Machines That Will Think and Feel

Artificial intelligence is still in its infancy—and that should scare us

What will artificial intelligence accomplish and when?

By DAVID GELERNTER

Wall Street Journal, March 18, 2016 10:36 a.m. ET

Artificial intelligence is breathing down our necks: Software built by Google startled the field last week by easily defeating the world's best player of the Asian board game Go in a five-game match. Go resembles chess in the deep, complex problems it poses but is even harder to play and has resisted AI researchers longer. It requires mastery of strategy and tactics while you conceal your own plans and try to read your opponent's.

Mastering Go fits well into the ambitious goals of AI research. It shows us how much has been accomplished and forces us to confront, as never before, AI's future plans. So what will artificial intelligence accomplish and when?

AI prophets envision humanlike intelligence within a few decades: not expertise at a single, specified task only but the flexible, wide-ranging intelligence that Alan Turing foresaw in a 1950 paper proposing the test for machine intelligence that still bears his name. Once we have figured out how to build artificial minds with the average human IQ of 100, before long we will build machines with IQs of 500 and 5,000. The potential good and bad consequences are staggering. Humanity's future is at stake.

Suppose you had a fleet of AI software apps with IQs of 150 (and eventually 500 or 5,000) to help you manage life. You download them like other apps, and they spread out into your phones and computers—and walls, clothes, office, car, luggage—traveling within the dense computer network of the near future that is laid in by the yard, like thin cloth, everywhere.

AI apps will read your email and write responses, awaiting your nod to send them. They will escort your tax return to the IRS, monitor what is done and report back. They will murmur (from your collar, maybe) that the sidewalk is icier than it looks, a friend is approaching across the street, your gait is slightly odd—have you hurt your back? They will log on for you to 19 different systems using 19 different ridiculous passwords, rescuing you from today's infuriating security protocols. They will answer your phone and tactfully pass on messages, adding any reminders that might help.

In a million small ways, next-generation AI apps will lessen the friction of modern life. Living without them will seem, in retrospect, like driving with no springs or shocks.

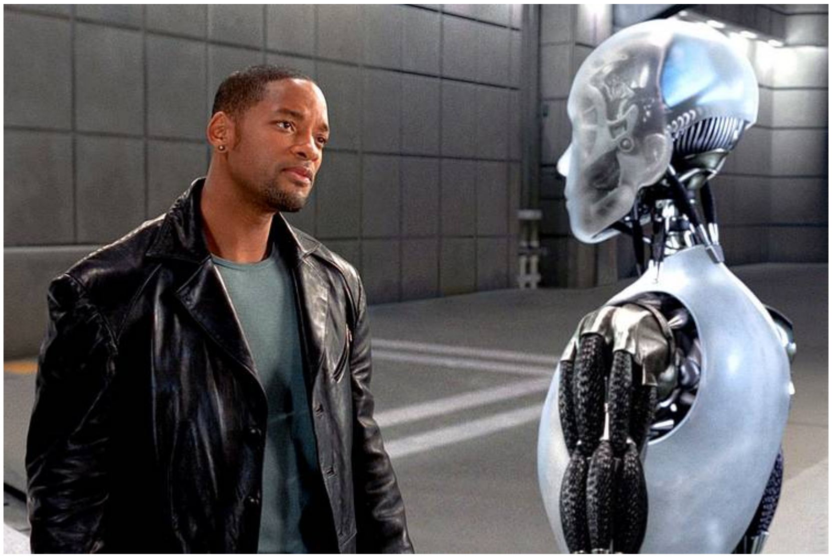
[A scene from 'Blade Runner,' 1982]

But we don't have the vaguest idea what an IQ of 5,000 would mean. And in time, we will build such machines—which will be unlikely to see much difference between humans and houseplants. The software Go master is a potent reminder of how far we've already come down this road. How would we fare in a world of superhuman robots?

Consider a best case. Human beings like butterflies, encourage their presence, don't push them around. Still, we can't tell one butterfly of any given species from another. They are so unlike us that their individual characteristics are beyond our discernment.

Some children still collect butterflies. It's a hobby to encourage (particularly as an alternative to videogames and social network blather); you get outside and learn about nature. If supersmart robots treat us like butterflies, we will be lucky—but don't count on it. We can't support ourselves by browsing flowers, decoratively. Why, then, are we building machines whose intelligence could swamp and overwhelm ours? Machines whose behavior and attitude to mere humans are impossible to foresee?

Because it is human nature to learn as much as we can and to build the best machines we can. Of course, that doesn't make it smart in any particular case. Robots with superhuman intelligence are as dangerous, potentially, as nuclear bombs. Yet they are the natural outcome of robots with "ordinary" intelligence—which are (in turn) the goal of AI researchers working hard, right now, all over the world.



