JOOST SWARTE

Intelligent Machines

The Seven Deadly Sins of AI Predictions

Mistaken extrapolations, limited imagination, and other common mistakes that distract us from thinking more productively about the future.

by Rodney Brooks   October 6, 2017

We are surrounded by hysteria about the future of artificial intelligence and robotics—hysteria about how powerful they will become, how quickly, and what they will do to jobs.

I recently saw a story in MarketWatch that said robots will take half of today's jobs in 10 to 20 years. It even had a graphic to prove the numbers.

The claims are ludicrous. (I try to maintain professional language, but sometimes …) For instance, the story appears to say that we will go from one million grounds and maintenance workers in the U.S. to only 50,000 in 10

to 20 years, because robots will take over those jobs. How many robots are currently operational in those jobs? *Zero.* How many realistic demonstrations have there been of robots working in this arena? *Zero.* Similar stories apply to all the other categories where it is suggested that we will see the end of more than 90 percent of jobs that currently require physical presence at some particular site.

This story is part of our November/December 2017 Issue

See the rest of the issue

Subscribe

Mistaken predictions lead to fears of things that are not going to happen, whether it's the wide-scale destruction of jobs, the Singularity, or the advent of AI that has values different from ours and might try to destroy us. We need to push back on these mistakes. But why are people making them? I see seven common reasons.

JOOST SWARTE

1. Overestimating and underestimating

Roy Amara was a cofounder of the Institute for the Future, in Palo Alto, the intellectual heart of Silicon Valley. He is best known for his adage now referred to as Amara's Law:

> We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run.

There is a lot wrapped up in these 21 words. An optimist can read it one way, and a pessimist can read it another.

A great example of the two sides of Amara's Law is the U.S. Global Positioning System. Starting in 1978, a constellation of 24 satellites (now 31 including spares) were placed in orbit. The goal of GPS was to allow precise delivery of munitions by the U.S. military. But the program was nearly canceled again and again in the 1980s. The first operational use for its intended purpose was in 1991 during Desert Storm; it took several more successes for the military to accept its utility.


Mistaken predictions lead to fears of things that are not going to happen.

Today GPS is in what Amara would call the long term, and the ways it is used were unimagined at first. My Series 2 Apple Watch uses GPS while I am out running, recording my location accurately enough to see which side of the street I run along. The tiny size and price of the receiver would have been incomprehensible to the early GPS engineers. The technology synchronizes physics experiments across the globe and plays an intimate role in synchronizing the U.S. electrical grid and keeping it running. It even allows the high-frequency traders who really control the stock market to mostly avoid disastrous timing errors. It is used by all our airplanes, large and small, to navigate, and it is used to track people out of prison on parole. It determines which seed variant will be planted in which part of many fields across the globe. It tracks fleets of trucks and reports on driver - performance.

GPS started out with one goal, but it was a hard slog to get it working as well as was originally expected. Now it has seeped into so many aspects of our lives that we would not just be lost if it went away; we would be cold, hungry, and quite possibly dead.

We see a similar pattern with other technologies over the last 30 years. A big promise up front, disappointment, and then slowly growing confidence in results that exceed the original expectations. This is true of computation, genome sequencing, solar power, wind power, and even home delivery of groceries.

AI has been overestimated again and again, in the 1960s, in the 1980s, and I believe again now, but its prospects for the long term are also probably being underestimated. The question is: *How long is the long term*? The next six errors help explain why the time scale is being grossly underestimated for the future of AI.
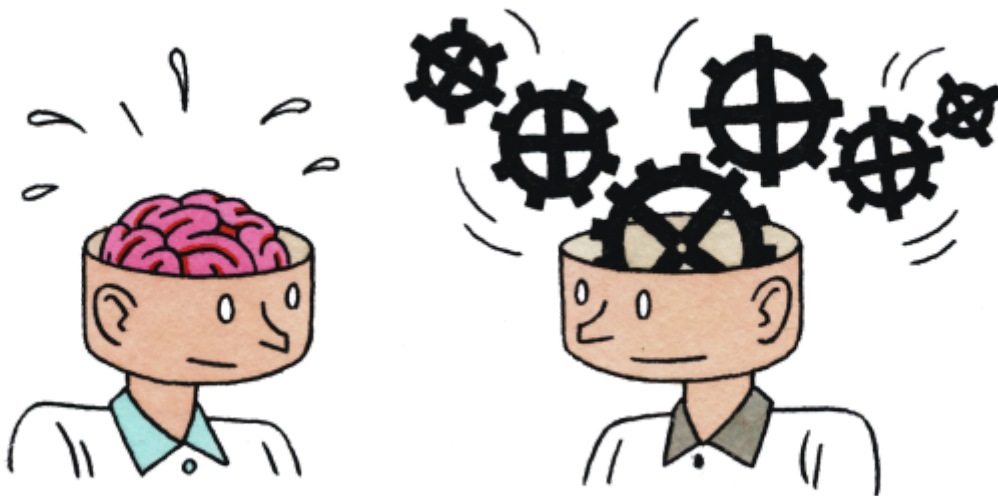
JOOST SWARTE

