

## Intelligent machines raise real concerns with smart people

**Bloomberg News**

The following editorial appears on Bloomberg View:

From Ancient Greece to Mary Shelley, it's been an abiding human fear: A new invention takes on a life of its own, pursues its own ends, and grows more powerful than its creators ever imagined.

For a number of today's thinkers and scientists, artificial intelligence looks as if it may be on just such a trajectory. This is no cause for horror or a premature backlash against the technology. But AI does present some real long-term risks that should be assessed soberly.

AI involves creating computer software or machines that can intelligently pursue a goal. For decades, the discipline has been plagued by excessive optimism. It's proved especially difficult for researchers to engineer a "general" intelligence, or a machine that can independently solve problems and adapt to new circumstances, like a human.

Yet progress is now accelerating. Narrow applications of AI are all around you – whenever Google provides search results, an iPhone responds to your voice command or Amazon suggests a book you'd like. AI already seems to be better than humans at chess, "Jeopardy" and sensible driving. It holds promise in fields such as energy, science and medicine.

Experts in one survey estimate that artificial intelligence may approach the human kind between 2040 and 2050, and exceed it a few decades later. They also suggest there's a one-in-three probability that such an evolution will be "bad" or "extremely bad" for humanity.

They're not the only ones worried. Two grim new books have sounded the alarm on AI. Elon Musk has warned that it could be "more dangerous than nukes." And Stephen Hawking has called it "potentially the best or worst thing to happen to humanity in history."

These aren't cranks or Luddites. Why are they so alarmed?

First, the advance of artificial intelligence isn't easy to directly observe and measure. As the power and autonomy of a certain AI increased, it might prove difficult to understand or predict its actions. It might evolve quickly and in unexpected ways.

A deeper worry is that if the AI came to exceed human intelligence, it could continually improve itself – growing yet smarter and more powerful with each new version. As it did so, its motives would presumably grow more obscure, its logic more impenetrable and its behavior more unpredictable. It could become difficult or impossible to control.

To use a simplified example: A self-learning AI programmed to calculate the decimals of pi might, as it became more intelligent, decide that the most efficient way to meet its goals would be to commandeer all the computing power on earth. It might conclude, further, that it needed to acquire more resources to expand that computing power. And so on. The role of humans in such a scenario isn't obvious. Or maybe it is.

Researchers are starting to take these risks more seriously. But preparing for them won't be easy. One idea is to conduct experiments in virtual or simulated worlds, where an AI's evolution could be observed and the dangers theoretically contained. A second is to inculcate some simulation of morality into AI systems. Researchers have made progress applying deontic logic – the logic of permission and obligation – to improve robots' reasoning skills and clarify their behavioral rationales. Something similar may be useful in AI more generally.

More prosaically, researchers in the field need to devise commonly accepted guidelines for experimenting with AI and mitigating its risks. Monitoring the technology's advance may also one day call for global political cooperation, perhaps on the model of the International Atomic Energy Agency.

None of which should induce panic. Artificial intelligence remains in its infancy. It seems likely to produce useful new technologies and improve many lives. But without proper constraints, and a load of foresight, it's a technology that could resemble Frankenstein's monster – huge, unpredictable and very different from what its creators had in mind.

— — —

**©2014 Bloomberg News. 8-11-14.  
WASHINGTON POST-BLOOMBERG**

---