[Will Smith and Sonny the Robot in ‘I, Robot,’ 2004]

Thoughtful people everywhere ought to resolve that it would be unspeakably stupid to allow technologists to fool around with humanlike and superhuman machines—except with the whole world’s intense scrutiny. Technologists are in business to build the most potent machines they can, not to worry about consequences.
So where do we stand? How long before we arrive at these hugely dangerous computers?

For now, we are not even headed in the right direction. AI today is nowhere near understanding the human mind. It’s trying to get to California (so to speak) without ever leaving I-95.

But a breakthrough won’t be hard. We only need to look at things from a slightly different angle—which might happen in a hundred years or this afternoon. Once it does, we had better start following developments as carefully as we would monitor lab technicians playing with plague bacteria.

The issues are surprisingly simple. Most scientists and philosophers identify human thought with rationality, reasoning, logic. Even if they see emotion as important, they rarely see it as central—and even if they do that, they seldom grasp the remarkably simple way that it relates to rational thought and the mind at large.

Filling in this gap in our knowledge is dangerous: It moves us a step closer to superhuman robots. But learning is our fate. It’s impossible to stop. Our best bet is to discover all that we can and to move forward with our eyes wide open.
The human mind is no static, rational machine. Nor is it sometimes rational and sometimes emotional, depending on mood or whim or chance. The mind moves down a continuous spectrum every day, from moments of pure thinking about things to moments of pure being, experience or feeling.

At the start of the day, we tend to be mentally energetic; we tend to start near the spectrum’s top. At the day’s end, we’re at the other end of the spectrum. When we grow sleepy, we are down-spectrum, moving steadily lower. Asleep and dreaming, we are near the bottom.

During the day, we generally go through two main oscillations—drifting downward toward a temporary low point (often in middle or late afternoon), then drifting higher for a few hours, and finally downward again straight into sleep. Everyone has his own pattern, but our general course over a day is from up-spectrum to down. Software could imitate this spectrum—and will have to, if it is ever to achieve humanlike thought.

The spectrum's top edge is what we might call thinking-about—pondering the morning news, or the daffodils outside or the future of American colleges. At the opposite end, you reach a state of pure being or feeling—sensation or emotion—that is about nothing. Chill or warmth, seeing violet or smelling cut grass, uneasiness or thirst, happiness or euphoria—each must have a cause, but they are not about anything. The pleasant coolness of your forearm is not about the spring breeze. Over the day, the mind moves from one kind of mental state to a very different kind, from mental apples to mental oranges.

We can only reproduce this spectrum in software if we can reproduce the endpoints. Thinking-about can be simulated on a computer. But no computer will ever feel anything. Feeling is uncomputable. Feeling and consciousness can only happen (so far as we know) to organic, Earth-type animals—and can't ever be produced, no matter what, by mere software on a computer.

But that doesn't necessarily limit artificial intelligence. Software can simulate feeling. A robot can tell you that it's depressed and act depressed, although it feels nothing. AI could, in principle, build a simulated mind that reproduced all the nuances of human thought and dealt with the world in a thoroughly human way, despite being unconscious. After all, your best friend might be a zombie. (Would it matter? Of course! But only to him.)

Still: No artificial mind will ever be humanlike unless it imitates not just feeling but the whole spectrum. Consider how your mind changes as you slide down-spectrum—gradually mixing together feeling and logical analysis, like a painter mixing ultramarine blue into cool red, producing reds, red-violets, purples, purple-blues. The human mind is this whole range of shades, not just rationality, nor rationality plus a side-order of emotion, nor the definitive list of five or six kinds of mental state some textbooks give. The range is infinite, but the formula for creating this infinity is simple.

At the spectrum's top, we concentrate readily, focus on external reality, solve problems. Memory is docile. It supplies facts, figures and methods without distracting us with fascinating recollections. Emotion lies low. Early mornings are rarely the time for dramatic arguments or histrionic scenes.

As we move down-spectrum, we lose concentration and reasoning power and come gradually to rely more on remembering than on reasoning—on the "wisdom of experience”—to solve problems. Lower still, the mind starts to wander. We daydream: Our memories, fantasies, ideas are growing vivid and distracting as the mind’s focus moves from outside the self to inside.

Finally, we grow sleepy and find ourselves free-associating—memory now controls the mental agenda—and we pass through the strange state of “sleep onset thought” on the way to sleep and dreaming. (Sleep onset thought is usually a series of vivid, static, memories, some or all of them hallucinations—like dreams themselves.)