2. Imagining magic

When I was a teenager, Arthur C. Clarke was one of the "big three" science fiction writers, along with Robert Heinlein and Isaac Asimov. But Clarke was also an inventor, a science writer, and a futurist. Between 1962 and 1973

he formulated three adages that have come to be known as Clarke's Three Laws:

> When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.

> The only way of discovering the limits of the possible is to venture a little way past them into the impossible.

> Any sufficiently advanced technology is indistinguishable from magic.

Personally, I should probably be wary of the second sentence in his first law, as I am much more conservative than some others about how quickly AI will be ascendant. But for now I want to expound on Clarke's Third Law.

Imagine we had a time machine and we could transport Isaac Newton from the late 17th century to today, setting him down in a place that would be familiar to him: Trinity College Chapel at the University of Cambridge.

Now show Newton an Apple. Pull out an iPhone from your pocket, and turn it on so that the screen is glowing and full of icons, and hand it to him. Newton, who revealed how white light is made from components of different-colored light by pulling apart sunlight with a prism and then putting it back together, would no doubt be surprised at such a small object producing such vivid colors in the darkness of the chapel. Now play a movie of an English country scene, and then some church music that he would have heard. And then show him a Web page with the 500-plus pages of his personally annotated copy of his masterpiece *Principia*, teaching him how to use the pinch gesture to zoom in on details.

Watch out for arguments about future technology that is magical.

Could Newton begin to explain how this small device did all that? Although he invented calculus and explained both optics and gravity, he was never able to sort out chemistry from alchemy. So I think he would be flummoxed, and unable to come up with even the barest coherent outline of what this

device was. It would be no different to him from an embodiment of the occult—something that was of great interest to him. It would be indistinguishable from magic. And remember, Newton was a really smart dude.

If something is magic, it is hard to know its limitations. Suppose we further show Newton how the device can illuminate the dark, how it can take photos and movies and record sound, how it can be used as a magnifying glass and as a mirror. Then we show him how it can be used to carry out arithmetical computations at incredible speed and to many decimal places. We show it counting the steps he has taken as he carries it, and show him that he can use it to talk to people anywhere in the world, immediately, from right there in the chapel.

What else might Newton conjecture that the device could do? Prisms work forever. Would he conjecture that the iPhone would work forever just as it is, neglecting to understand that it needs to be recharged? Recall that we nabbed him from a time 100 years before the birth of Michael Faraday, so he lacked a scientific understanding of electricity. If the iPhone can be a source of light without fire, could it perhaps also transmute lead into gold?

This is a problem we all have with imagined future technology. If it is far enough away from the technology we have and understand today, then we do not know its limitations. And if it becomes indistinguishable from magic, anything one says about it is no longer falsifiable.

This is a problem I regularly encounter when trying to debate with people about whether we should fear artificial general intelligence, or AGI—the idea that we will build autonomous agents that operate much like beings in the world. I am told that I do not understand how powerful AGI will be. That is not an argument. We have no idea whether it can even exist. I would like it to exist—this has always been my own motivation for working in robotics and AI. But modern-day AGI research is not doing well at all on either being general or supporting an independent entity with an ongoing existence. It mostly seems stuck on the same issues in reasoning and common sense that AI has had problems with for at least 50 years. All the evidence that I see says we have no real idea yet how to build one. Its

properties are completely unknown, so rhetorically it quickly becomes magical, powerful without limit.

Nothing in the universe is without limit.

Watch out for arguments about future technology that is magical. Such an argument can never be refuted. It is a faith-based argument, not a scientific argument.

JOOST SWARTE

3. Performance versus competence

We all use cues about how people perform some particular task to estimate how well they might perform some different task. In a foreign city we ask a stranger on the street for directions, and she replies with confidence and with directions that seem to make sense, so we figure we can also ask her about the local system for paying when you want to take a bus.

