algorithm which will now essentially create derivative work of the original work. That's much more problematic.

There needs to be an entire new field (here's a framework we use at Pardee RAND) focused on creating emerging technology that has equity and ethics at its core, instead of toxicity and disinformation.

Vanessa Parli:
I would add that ethics needs to start from the very beginning. Computer science graduates working on these systems also need to be trained in ethics. At Stanford there's a program called Embedded Ethics, where all the computer science students in their core courses are taught ethics modules in the hope that that impacts their thinking as they go on and develop these technologies.

For the grant funding we do at HAI, applicants need to write an ethics statement as part of their application. They have to show that they're thinking about the ethical implications, the societal implications. If this technology were to be ubiquitous, how might it play out and what adjustments are made in the research approach to be sure outcomes are positive? Those statements are reviewed by an interdisciplinary panel of experts again from medicine, philosophy, computer science, etc. and a lot of times, there is iteration on the research methodology. Sometimes it's decided that part of the research maybe should not go forward.

Question: *What do you think a strong governance structure for generative AI should look like?*

Meredith Broussard
It's important to keep in mind the context where AI is used. Let's take facial recognition AI. A low-risk use might be using facial recognition to unlock your phone. A high-risk use might be law enforcement using facial recognition on real-time video feeds as part of surveillance, because it's going to misidentify people with darker skin, thereby contributing to harassment and over-policing. The context is key.

I would really like to see more algorithmic auditing. Algorithmic auditing is something that we talk a lot about in PIT circles. It's the work of opening up black boxes, interrogating algorithms to look at where the biases are. All you have to do is look for biases, and you'll find them. I would love to see algorithmic auditing integrated into regular ethics reviews, and to have ongoing monitoring of technology as new iterations are rolled out. We need to monitor the technologies to make sure not only that the technologies are not biased to begin with, but that the bias that is mathematically mediated is then not added back in in future iterations. So that's a kind of technical feature of algorithmic governance that is going to help in implementing high-level policies.

Todd Richmond
The key question is, can you trace it back? Can you open up the black box, and can you start to trace how the algorithms are working? Because when you have algorithms rewriting themselves, the people who set them into motion don't really know what's going on under the hood.

We're big fans of red teaming, and I was heartened to see that Open AI is doing red teaming on their algorithms. The red team process is usually done to try and figure out where your vulnerabilities are, and commercial companies do it for a competitive advantage.

A few years ago, we started arguing for narrative red teaming. LA City announced that they were going to release all their 311 data to the public as part of a transparency effort, and I was immediately horrified because the 311 data, taken out of context, allows you to construct very toxic narratives when you combine it with demographic census data. We have to think about how data might be weaponized. With narrative red-teaming, if you're going to release a report or you're going to release data, you work through how people could weaponize that data, and then prepare messaging in advance, and maybe conduct narrative inoculation. We have a grad student working on this with Russian troll posts and how you inoculate against disinformation campaigns.

For the governance piece, the thing that I find most challenging is that most of us are sitting in the U.S., and we have a very U.S.-centric viewpoint of this. This is a global phenomenon. At Pardee RAND, we do a lot of national security work, where we worry about adversaries. Not all of the countries that are developing these technologies have the same moral compass that we

have. There's an asymmetric governance problem. It's great if we want to pause AI development, but the problem is that other people are not going to pause it. It's a very messy, complex global problem set when you talk about governance for these emerging technologies.

Vanessa Parli
There's also this aspect of developing community norms. These technologies are moving so fast. Our governments don't move so fast.What do we as a community of researchers and computer scientists want for this technology? There's the example of CRISPR, where once those in the development phase of that technology realized its impact, they developed their own community norms which had nothing to do with the federal government. When we develop these technologies, what is appropriate to release? What types of documentation should be released with these models?

Todd Richmond: One thing I just want to tag on is that we should be very careful when we say "community". We need to cast a wide net for our stakeholders that includes artists. My wife is faculty at Cal Arts, and I've spoken to her students about generative image algorithms. They stand to lose a lot in this, and that community of practice needs to not just be the computer scientist and the technical folks, but it needs to bring in the arts and the humanities because they are very real stakeholders in this equation.

Audience Question:
Some of the systemic problems precipitated by the Internet were unforeseen. What systemic problems do you anticipate? For example, if we all use ChatGPT for intellectual work?

Todd Richmond
Plagiarism is nothing new, and what we're seeing in schools is a supersonic version of plagiarism. Humans do what humans do. So we can look to the past to see how humans behave badly. It's just that. The problem is that digital technology scales in a way we've never seen before. So the speed at which those problems propagate and the scope of those problems is drastically different now. It's going to be faster, and it's going to be on a bigger scale.

