June 21, 2023
[Sonia Fernandez](#)

Researchers at the 2023 Mellichamp Mind & Machine Intelligence Summit contemplate and discuss the state of artificial intelligence

With the rise of artificial intelligence in our daily lives (think search engines, e-commerce and ChatGPT), it's high time we ask ourselves some of the hard questions: How is AI changing how we work? How do we deal with one another? And even how do we think?

Which is why some of the finest minds in the world of human cognition and AI convened at UC Santa Barbara for the 2023 Mellichamp Mind and Machine Intelligence Annual Summit to discuss the impact of artificial intelligence on our lives, as well as the ways it could benefit or hinder society.

"The idea of human-centered AI is to try to design it so it actually potentiates humans rather than hinders them," said [Miguel Eckstein](#), a UCSB professor of psychology who with computer science professor [William Wang](#) directs the interdisciplinary Mind and Machine Intelligence initiative.

The symposium covered a wide range of topics, from issues of bias and lack of transparency, to how AI can better reflect human priorities and generate beneficial predictions. Artificial intelligence, according to the speakers, is ultimately a tool,

albeit one whose power and mystery has led to overreliance in some cases, and in others, fear and suspicion.

The different tools we use change the way we think," Danny Oppenheimer, a Carnegie Mellon University psychology professor, on the effects of outsourcing human cognition to external tools. While AI could ease the cognitive burden of some tasks, he noted, it also has the potential to dampen human ability to store and retrieve information, and to impede development of diverse and innovative solutions and options.

But AI's ability to sift through vast amounts of complex, variable data also gives humans the ability to see wider patterns, allowing, for instance, the detection and description of corruption in government, or racial bias in policing. The challenge, according to the researchers, is to create robust and reliable models and algorithms that more closely and realistically match human priorities and preferences — such as in the case of recommendation systems and search engines — as well as the values of society in general, for more fair and equitable outcomes.

"We need to build better user models so we can change how the algorithms are fundamentally created," said University of Chicago computation and behavioral science professor Sendhil Mullainathan in his keynote address.

Human expertise and artificial intelligence's analytical power would make a formidable problem-solving team, benefiting decision makers in areas such as business and medicine. Developing this type of close collaboration requires accurate assumptions about what both AI and humans can and can't do, relative to each other.

In the absence of knowledge about performance or accuracy, said UC Irvine cognitive scientist Mark Steyvers, people tend to have high expectations of AI. "People believe that AI can do everything, (and) is great at all tasks," he said, independent of how easy or difficult the task is perceived to be. In contrast, he added, people assume the ease or difficulty of a task for another person based on their own experience with the task. Other obstacles to a fruitful hybrid system of decision making may exist where there is a wide discrepancy in the accuracy of the model and human. In those situations, it can become difficult for a joint human/AI system to exceed the performance of the more accurate agent.

But how about the hidden and not-so-hidden pitfalls of interacting with machines and artificial intelligence? It's no secret that malicious actors often use technology to provoke other people into believing untruths and errors, which causes them to act on those unwarranted beliefs with damaging results. What you see is often not what you get, said S. Shyam Sundar, the Jimirro Professor of Media Effects at Penn State University. The human tendency to believe what one sees is a heuristic — a cognitive shortcut — that unscrupulous people can use to bypass our truth filters.

"The scourge of deepfake right now is so severe, that we're really worried about how people might be falling for this heuristic, because nobody wants to disbelieve their eyes," he said. Coupled with our tendency to cast machines in an overwhelmingly positive light — they are "more objective" or "more precise" or "more secure" — these shortcuts and other human cognitive tendencies can lead us to over-trust our machines, according to Sundar, who discussed the calibration of user trust in AI.

Meanwhile, UCSB computer scientist Ambuj Singh examined the more adversarial aspect of artificial intelligence in his discussion of how AI might hurt group decision making. It's an effort to anticipate problems as modern society outsources more of its security and high stakes decisions to artificial intelligence.

"If we could build these models it is possible that an AI agent can attack the decision-making of human-AI groups," he said, noting behaviors that humans tend toward in group decision-making settings that can be gamed by stealthy AI. Singh's presentation discussed two human group decision making tendencies: that people tend to reach consensus when getting together to decide what to do, and also that a person who might have made a low-risk decision on an issue independently will likely make a higher-risk decision on the topic when in a group (i.e. lower sensitivity to risk). Singh found that it was possible for an adversarial AI to skew decisions without interrupting areas where the humans reached consensus, thus bypassing detection. Such behaviors, he found, can be generated by an AI agent that observed the group and had prior knowledge of the individuals' behaviors. This field of "evasion attacks" is fertile ground for further research in the realm of human-AI decision-making.

"While we are looking at questions of how AI can improve a group, we should also look at how AI agents can undermine group behavior," he said.

Media Contact

Sonia Fernandez

Senior Science Writer

(805) 893-4765

sonia.fernandez@ucsb.edu

About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.