Now suppose a person tells us that a particular photo shows people playing Frisbee in the park. We naturally assume that this person can answer questions like *What is the shape of a Frisbee? Roughly how far can a person throw a Frisbee? Can a person eat a Frisbee? Roughly how many people play Frisbee at once? Can a three-month-old person play Frisbee? Is today's weather suitable for playing Frisbee?*

Computers that can label images like "people playing Frisbee in a park" have no chance of answering those questions. Besides the fact that they can only label more images and cannot answer questions at all, they have no idea what a person is, that parks are usually outside, that people have ages, that weather is anything more than how it makes a photo look, etc.

This does not mean that these systems are useless; they are of great value to search engines. But here is what goes wrong. People hear that some robot or some AI system has performed some task. They then generalize from that performance to a competence that a person performing the same task could be expected to have. And they apply that generalization to the robot or AI system.

Today's robots and AI systems are incredibly narrow in what they can do. Human-style generalizations do not apply.
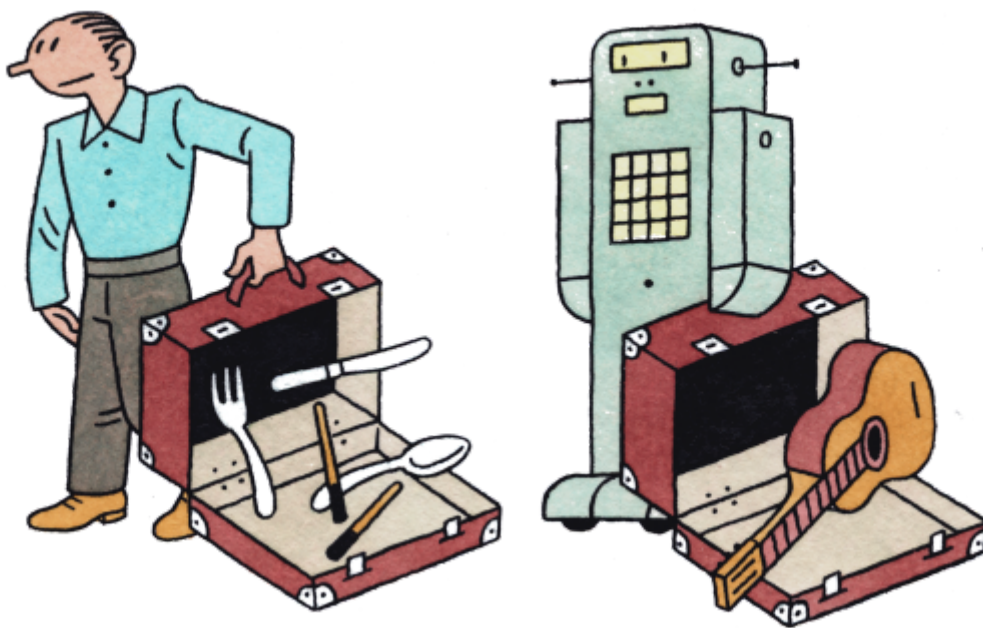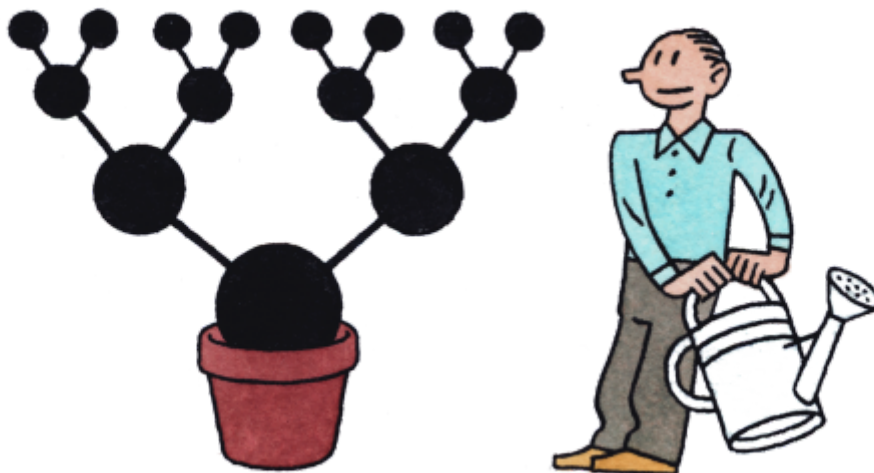

JOOST SWARTE

4. Suitcase words

Marvin Minsky called words that carry a variety of meanings "suitcase words." "Learning" is a powerful suitcase word; it can refer to so many different types of experience. Learning to use chopsticks is a very different experience from learning the tune of a new song. And learning to write code is a very different experience from learning your way around a city.

