

<https://www.wsj.com/articles/will-ai-help-us-or-leave-us-behind-a1dfa5d3>

ESSAY

## Will AI Help Us or Leave Us Behind?

The most urgent question for the revolution unfolding before our eyes is whether we can keep the technology connected to our humanity.

By Walter Isaacson

May 11, 2023 9:59 am ET

Technology revolutions usually start with little fanfare. No one woke up one morning in 1760 and shouted, “OMG, the Industrial Revolution has just begun!” Even the Digital Revolution chugged away for many years in the background, with hobbyists cobbling together personal computers to show off at geeky gatherings such as the Homebrew Computer Club, before people noticed that the world was being fundamentally transformed.

The Artificial Intelligence Revolution is different. Within a few weeks in the spring of 2023, millions of tech-aware people and then ordinary folks noticed that a transformation was happening with head-snapping speed that would change the nature of work, learning and creativity and the tasks of daily life.

### Is it inevitable that machines will become super-intelligent on their own?

The advent of chatbots and other forms of generative AI—computers that can generate original text or images by training themselves on enormous sets of data—raises a question that has been central to the history of artificial intelligence. Should our goal be to make humans and

machines tight partners, so that they make progress by working in symbiosis? Or is it inevitable that machines will become super-intelligent on their own, eventually leaving us mere mortals behind?

The patron saint of the first school is the 19th-century English writer and mathematician Ada Lovelace. In the 1840s, she described how the use of punch cards in a numerical calculator being built by her friend Charles Babbage could allow the machine to deal not only with numbers but with anything that could be noted in symbols, including words, music and images. In other words, she envisioned a general-purpose computer. It would even be able to perform “abstract mental processes,” she said. But she added one major caveat: Unlike humans, it would lack the capacity to be creative or to come up with ideas on its own. “It can do whatever we know how to order it to perform,” she wrote, but it “has no pretensions whatever to originate anything.”

A century later, the English mathematician and computer scientist Alan Turing became the patron saint of the second school. His seminal 1950 paper about AI, “Computing Machinery and Intelligence,” posed the question, “Can machines think?” In a section called “Lady Lovelace’s objection,” he rebutted her contention by saying that someday we would develop machines that would learn on their own, rather than merely execute programs. They would be able to originate ideas. “Instead of trying to produce a program to simulate the adult mind, why not rather try to produce one which simulates the child’s?” he asked. “If this were then subjected to an appropriate course of education, one would obtain the adult brain.”

He famously proposed an “imitation game” in which a computer would try to answer questions and hold a conversation with such skill that a person would not be able to guess whether it was a machine or a human. With the release of OpenAI’s GPT-4 and Google’s Bard in March, it

became clear that there are now machines that can easily pass this Turing Test.

Up until now, the Lovelace approach of trying to produce a tighter partnership between human creativity and machine processing power—augmented intelligence rather than artificial intelligence—has prevailed. The digital revolution has been driven by improvements in human-computer interfaces, which are the methods we use to exchange information with our machines as seamlessly and quickly as possible.

The pioneer of this approach was the MIT psychologist and tech visionary J.C.R. Licklider. While working on ways that humans could instantly interact with air defense computers, he helped devise the use of video screens to output information and provide ways for humans to input responses. This experience formed the basis for one of the most influential papers in the history of modern technology, titled “Man-Computer Symbiosis,” which Licklider published in 1960. “The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly,” he wrote, “and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.”

MIT students turned some of these early computers into video games with joysticks and pointers. Other advances in connecting machines and humans quickly followed, such as the point-and-click mouse and the graphic user interface that we now use on our personal computers. At his last meeting at Apple in 2011, Steve Jobs got to see and test Siri, which was an example of the next great advance: the ability to use voice to interact with our machines.

Elon Musk is now pursuing what would be the ultimate step in human-machine mind-melding: a microchip with neural sensors that can be implanted into the brain and allow almost instant sharing of information and signals between us and our computers. This month his company Neuralink completed a final round of animal studies and prepared an application to the Food and Drug Administration to allow chips to be implanted into the brains of human test subjects. “The best way to make sure that AI does not turn against us or destroy humanity is to tightly connect it to human agency,” he says.

What’s potentially unnerving about the latest advances in AI is that these systems, following the vision of Turing, are being designed so that they can learn on their own. They even appear to be developing their own agency rather than being tightly tethered to human instructions and intentions. Most frighteningly, their creators, as was the case with Dr. Frankenstein, sometimes don’t fully understand how and why they produce certain answers, have hallucinations, make false assertions and even seem to express unsettling emotions. This raises the specter of an impending “singularity,” a term used by the mathematician John von Neumann and the sci-fi writer Vernor Vinge to describe the moment when artificial intelligence can forge ahead on its own at an uncontrollable pace and leave us humans behind.

AI safety will not come merely from convening people with knitted brows to write international regulations or codes of AI conduct, as worthy as those efforts may be. Instead, I think it will require that we go back to the approach of finding new and better technologies to tie humans and computers ever closer together. The goal should be to make sure our machines are always connected to human agency. At the very least, that would assure that real humans are responsible and accountable for what the machines do. And at best, it would make it less

likely that these systems run amok and develop, so to speak, a mind of their own.

This semester, in the class I teach at Tulane on the history of digital technology, my students and I got to share the excitement of seeing a new revolution suddenly being born. I had them read William Wordsworth's poem "The French Revolution as It Appeared to Enthusiasts at Its Commencement," with its exhilarating line, "Bliss was it in that dawn to be alive!" But I also pointed out that there was an edgy irony to the title. The French Revolution did not end well. It's fine to be an enthusiast during this dawn, but it's also important to focus on keeping the revolution connected to our humanity.

*Mr. Isaacson is writing a biography of Elon Musk.*