

How Humans Outshine AI in Adapting to Change

by Rachel Layne

Could artificial intelligence systems eventually perform surgeries or fly planes?

First, AI will have to learn to navigate shifting conditions as well as people do.

Julian De Freitas and colleagues pit humans against machines in a video game to study AI's current limits and mine insights for the real world.

You've probably never thought about all the split-second adjustments you make in a single day to perform different tasks. Wake up in a hotel room, walk into a library, sit behind the wheel of a car, or swipe up to access your phone apps. Each time, you automatically "self-orient" before you even begin a task, pivoting your perspective of where you are and what you can do as your environment changes.

Artificial intelligence can't do that yet—and the machines may have a long way to go before they can truly replicate this near-instant flexibility that is typically second nature for humans, says Julian De Freitas, an assistant professor at Harvard Business School, in the article "[Self-Orienting in Human and Machine Learning](#)," recently published in the journal *Nature Human Behaviour*.

"OUR RESEARCH SHOWS THAT A KEY INGREDIENT THAT MAKES US FLEXIBLE IS HAVING A NOTION OF THE 'SELF,' AND WE CONCRETELY SHOW WHAT THIS BUYS HUMANS OVER AI."

With many companies looking to AI to streamline processes and increase productivity, the research shines a light on the limitations of the technology, says De Freitas, who is also director of the Ethical Intelligence Lab at HBS. Unlike humans, AI can't flexibly navigate changing environments yet because it does not have a notion of its "self" and what it can do with it. This shortcoming raises questions about whether it's safe to rely on AI in certain circumstances, such as an autonomous car that needs to figure out that it has a new problem to solve other than navigation when it unexpectedly gets stuck in a ditch.

"Algorithms can be very good at specialized tasks, and sometimes even have almost superhuman capabilities when confined to specific domains," says De Freitas, who studies automation in marketing. "But what makes humans so effective is that we can do many things. We're pretty flexible. And this is, of course, of immense commercial value as well. Our research shows that a key ingredient that makes us flexible is having a notion of the 'self,' and we concretely show what this buys humans over AI."

De Freitas coauthored the research with Ahmet Kaan Uğuralp and Zeliha Oğuz-Uğuralp of Turkey's Bilkent University; L. A. Paul of Yale University; Joshua Tenenbaum of the Massachusetts Institute of Technology; and Tomer D. Ullman, an assistant professor in Harvard's Psychology Department.

How human responses compare to AI

To test the flexibility of AI versus humans in adjusting to new situations, the authors set up four video games, outlining certain tasks for humans and several popular game-playing AI algorithms to complete. The tasks tested the players' ability to find themselves and respond appropriately amid environments that required increasingly more flexible self-orienting.

Like a simplified version of a four-player scenario of the classic video game Mario Kart, each game included four “possible selves,” which were indicated by red squares. Yet, only one avatar (also known as the “digital self”) was controlled by a player’s keypress. To complete the game, the player—human or machine—had to navigate the digital self to a goal using four moves: up, down, right, or left. Human players used arrow keys. Each of the game versions interfered with the straightforward ability of the human or machine to find its avatar and navigate to a goal.

“PEOPLE WERE SOLVING EVERYTHING FASTER; SELF-ORIENTATION DOESN’T SEEM TO EXIST AT ALL FOR AI.”

The games were designed so that, in principle, a player could solve them without self-orienting, for example, by noticing whichever avatar is closest to the goal, and trying to navigate that avatar to the reward. Yet, the researchers hypothesized that human players would solve the games by “self-orienting”—that is, first figuring out which avatar was their digital self, then proceeding to navigate their digital selves to the “rewarding goal.”

On the AI side, researchers tested six common types of reinforcement learning algorithms that had been designed to learn from frame-by-frame images of the game. The four games were successively harder, going from a simple logic game to one in which embodiments rapidly switched, seemingly at random.

The final score: 4-0 for humans. “People were solving everything faster; self-orientation doesn’t seem to exist at all for AI,” De Freitas says.

How does the technology need to improve?

Developers still need to figure out how and where AI can learn to successfully deal with the unexpected, taking inspiration from how humans naturally solve problems by filling in gaps for situations they’ve never encountered, he says.

Consider, for example, a doctor dealing with a disabled elderly patient in an Emergency Room, after just seeing a healthy young patient. Good doctors know that they have to reorient themselves to a different problem—not just treating the patient but making sure the older person is helped to the room and assisted throughout the examination. Approaching this situation successfully requires recognizing the problem has changed and reorienting to the new task, says De Freitas.

“IN CONTRAST, HUMANS ADAPT; THEY CONTINUOUSLY UNDERSTAND WHERE THEY ARE IN THE WORLD AND WHAT PROBLEM THEY ARE SOLVING IN RESPONSE TO CHANGING CIRCUMSTANCES FAR BETTER THAN CURRENT AI DOES.”

“The current way to achieve this feat with AI is to throw a lot of data at it and hope that AI sees everything it needs to see to learn what it should learn. But I don’t think that’s a flexible, fail-safe approach,” De Freitas says. “In contrast, humans adapt; they continuously understand where they are in the world and what problem they are solving in response to changing circumstances far better than current AI does.”

De Freitas is working with collaborators to give AI “the same self-orienting capabilities as humans, so they behave in the right way, no matter what they see,” he says. “But that’s a hard problem to solve.”

Assessing the capabilities of AI

So how can companies apply the research findings when considering when and how to fold AI into everyday work tasks? De Freitas offers some suggestions:

For now, proceed cautiously when using AI in fast-changing conditions. Managers should be aware of when using an algorithm will speed processes and when it will slow them down and/or be more likely to fail. The research shows AI is more likely to struggle in situations where environments shift enough to require a pivot of the self.

“In any sort of changing environment setting—like shifting between different workflows, providing personalized care to a wide range of patients with various problems, or the example of an automated vehicle having to respond to changing environments—this is where humans are going to shine more than automation systems,” De Freitas says. “If you more deeply understand *why* your AI systems are limited, you are probably better equipped to know when and how to deploy them in practice.”

Acknowledge the gap in ability between AI and humans. “Just identifying and acknowledging the gap is the first step in addressing it in whatever way makes sense for the way that you're leveraging automation, such as improving the system itself or supplementing it with human decision-making,” De Freitas says. “All managers want these systems to be adaptive, intuitive, and have broad applications. Our work identifies a key reason why that's still hard.”

You Might Also Like:

- How Will the Tech Titans Behind ChatGPT, Bard, and LLaMA Make Money?
- Want to Leave a Lasting Impression on Customers? Don't Forget the (Proverbial) Fireworks
- Can Autonomous Vehicles Drive with Common Sense?

Feedback or ideas to share? Email the Working Knowledge team at hbswk@hbs.edu.

Designer's note: Image created uses elements from AdobeStock/scharfsinn86, AdobeStock/Shamil, and generated by Midjourney, an artificial intelligence tool
