

## Review and Excerpts: The Doomsday Invention

2018-12-18 21:44:25

The people who say that artificial intelligence is not a problem tend to work in artificial intelligence. Many prominent researchers regard Bostrom's basic views as implausible, or as a distraction from the near-term benefits and moral dilemmas posed by the technology—not least because A.I. systems today can barely guide robots to open doors. – The Doomsday Invention

AI researchers (myself included) tend to see the problems and bottlenecks in their research and systems more than the path they have progressed. When you know the tricks you added to look your robot intelligent, that seems less intelligent.

But people forget that the problem with machines is that they make people dependent. People forget to use logarithm tables after the invention of calculator and they will probably lose interest in any kind of deep thinking when a suitable machines becomes available.

When machines become more like human, we become more like animals.

---

“People think it is about technology, but it is really about religion, people turning to metaphysics to cope with the human condition. They have a way of dramatizing their beliefs with an end-of-days scenario—and one does not want to criticize other people's religions.”  
– The Doomsday Invention

Certainly the problem is about our *existential crisis* as a species. Religion provides a frame to speak about the universe and implicitly we consider ourselves unique, whether you believe in science or *a God that created heavens and earth*.

When we consider earthly machines may become a threat to our earthly existence, it requires us to find another *meaning* for our existence. We need to find a way that humans are still relevant in a future world, but so far, it seems rather difficult.

No matter how improbable extinction may be, Bostrom argues, its consequences are near-infinitely bad; thus, even the tiniest step toward reducing the chance that it will happen is near-infinitely valuable. – The Doomsday Invention

There is more *greater than 0* probability that robots will make us irrelevant and when a species becomes irrelevant, it's very difficult for it to survive. We may find a niche, like an old aristocrat after a revolution, but that won't suffice to keep humanity to flourish.

---

Many of those Earth-like planets are thought to be far, far older than ours. One that was recently discovered, called Kepler 452b, is as much as one and a half billion years older. Bostrom asks: If life had formed there on a time scale resembling our own, what would it look like? What kind of technological progress could a civilization achieve with a head start of hundreds of millions of years? – The Doomsday Invention

Earth may have had a technological progress in the past. We cannot see its effects because maybe it's too old that all evidence are lost.

What if dinosaurs became highly intelligent and lived an industrial revolution and collapse within a thousand years? Could we really detect this after millions of years?

Suppose humanity became extinct next year and a few hundred million years from now, a new intelligent species appeared. Could they catch our existence?

---

“Perhaps the most likely type of existential risks that could constitute a Great Filter are those that arise from technological discovery. It is not far-fetched to suppose that there might be some possible technology which is such that (a) virtually all sufficiently advanced civilizations eventually discover it and (b) its discovery leads almost universally to existential disaster.”– The Doomsday Invention

Half-baked Artificial Intelligence seems a most likely candidate, because it won't be able to survive on its own but will be enough to make the species that invented it extinct.

---

Mostly, there was skepticism about the intelligence-explosion idea, which assumed answers to many unresolved questions. No one fully understands what intelligence is, let alone how it might evolve in a machine. Can it grow as Good imagined, gaining I.Q. points like a rocketing stock price? If so, what would its upper limit be? And would its increase be merely a function of optimized software design,

without the difficult process of acquiring knowledge through experience? Can software fundamentally rewrite itself without risking crippling breakdowns? No one knows. In the history of computer science, no programmer has created code that can substantially improve itself.— The Doomsday Invention

I think this *sky-rocketing I.Q.* idea is a bit far fetched. Instead we will see much more dependency, hence fragility because of A.I. When you keep paper records along with computers, that's fine, you can run the government without electricity, however when you go *completely digital*, in case of an emergency, everything is more likely to collapse.

Our dependency to robots will probably be like this. We will decide *living without robots* is *impossible* anyway and an energy crisis that make them unfeasible will send us to stone age.

Otherwise there are basic, theoretical problems with respect to *infinite IQ* (or *singularity*) idea. World is not a sum of optimization problems, real world is inherently randomized and overcoming this requires probably requires more than exponential mental power and exponential energy. I tend to think that really strong robots will be living outside earth, in deep space, because their energy needs won't be satisfied in earth.

---

Can a digital god really be contained? He imagines machines so intelligent that merely by inspecting their own code they can extrapolate the nature of the universe and of human society, and in this way outsmart any effort to contain them.— The Doomsday Invention

Our chance is that we don't really pose a threat to a *digital god* and it will probably just leave us as is, fixing our environmental damage and give us toys that will make us happy until we die.

Is fighting with humanity necessary for a *digital god*? I don't think so.

---

Last October, Tomaso Poggio, an M.I.T. researcher, gave a skeptical interview. "The ability to describe the content of an image would be one of the most intellectually challenging things of all for a machine to do," he said. "We will need another cycle of basic research to solve this kind of question." The cycle, he predicted, would take at least twenty years. A month later, Google announced that its deep-learning network could analyze an image and offer a caption of what it saw: "Two pizzas sitting on top of a stove top," or "People shopping at an outdoor market."— The Doomsday Invention

Poggio probably considered *content of an image* a bit more than Google has achieved but yes, it's advancing rapidly.