



I, for One, Welcome Our New Computer Overlords

Grady Booch

AS IT TURNS OUT, I'm in very good company: Elon Musk, Bill Gates, Clive Sinclair, Stephen Hawking, and even Steve Wozniak all agree that the use and power of AI will increase.

But beyond that, I would argue that they're all wrong.

The thing on which we disagree is a fear of the rise of extrahuman AI, an event about which each has expressed strong concern.

Those Against

In an interview with the *Guardian*, Elon said, "I think we should be very careful about artificial intelligence. If I had to guess at what our biggest existential threat is, it's probably that."¹

During a Reddit Ask Me Anything interview, Bill noted, "I am in the camp that is concerned about super intelligence. First, the machines will do a lot of jobs for us and not be super intelligent. That should be positive if we manage it well. A few

decades after that though, the intelligence is strong enough to be a concern. I agree with Elon Musk and some others on this and don't understand why some people are not concerned."²

Clive Sinclair has posited, "Once you start to make machines that are rivaling and surpassing humans with intelligence, it's going to be very difficult for us to survive."³

In an interview with the BBC, Stephen stated, "The development of full artificial intelligence could spell the end of the human race."⁴

Woz expressed these same concerns in a wonderfully playful way, stating,

I agree that the future is scary and very bad for people. If we build these devices to take care of everything for us, eventually they'll think faster than us and they'll get rid of the slow humans to run companies

more efficiently. Will we be the gods? Will we be the family pets? Or will we be ants that get stepped on? I don't know about that. But when I got that thinking in my head about if I'm going to be treated in the future as a pet to these smart machines, well I'm going to treat my own pet dog really nice.⁵

Nick Bostrom, a University of Oxford philosopher, explores the existential threat of AI in his compelling, well-reasoned book *Superintelligence: Paths, Dangers, Strategies*.⁶ I'll leave it to you, dear reader, to metabolize his entire work, but I'd offer that the essence of his argument lies in this statement:

At some point, the AI becomes better at AI design than the human programmers. Now when the AI improves itself, it improves the thing that does the improving,

*an intelligence explosion results—a rapid cascade of recursive self-improvement cycles causing the AI’s capability to soar. ... At the end of the recursive self-improvement phase, the system is strongly superintelligent.*⁶

Thus, argues Nick, some threshold exists beyond which an AI would become an existential threat to humanity; therefore, our immediate action should be to take steps to restrain, control, and otherwise limit the reach of such a superintelligent agent. This is one reason for the formation of the Future of Life Institute (<http://futureoflife.org>). Perhaps somewhat in the spirit of Pascal’s wager, Elon has provided significant funding for the institute’s research.

Those For

Before I go on, I should offer my full disclosure: I’m an early signatory to the Institute’s “Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter.”⁷ As the letter states, “We believe that research on how to make AI systems robust and beneficial is both important and timely, and that there are concrete research directions that can be pursued today.”⁷

Indeed, there’s no denying that those of us who build this technology must do so intentionally and with full consideration of our work’s potential consequences.

Rather than dive into a detailed analysis of Nick’s work, I instead refer you to others who have taken on this task. Paul Ford offers a useful take on the matter in “Our Fear of Artificial Intelligence.”⁸

I’m particularly impressed by Sebastian Benthall’s deep analysis and how he questions the potential of recalcitrance.^{9,10}

Rodney Brooks also has a say on the matter:

*I say relax everybody. If we are spectacularly lucky we’ll have AI over the next thirty years with the intentionality of a lizard, and robots using that AI will be useful tools. And they probably won’t really be aware of us in any serious way. Worrying about AI that will be intentionally evil to us is pure fear mongering. And an immense waste of time.*¹¹

Fear sells. In fact, fear fuels a great deal of contemporary news reporting and politics (if it bleeds, it leads, as the saying goes) because fear touches us on a fundamental, emotional level. It’s easier to talk about what we fear than to act on that fear. This is why it’s important to cut through the fear-mongering

things. At the worst, it’s a highly misguided fear.

Rodney observes that we’re a long, long way off, and I agree with him. He notes, “I think it is a mistake to be worrying about us developing malevolent AI anytime in the next few hundred years. I think the worry stems from a fundamental error in not distinguishing the difference between the very real advances in a particular aspect of AI and the enormity and complexity of building sentient volitional intelligence.”¹¹

These days, I’m deeply involved in creating cognitive systems. I’ve seen how the sausage is made. It’s hard work, it’s not always pretty, and although some exciting breakthroughs have occurred in perception and deep learning, we still have a long way to go.

Might a superintelligent AI emerge? In some distant future,

Fear touches us on a fundamental, emotional level.

and examine why we fear the rise of superintelligence: we fear it because it calls into question what it means to be human.

To be clear, I’m not accusing Nick, Elon, Bill, Clive, Stephen, or Woz of being fearmongers: they’re all speaking from their heart and their experience. Therefore, I accept that this fear of the rise of superintelligence isn’t irrational.

