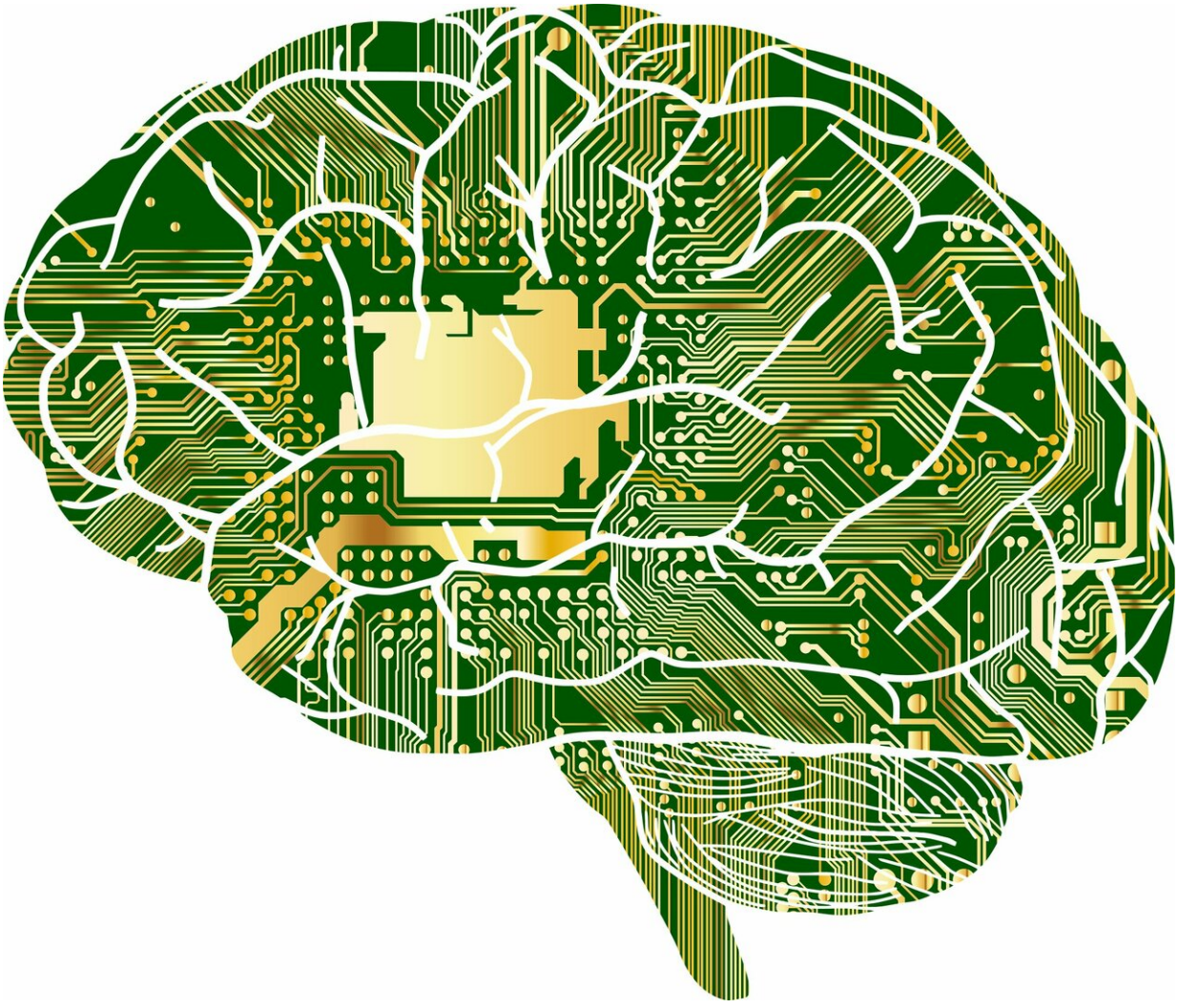
Researchers develop hybrid human-machine framework for building smarter AI

March 7 2022



Credit: CC0 Public Domain

From chatbots that answer tax questions to algorithms that drive autonomous vehicles and dish out medical diagnoses, artificial intelligence undergirds many aspects of daily life. Creating smarter, more accurate systems requires a hybrid human-machine approach, according to researchers at the University of California, Irvine. In a study published this month in *Proceedings of the National Academy of Sciences*, they present a new mathematical model that can improve performance by combining human and algorithmic predictions and confidence scores.

"Humans and machine algorithms have complementary strengths and weaknesses. Each uses different sources of information and strategies to make predictions and decisions," said co-author Mark Steyvers, UCI professor of cognitive sciences. "We show through empirical demonstrations as well as theoretical analyses that humans can improve the predictions of AI even when human accuracy is somewhat below [that of] the AI—and vice versa. And this accuracy is higher than combining predictions from two individuals or two AI algorithms."

To test the framework, researchers conducted an image classification experiment in which human participants and computer algorithms worked separately to correctly identify distorted pictures of animals and everyday items—chairs, bottles, bicycles, trucks. The human participants ranked their confidence in the accuracy of each image identification as low, medium or high, while the machine classifier generated a continuous score. The results showed large differences in confidence between humans and AI algorithms across images.

"In some cases, human participants were quite confident that a particular picture contained a chair, for example, while the AI algorithm was confused about the image," said co-author Padhraic Smyth, UCI Chancellor's Professor of computer science. "Similarly, for other images, the AI algorithm was able to confidently provide a label for the

object shown, while human participants were unsure if the distorted picture contained any recognizable object."

When predictions and confidence scores from both were combined using the researchers' new Bayesian framework, the hybrid model led to better performance than either human or machine predictions achieved alone.

"While past research has demonstrated the benefits of combining machine predictions or combining human predictions—the so-called 'wisdom of the crowds' – this work forges a new direction in demonstrating the potential of combining human and machine predictions, pointing to new and improved approaches to human-AI collaboration," Smyth said.

This interdisciplinary project was facilitated by the Irvine Initiative in AI, Law, and Society. The convergence of cognitive sciences—which are focused on understanding how humans think and behave—with computer science—in which technologies are produced—will provide further insight into how humans and machines can collaborate to build more accurate artificially intelligent systems, the researchers said.

Additional co-authors include Heliodoro Tejada, a UCI graduate student in cognitive sciences, and Gavin Kerrigan, a UCI Ph.D. student in computer science.

More information: Bayesian modeling of human–AI complementarity, *Proceedings of the National Academy of Sciences* (2022). DOI: 10.1073/pnas.2111547119.

Provided by University of California, Irvine