When people hear that machine learning is making great strides in some new domain, they tend to use as a mental model the way in which a person would learn that new domain. However, machine learning is very brittle, and it requires lots of preparation by human researchers or engineers, special-purpose coding, special-purpose sets of training data, and a custom learning structure for each new problem domain. Today's machine learning is not at all the sponge-like learning that humans engage in, making rapid progress in a new domain without having to be surgically altered or purpose-built.

Likewise, when people hear that a computer can beat the world chess champion (in 1997) or one of the world's best Go players (in 2016), they tend to think that it is "playing" the game just as a human would. Of course, in reality those programs had no idea what a game actually was, or even that they were playing. They were also much less adaptable. When humans play a game, a small change in rules does not throw them off. Not so for AlphaGo or Deep Blue.

Suitcase words mislead people about how well machines are doing at tasks that people can do. That is partly because AI researchers—and, worse, their institutional press offices—are eager to claim progress in an instance of a suitcase concept. The important phrase here is "an instance." That detail soon gets lost. Headlines trumpet the suitcase word, and warp the general understanding of where AI is and how close it is to accomplishing more.

JOOST SWARTE

5. Exponentials

Many people are suffering from a severe case of "exponentialism."

Everyone has some idea about Moore's Law, which suggests that computers get better and better on a clockwork-like schedule. What Gordon Moore actually said was that the number of components that could fit on a microchip would double every year. That held true for 50 years, although the time constant for doubling gradually lengthened from one year to over two years, and the pattern is coming to an end.

Doubling the components on a chip has made computers continually double in speed. And it has led to memory chips that quadruple in capacity every two years. It has also led to digital cameras that have better and better resolution, and LCD screens with exponentially more pixels.

The reason Moore's Law worked is that it applied to a digital abstraction of a true-or-false question. In any given circuit, is there an electrical charge or voltage there or not? The answer remains clear as chip components get smaller and smaller—until a physical limit intervenes, and we get down to components with so few electrons that quantum effects start to dominate. That is where we are now with our silicon-based chip technology.

When people are suffering from exponentialism, they may think that the exponentials they use to justify an argument are going to continue apace. But Moore's Law and other seemingly exponential laws can fail because they were not truly exponential in the first place.

Back in the first part of this century, when I was running MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and needed to help raise money for over 90 different research groups, I tried to use the memory increase on iPods to show sponsors how things were continuing to change very rapidly. Here are the data on how much music storage one got in an iPod for $400 or less:

year	gigabytes

2002	10
2003	20
2004	40
2006	80
2007	160

Then I would extrapolate a few years out and ask what we would do with all that memory in our pockets.

Extrapolating through to today, we would expect a $400 iPod to have 160,000 gigabytes of memory. But the top iPhone of today (which costs much more than $400) has only 256 gigabytes of memory, less than double the capacity of the 2007 iPod. This particular exponential collapsed very suddenly once the amount of memory got to the point where it was big enough to hold any reasonable person's music library and apps, photos, and videos. Exponentials can collapse when a physical limit is hit, or when there is no more economic rationale to continue them.

Similarly, we have seen a sudden increase in performance of AI systems thanks to the success of deep learning. Many people seem to think that means we will continue to see AI performance increase by equal multiples on a regular basis. But the deep-learning success was 30 years in the making, and it was an isolated event.

That does not mean there will not be more isolated events, where work from the backwaters of AI research suddenly fuels a rapid-step increase in the performance of many AI applications. But there is no "law" that says how often they will happen.
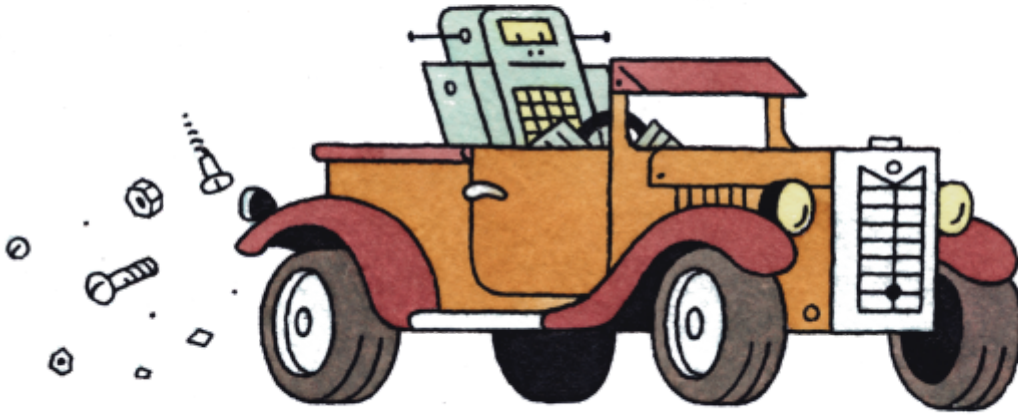
JOOST SWARTE

6. Hollywood scenarios

The plot for many Hollywood science fiction movies is that the world is just as it is today, except for one new twist.

