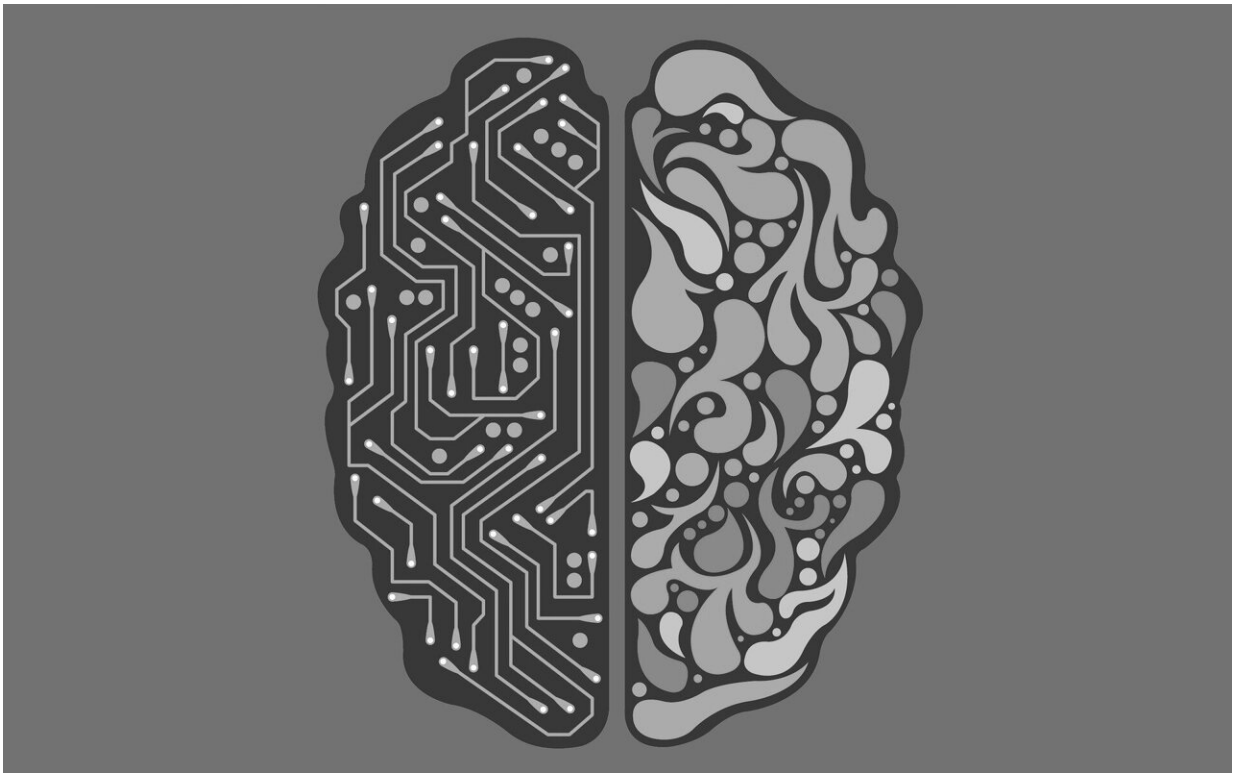
Researchers aim to bring humans back into the loop as AI use and misuse rises

February 20 2023



Credit: Pixabay/CC0 Public Domain

Artificial intelligence is dominating headlines—enabling new innovations that drive business performance—yet the negative implications for society are an afterthought.

How can humans get back into the loop in the quest toward a better society for all?

A trans-Atlantic team of researchers, including two from the University of Michigan, has reviewed information systems research on what's known as the "Fourth Industrial Revolution" and found an overwhelming focus on technology-enabled business benefits.

The focus means far less attention is being paid to societal implications—what the researchers refer to as "the increasing risk and damage to humans."

"We're talking about AI the wrong way—focusing on technology not people—moving us away from the things we want, such as better medications, elder care and safety regulations, and toward the things we don't, like harmful deepfakes, job losses and biased decision making," said Nigel Melville, associate professor of technology and operations at U-M's Ross School of Business and design science program director.

"Our new framework is a theory-based attempt to go in a new direction by recentering humans in actions and outcomes in the discourse about ever-smarter machines."

Melville co-authored the study, which appears in the *Information Systems Journal*, with Lionel Robert, professor at the U-M School of Information, and Xiao Xiao, associate professor at the Copenhagen Business School.

Taking ChatGPT as an example, the researchers say developers likely never intended to enable student cheating or AI-generated recommendation letters. But it's clear they didn't fully consider the social implications of their AI chatbot, focusing instead on benefits such as reduced operational costs.

Their study aims to help society, including regulators considering AI safety regulations and organizations considering adopting AI, understand the societal implications of ever-smarter machines.

For example, the results suggest a shift in the way that policy is made. Right now, lawmakers are reactively considering regulations about technologies that are already entering the market, so laws tend to adopt a narrow view of AI. Instead, the study suggests they need to turn their attention to the bigger picture and write proactive laws that address four emerging machine capabilities:

- Decision-making: Where do we draw the line on machine decision-making, how do we move that line as machines advance and how do we proactively manage bias and other forms of invalid decisions?
- Creativity Automation: How will we manage job losses in creative fields such as visual illustration and music, who owns the rights to AI-generated work and what to do about perfect and cheap deep fakes of anyone?
- Machine-human relationships: As humans increasingly interact with machines with human-like capabilities, how will workers react to AI bosses, how might "AI friends" influence democracy and how might human relationships themselves change over time as a result?
- Machines teaming with other machines: How do we manage exponential capabilities of intermachine teaming, craft rules about machine interaction to balance positive outcomes with risk and determine how much autonomy is appropriate when considering the risks of intermachine teaming as complex systems?

Their work suggests the research community can support this endeavor by framing AI as human-like machine capabilities of cognition and

communication. This human-centered framing steps away from the obscurity of technology, instead focusing on what machines actually do in human terms as well as the implications for humans, society and business.

Still, the study says, much more research is needed to build on these ideas, such as the application of ethical frameworks to support policies for machine capabilities.

The researchers say everybody will be increasingly affected by so-called smart machines that emulate human capabilities, at home, in the workplace, in the legal system and other spheres of society. There is no reason these machines need to be operating in the world before problems are recognized—they could be corrected or even abandoned if caught early, or designed to mitigate their occurrence from the start.

"In the short run, our approach may simply limit their negative impacts, but in the long run, it may lead to the development and deployment of AI systems based on their benefits and costs for our society," Robert said.

More information: Nigel P. Melville et al, Putting humans back in the loop: An affordance conceptualization of the 4th industrial revolution, *Information Systems Journal* (2022). DOI: 10.1111/isj.12422

Provided by University of Michigan