▼

DOW JONES, A NEWS CORP COMPANY

DJIA Futures **26079** -0.02% ▼	S&P 500 F **2899.50** 0.01% ▲	Stoxx 600 **385.76** 0.08% ▲	U.S. 10 Yr **-1/32 Yield** 2.885% ▼	Crude Oil **69.17** 0.93% ▲

COMMENTARY

Natural Intelligence Still Has Its Advantages

AI is disruptive, but it hasn't rendered humanity obsolete.

"Jeopardy!" contestant Ken Jennings and his opponent, an IBM computer called Watson, in Yorktown Heights, N.Y., Jan. 13, 2011.
PHOTO: SETH WENIG/ASSOCIATED PRESS

By Vincent Conitzer
Aug. 28, 2018 6:27 p.m. ET

The late Stephen Hawking warned that artificial intelligence could "spell the end of the human race." Already, AI systems are starting to outperform people in domains ranging from board games to speech recognition. Is humanity on the way out?

For those not working in AI, it can be difficult to interpret achievements in the field. Take, for example, Watson's 2011 victory over human "Jeopardy!" champions Brad Rutter and Ken Jennings. This was a stunning achievement because while it should surprise nobody that Watson had access to an encyclopedic amount of knowledge, "Jeopardy!" requires more than that. The hard part—at least for AI systems, but often for humans as well—is understanding the clue well enough to guess the correct answer. Many AI researchers, myself included, thought this would remain beyond the capabilities of AI systems for a while to come. We were wrong.

But does this mean Watson obtained a human-level understanding of the world? No. Watson also produced some cringeworthy responses, for example "What is Toronto?" for a clue in the "U.S. cities" category.

This is part of a broader pattern of AI systems achieving superhuman levels of performance and yet making blunders that leave us scratching our heads. Researchers from Carnegie Mellon were able to fool a facial recognition system consistently into thinking that one of them, clearly a man, was actress Milla Jovovich, by wearing carefully designed eyeglass frames.

In both cases, the cause of the mistake is that the AI system solves problems in a very different way than humans. Often, this involves picking up on some statistical pattern that can be used to great effect, but it sometimes produces answers that lack common sense. Moreover, if something changes about how the data is produced, performance may plummet. This occurs especially when the change is intended to mislead the system, as with the eyeglass frames.

From this information, we can gather insight into which jobs, or parts of jobs, AI systems will take from us. Tasks that require responding to the same kind of standardized input over and over, with a clear measure of success, are a natural fit. Such tasks range from the diagnosis of

medical images to flipping burgers. On the other hand, jobs that are messy and unpredictable and require an understanding of people and the broader world—I like to think of kindergarten teachers—will likely remain safe for a long time.

Much progress has been made in AI in a short time, so future breakthroughs are not unthinkable. For now, humans remain unsurpassed in their broad, integrated, flexible and robust understanding of the world. If AI starts to catch up with us on that, some of the most intractable problems in philosophy, such as the nature of consciousness, will become very pertinent. But currently there is no clear path toward building such systems. The AI systems we know how to build today are likely to be disruptive in many domains—the labor market, our social fabric, the nature of warfare. But they do not make humanity obsolete.

Mr. Conitzer is a professor of computer science at Duke University.

Appeared in the August 29, 2018, print edition.