Meredith Broussard
I don't know if I agree with the idea that it's scaling in a way we've never seen before. We've been doing this for 30 years now. Digital technology scales, yes– it's not dramatic, it's not new.

But in terms of plagiarism, you're absolutely right, plagiarism is nothing new. I've seen some interesting work about how to inoculate students against cheating with ChatGPT, like creating iterative assignments. So, you can do in-class writing, and have assignments that build on the work done previously. That's a little bit more challenging.

Another assignment that I've seen a lot is when instructors have the students use ChatGPT and then critique the output. That's been a really useful exercise for critical thinking around technology. I don't think we can eliminate cheating entirely.

We *can* reevaluate what we are trying to do by asking students to have closed book exams. Are we requiring them to memorize something for the sake of memorizing it? Can we design experiments and exams that are, say, open book? Can we design assignments that acknowledge that there is generative AI out there and that the students use it?

Audience Question
What would need to happen to guarantee transparency around the data sources being fed into AI?

Vanessa Parli
I won't comment on what needs to be done, but I do want to point out a resource that Stanford's Center for Research on Foundation Models has recently developed, called ecosystem graphs, where they try to map each of these different generative AI systems to find where is the data coming from, what system is built upon, so that you can can see what's going on, and perhaps better identify where some of the bias, etc. might be.

Meredith Broussard
If you want to know what's in GPT3, you can go and read the academic papers. What you'll find there is that this is trained on common crawl, and for its self-censorship it's using Real Toxicity Prompts to find bad words in the data set. The information is not super secret. People like to pretend that it's super secret, but it's not all that secret.

Todd Richmond
Well, the weighting is the secret sauce.

Meredith Broussard
The weighting is secret. But most people are not messing around with weights, most people are just interested in what generative AI is being trained on. The fact that it's being trained on Reddit data is really helpful to know, because you can look at Reddit, and you can say, "oh that's a cesspool. It's pretty interesting, but it's also a cesspool. So maybe this generative AI is going to spew some filth that I do not agree with." That's the level of transparency that I think a lot of people would be pretty happy with.

Audience Question
Apart from summarizing articles, are students and faculty in your programs using ChaGPT in other ways to engage in learning, for example for running experiments?

Todd Richmond:
Not only our students, but our researchers are using it to write code. It is very good at Python, and if it works, if it's good enough, then it will get adopted. Some students are using it to do literature reviews. Sometimes it's good at lit reviews, and sometimes it's not good. They're using it to do outlining for their dissertations, to help them think through how they sequence things. Some of them are using it for summation analysis. You can put text in and ask it, “What are the takeaways?” And if ChatGPT doesn't synthesize your prose the way you expect,, then maybe you didn't write very clearly in the first place. So it's kind of like having another writer to check your work.

We've also been connecting it to agent-based modeling systems. You can run really interesting experiments with agent-based models that are driven by ChatGPT inputs and outputs. So there's a lot of really interesting stuff that can be done, and we're just scratching the surface at this point.

Audience Question
How do you think AI can help society as opposed to harm society?

Todd Richmond
It holds the promise of doing tasks that are boring, rote, and not stimulating, and giving humans more free time. That said, I have yet to see “more free time” as an outcome of any technology advancement in the past.

We're seeing amazing capabilities. When I was a biochemist, I did protein structure function. We wished that we could imagine, given a DNA sequence, what a protein structure looked like. AI is solving those problems. Is it 100 percent correct? No. But it's figured out thousands of structures, and it gives us a starting point where the humans can come in and do really interesting work that builds upon that. It has the power to do a lot of stuff that we wish we were able to do, but don’t have the time or don't have the patience to do.

The challenge will be, is it accurate? Is it equitable? And is it good enough for the humans to then use and build on? That's an open question.

Meredith Broussard
I’ll push back on that notion a little bit. We are about 30 years into the technological era, so we need to add nuance to our assumptions about technology. I wouldn't assume that there is a “thing” about AI that is going to be helpful for a society. I would not assume that AI is going to be all good or all bad. I would just encourage people to add nuance to your understanding of it. It's not about binaries anymore. And in general, I'm really optimistic about the field of public interest technology as a way of helping people understand all of the nuances and all of the potential implications of new technologies.

Vanessa Parli
There is a lot of promise, but we don't know what we don't know. We really need to think about how we want to use these tools. What are humans not good at, that maybe the tech is better? And vice versa, what are humans better at? How do we want to pair up and use these tools to create exciting work and opportunities? We should be thinking in that way especially since there's a lot of hype out there, as Meredith said.