However, it’s not, as Rodney observes, an immediate fear, nor is it, in my estimation, a fear of probable

perhaps, but we must realize that the journey of crafting such an AI changes the world along the way. So, for us to project our fears through the lens of who we are today misses the reality that when such an AI forms, we won’t be the same as we are now.

What I Do Fear

That’s why I suggest that we need to stop worrying and learn to love the AI, because otherwise we’ll be distracted from a very clear and present

danger. I don't fear the rise of super-intelligent AI as do Nick, Elon, and Bill; what I do fear is the fragile software on which society relies.

We've surrendered ourselves to computing. Software isn't just eating the world, says Marc Andreessen; it's now the foundation on which modern civilization flourishes.¹² As a computing insider, I recognize that this digital edifice is exquisitely complex and fragile. The many security breaches we encounter daily are only the visible fracture lines in this infrastructure; the real fault lines lie much deeper, in the vast amounts of legacy code that run the world.

Peter Neumann gives us an accounting of some of the risks to the public that arise from computing,¹³ and I suspect his work touches only a small fraction of the tectonic forces. The loss of personal privacy, the growth of the digital divide between the haves and the have-nots, the impact of computing on employment, the disruption of industries at a speed faster than society can easily metabolize, the way that computing is changing not only social intercourse but also the way of war—these are the things that keep me awake at night, not the potential rise of a superintelligent AI of our own making.

Earlier, I quoted Clive, who said, “The development of full artificial intelligence could spell the end of the human race.”³

My immediate reaction to his point of view is this: Clive, you say that as if it's a bad thing. Perhaps our ultimate fate is the creation of a successor species that supersedes the human race. Why should we as humans expect a privileged place in the cosmos, simply because we're human? Instead, perhaps we should embrace the potential of the journey of what we might become in this coevolution of computing and humanity. I expect that we'll never notice a singularity because we'll slowly, irreversibly, inevitably become machines ourselves.

I do not fear our new computer overlords. Indeed, I welcome them. ☺

References

1. S. Gibbs, “Elon Musk: Artificial Intelligence Is Our Biggest Existential Threat,” *Guardian*, 27 Oct. 2014; www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat.
2. P. Holley, “Bill Gates on Dangers of Artificial Intelligence: ‘I Don’t Understand Why Some People Are Not Concerned,’” *Washington Post*, 29 Jan. 2015; www.washingtonpost.com/news/the-switch/wp/2015/01/28/bill-gates-on-dangers-of-artificial-intelligence-dont-understand-why-some-people-are-not-concerned.
3. “Out of Control AI Will Not Kill Us, Believes Microsoft Research Chief,” *BBC News*, 28 Jan. 2015; www.bbc.com/news/technology-31023741.
4. R. Cellan-Jones, “Stephen Hawking Warns Artificial Intelligence Could End Mankind,” *BBC News*, 2 Dec. 2014; www.bbc.com/news/technology-30290540.
5. D. Storm, “Steve Wozniak on AI: Will We Be Pets or Mere Ants to Be Squashed Our Robot Overlords?,” *Computerworld*, 25 Mar. 2015; www.computerworld.com/article/2901679/steve-wozniak-on-ai-will-we-be-pets-or-mere-ants-to-be-squashed-our-robot-overlords.html.
6. N. Bostrom, *Superintelligence: Paths, Dangers, Strategies*, Oxford Univ. Press, 2014.
7. “Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter,” Future of Life Inst.; http://futureoflife.org/AI/open_letter.
8. P. Ford, “Our Fear of Artificial Intelligence,” *MIT Technology Rev.*, 11 Feb. 2015; www.technologyreview.com/review/534871/our-fear-of-artificial-intelligence.
9. S. Benthall, “Bostrom’s Superintelligence: Definitions and Core Argument,” blog, 20 Aug. 2015; <http://digifesto.com/2015/08/20/bostroms-superintelligence-definitions-and-core-argument>.
10. S. Benthall, “The Recalcitrance of Prediction,” blog, 28 Aug. 2015; <http://digifesto.com/2015/08/28/the-recalcitrance-of-prediction>.
11. R. Brooks, “Artificial Intelligence Is a Tool, Not a Threat,” blog, 10 Nov. 2014; www.rethinkrobotics.com/blog/artificial-intelligence-tool-threat.
12. “Why Software Is Eating the World,” *Wall Street Journal*, 20 Aug. 2011; www.wsj.com/articles/SB10001424053111903480904576512250915629460.
13. “ACM Forum on Risks to the Public in Computers and Related Systems,” *Risks Digest*, vol. 28, no. 93, 3 Sept. 2015; www.csl.sri.com/users/risko/risks.txt.

GRADY BOOCH is an IBM Fellow and one of UML's original authors. He's currently developing *Computing: The Human Experience*, a major transmedia project for public broadcast. Contact him at grady@computingthehumanexperience.com and follow him on Twitter @grady_booch.



Visit the podcast for professional developers to hear in-depth interviews with some of the top minds in software engineering.

www.se-radio.net

Sponsored by





See www.computer.org/software-multimedia for multimedia content related to this article.