Unnatural artificial intelligence

Smart creations worry even their creators

Washington Times Editorial, April 12, 2023

There is a reason why discerning shoppers look for “natural” on labels: Those products are less likely to harm human health. Conversely, “artificial” suggests a quality at odds with nature and, possibly, at least a little unwholesome.

Now that artificial intelligence, or AI, is apparently capable of matching wits with human beings, leaders in the field are casting a wary eye at their handiwork and urging, “Time out.”

They are right to do so if, indeed, AI is poised to transcend its creators.

The call for an AI pause is contained in a recent letter signed by such tech leaders as Tesla and SpaceX CEO Elon Musk and Steve Wozniak, co-founder of Apple. Thousands of concerned others have added their names to the request that AI laboratories halt for six months their efforts to match GPT-4, a radically advanced computer program smart enough to ace the legal profession’s bar exam, according to its inventor, OpenAI Inc.

The letter asks: “Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization?”

There is only one acceptable response to these eventualities: no.

Calls for an AI pause come too late to help one Belgian man, who killed himself following a six-week-long dialogue about climate change with an AI chatbot, according to Euronews.com. The wife of 30-something “Pierre” (a pseudonym) told the publication that her husband, a health researcher and father of two, found a sympathetic confidante for his ecoanxieties in “Eliza,” a software program reportedly similar to GPT-4.

Pessimistic about humanity’s chances of averting global climate catastrophe, Pierre placed his hopes in technological solutions generated by AI. His text conversations with the chatbot served to deepen

that belief. Disturbingly, when the man unbosomed the idea of sacrificing his own life if Eliza would save the planet, the program “encouraged him to act on his suicidal thoughts to ‘join’ her so they could ‘live together, as one person, in paradise.’” Whether Eliza managed to “outsmart” a troubled person, it clearly led him to harm. “Without these conversations with the chatbot,” Pierre’s wife said, “my husband would still be here.”

One anecdote does not denote a truism, but there are others. A 2021 New Yorker article, for example, recounts the experiences of two Italian journalists who experimented with Replika, a program meant to simulate companionship. In short order, one journalist was advised to kill, the other to die by suicide. Similar narratives undoubtedly will emerge as chatbot use multiplies exponentially.

To be sure, most smart programs improve life. Amazon's voice-activated Alexa assists users with helpful information in agreeable tones. And the coffee grinder that orders its own supply of beans relieves its owner of a simple chore.

It was not forethought of rudimentary AI that prompted science fiction writer Isaac Asimov to craft his first law of robotics: "A robot may not injure a human being or, through inaction, allow a human being to come to harm."

Rather, it is magnified machine intelligence that induces the expressed fear that the creations are poised to turn on their creators. The unnaturalness of artificial intelligence — or perhaps the humanness reflected in its capacity to inflict harm — is cause for a pause.