In *Bicentennial Man*, Richard Martin, played by Sam Neill, sits down to breakfast and is waited upon by a walking, talking humanoid robot, played by Robin Williams. Richard picks up a newspaper to read over breakfast. A newspaper! Printed on paper. Not a tablet computer, not a podcast coming from an Amazon Echo–like device, not a direct neural connection to the Internet.

It turns out that many AI researchers and AI pundits, especially those pessimists who indulge in predictions about AI getting out of control and killing people, are similarly imagination-challenged. They ignore the fact that if we are able to eventually build such smart devices, the world will have changed significantly by then. We will not suddenly be surprised by the existence of such super-intelligences. They will evolve technologically over time, and our world will come to be populated by many other intelligences, and we will have lots of experience already. Long before there are evil super-intelligences that want to get rid of us, there will be somewhat less intelligent, less belligerent machines. Before that, there will be really

grumpy machines. Before that, quite annoying machines. And before them, arrogant, unpleasant machines. We will change our world along the way, adjusting both the environment for new technologies and the new technologies themselves. I am not saying there may not be challenges. I am saying that they will not be sudden and unexpected, as many people think.

JOOST SWARTE

7. Speed of deployment

New versions of software are deployed very frequently in some industries. New features for platforms like Facebook are deployed almost hourly. For many new features, as long as they have passed integration testing, there is very little economic downside if a problem shows up in the field and the version needs to be pulled back. This is a tempo that Silicon Valley and Web software developers have gotten used to. It works because the marginal cost of newly deploying code is very, very close to zero.

Deploying new hardware, on the other hand, has significant marginal costs. We know that from our own lives. Many of the cars we are buying today, which are not self-driving, and mostly are not software-enabled, will probably still be on the road in the year 2040. This puts an inherent limit on how soon all our cars will be self-driving. If we build a new home today, we

can expect that it might be around for over 100 years. The building I live in was built in 1904, and it is not nearly the oldest in my neighborhood.

Capital costs keep physical hardware around for a long time, even when there are high-tech aspects to it, and even when it has an existential mission.

The U.S. Air Force still flies the B-52H variant of the B-52 bomber. This version was introduced in 1961, making it 56 years old. The last one was built in 1962, a mere 55 years ago. Currently these planes are expected to keep flying until at least 2040, and perhaps longer—there is talk of extending their life to 100 years.

I regularly see decades-old equipment in factories around the world. I even see PCs running Windows 3.0—a software version released in 1990. The thinking is "If it ain't broke, don't fix it." Those PCs and their software have been running the same application doing the same task reliably for over two decades.

Almost all innovations in robotics and AI take far, far, longer to be really widely deployed than people in the field and outside the field imagine.

The principal control mechanism in factories, including brand-new ones in the U.S., Europe, Japan, Korea, and China, is based on programmable logic controllers, or PLCs. These were introduced in 1968 to replace electromechanical relays. The "coil" is still the principal abstraction unit used today, and PLCs are programmed as though they were a network of 24-volt electromechanical relays. Still. Some of the direct wires have been replaced by Ethernet cables. But they are not part of an open network. Instead they are individual cables, run point to point, physically embodying the control flow—the order in which steps get executed—in these brand-new ancient automation controllers. When you want to change information flow, or control flow, in most factories around the world, it takes weeks of consultants figuring out what is there, designing new reconfigurations, and then teams of tradespeople to rewire and reconfigure hardware. One of the

major manufacturers of this equipment recently told me that they aim for three software upgrades every 20 years.

In principle, it could be done differently. In practice, it is not. I just looked on a jobs list, and even today, this very day, Tesla Motors is trying to hire PLC technicians at its factory in Fremont, California. They will use electromagnetic relay emulation in the production of the most AI-enhanced automobile that exists.

A lot of AI researchers and pundits imagine that the world is already digital, and that simply introducing new AI systems will immediately trickle down to operational changes in the field, in the supply chain, on the factory floor, in the design of products.

Nothing could be further from the truth. Almost all innovations in robotics and AI take far, far, longer to be really widely deployed than people in the field and outside the field imagine.

Rodney Brooks is a former director of the Computer Science and Artificial Intelligence Laboratory at MIT and a founder of Rethink Robotics and iRobot. This essay is adapted with permission from a post *that originally appeared at rodneybrooks.com.*

Time is running out to register for EmTech Digital. You don't want to miss expert discussions on AI.
Learn more and register

Related Video   **More videos**