[http://youtu.be/3OlCzxFuV9c *From the film, ‘2010,’ HAL asks ‘Will I dream?’]*

Emotions grow more powerful as we move downward. Daydreams and fantasies often move us emotionally. Dry, emotionless daydreams are rare. Dreams sometimes have a vaguely unpleasant emotional tone, but they can also make us elated or terrified: The strongest emotions we know happen in dreams.

Up-spectrum, thought uses memory like a faithful, colorless assistant. Down-spectrum, memory increasingly goes off on its own. The flow of information from memory gradually supplants the flow from outside as we sink into ourselves like a flame sinking into a candle.

The spectrum shows that dreaming is no strange intrusion into ordinary thought. It is the normal outcome of descending the spectrum, like rolling down the runway once the airplane has landed. We experience our dreams but don’t think about them much as they unfold; we allow all sorts of improbabilities and absurdities to happen without minding. This means less self-awareness and less memory, which is one reason why down-spectrum thinking has been so hard to grasp, from Joseph’s work in Genesis to Freud’s in Vienna.

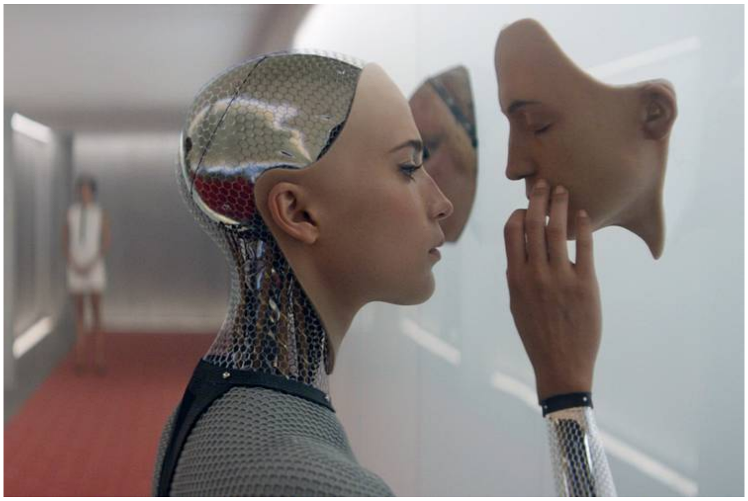
We need to understand all of this if we are to understand the mind. But does AI need to reproduce it all to achieve humanlike intelligence?

Every bit of it. One important down-spectrum activity is creativity. Reporters of personal experience tend to suggest that creativity happens when they are not thinking hard about a problem, when mental focus is diffuse—in other words, when they are down-spectrum. Creativity often centers on inventing new analogies, which allow us to see an old problem in the light of a new comparison.

But where does the new analogy come from? The poet Rilke compares the flight of a small bird across the evening sky to a crack in a smooth porcelain cup. How did he come up with that? Possibly by using the fact that these very different things made him feel the same way.

Emotion is a hugely powerful and personal encoding-and-summarizing function. It can comprehend a whole complex scene in one subtle feeling. Using that feeling as an index value, we can search out—among huge collections of candidates—the odd memory with a deep resemblance to the thing we have in mind.

Once AI has decided to notice and accept this spectrum—this basic fact about the mind—we will be able to reproduce it in software. The path from the spectrum to the superhuman robot is no short hop; it's more like a flight to the moon.



[Alicia Vikander in 'Ex Machina,' 2015]

But once we had rocket engines, radios, computers and a bunch of other technologies in hand, the moon flight was inevitable (though it also took genius, daring and endless work). The spectrum is just one among the technologies that are necessary to make AI work. But it's one that has been missing.

Humanlike and superhuman machines follow inevitably from the spectrum and other basic facts about the mind. That makes these fundamental ideas just as dangerous as the vision I started with: the building of super-intelligent computers. They are all natural outcomes of the human need to build and understand. We can't shut that down and don't want to.

The more we learn, the more carefully, critically and intelligently we can observe the dangerous doings of AI. Let normal people beware of AI researchers: "All who heard should see them there/ And all shout cry, Beware! Beware!/ Their flashing eyes, their floating hair!/ Weave a circle round them thrice/ And close your eyes with holy dread,/ For they on honeydew hath fed,/ And drunk the milk of Paradise." Is it strange to conclude a piece on artificial intelligence with Coleridge's "Kubla Khan"? But AI research is strange.

Mr. Gelernter is a professor of computer science at Yale and the author of "Tides of Mind: Uncovering the Spectrum of Consciousness," recently published by Liveright, a division of W.W. Norton.

http://www.wsj.com/articles/when-machines-think-and-feel-1458311760