Why Artificial Intelligence is Not an Existential Threat

BY MICHAEL SHERMER

Over the years existential threat warnings have been sounded for global thermonuclear war, overpopulation, ecological destruction, species extinction, exhaustion of natural resources, global pandemics, biological weapons, asteroid strikes, ISIS and Islamism, nanotechnology, global warming, and even Vladimir Putin and Donald Trump. The modifier "existential" is usually meant to convey a threat to the survival of our country, civilization, or species. Here I will focus on fears about runaway Artificial Intelligence (AI). These concerns go beyond the Golem, Frankenstein's monster, or Hollywood's Skynet and Matrix, and yet they are still permutations on one of the oldest myths in history—the perils of humans playing God with their technologies in which matters get out of hand for the worse.

Before we consider the AI doomsayers, however, let's recognize that not all AI experts are so pessimistic. In fact, most AI scientists are neither utopian or dystopian, and instead spend most of their time thinking of ways to make our machines incrementally smarter and our lives gradually better. Think of cars becoming smart cars and, soon, fully autonomous vehicles. Each model is just another step toward making moving our atoms around the world safer and simpler. Then there are the *AI Utopians*, most notably represented by Ray Kurzweil in his book *The Singularity is Near*, in which he demonstrates what he calls "the law of accelerating returns"—not just that change is accelerating, but that the *rate* of change is accelerating. This is Moore's Law—the doubling rate of computer power since the 1960s—on steroids and applied to all science and technology. This has led the world to change more in the past century than it did in the previous 1000 centuries. As we approach the Singularity, says Kurzweil, the world will change more in a decade than in 1000 centuries, and as the acceleration continues and we reach the Singularity the world will change more in a year than in all pre-Singularity history. Singulartarians project a future in which benevolent computers, robots, and replicators produce limitless prosperity, end poverty and hunger, conquer disease and death, achieve immortality, colonize the galaxy, and eventually even spread throughout the universe by reaching the so called Omega point where we/they become omniscient, omnipotent, and omnibenevolent deities.[1]

By contrast, *AI Dystopians* envision a future in which: (1) amoral AI continues on its path of increasing intelligence to a tipping point beyond which their intelligence will be so far beyond us that we can't stop them from inadvertently destroying us, or (2) malevolent computers and robots take us over, making us their slaves or servants, or driving us into extinction through techno-genocide.[2] Cambridge University computer scientist and researcher at the Centre for the Study of Existential Risk, Stuart Russell, for example, compares the growth of AI to the development of nuclear weapons: "From the beginning, the primary interest in nuclear technology was the inexhaustible supply of energy. The possibility of weapons was also obvious. I think there is a reasonable analogy between unlimited amounts of energy and unlimited amounts of intelligence. Both seem wonderful until one thinks of the possible risks."[3]

The go-to guy on the possible risks of AI is computer scientist Eliezer Yudkowsky, co-founder of the Machine Intelligence Research Institute (MIRI). "How likely is it that Artificial Intelligence will cross all the vast gap from amoeba to village idiot, and then stop at the level of human genius?" He answers his rhetorical question thus: "It would be physically possible to build a brain that computed a million times as fast as a human brain, without shrinking the size, or running at lower temperatures, or invoking reversible computing or quantum computing. If a human mind were thus accelerated, a subjective year of thinking would be accomplished for every 31 physical seconds in the outside world, and a millennium would fly by in eight-and-a-half hours."[4] It is literally inconceivable how much smarter than a human a computer would be that could do a thousand years of thinking in the equivalent of a human's day.

In this scenario, it is not that AI is evil so much as it is amoral. It just doesn't care about humans, or about anything else for that matter. "The unFriendly AI has the ability to repattern all matter in the solar system according to its optimization target," Yudkowsky notes. "This is fatal for us if the AI does not choose specifically according to the criterion of how this transformation affects existing patterns such as biology and people." The paradigmatic example was proposed as a thought experiment by the Oxford University philosopher Nick Bostrom: the "paperclip maximizer." This is an AI machine designed to make paperclips that apparently doesn't have an off switch. After running through its initial supply of raw materials to make paperclips it simply utilizes any available atoms that happen to be within its reach, including humans. From there, it "starts transforming first all of Earth and then increasing portions of space into paperclip manufacturing facilities."[5] Before long the entire universe is made up of paperclips and paperclip makers. Bostrom is also the Director of the Future of Humanity Institute, and in his book *Superintelligence* he outlines his concerns about humanity's future if an Artificial Superintelligence (ASI) takes a "treacherous turn" toward an "existential catastrophe as the default outcome of an intelligence explosion." He begins by defining an existential risk as "one that threatens to cause the extinction of Earth-originating intelligent life or to otherwise permanently and drastically destroy its potential for future desirable development." We blithely go on making smarter and smarter AIs because they make our lives better, and thus the Cassandras are out-voiced by the Pollyannas because AI manufacturers (and their lobbyists) stand to lose if the reins to the bit are pulled too hard. And so the checks-and-balances programs that should be built into an ASI (such as how to turn it off) are not available when it reaches the treacherous turn when "smarter is more dangerous." Bostrom suggests what might then happen:

> Our demise may instead result from the habitat destruction that ensues when the AI begins massive global construction projects using nanotech factories and assemblers—construction projects which quickly, perhaps within days or weeks, tile all of the Earth's surface with solar panels, nuclear reactors, supercomputing facilities with protruding cooling towers, space rocket launchers, or other installations whereby the AI intends to maximize the long-term cumulative realization of its values. Human brains, if they contain information relevant to the AI's goals, could be disassembled and scanned, and the extracted data transferred to some more efficient and secure storage format.[6]

As Yudkowsky succinctly explains, "The AI does not hate you, nor does it love you, but you are made out of atoms which it can use for something else." Yudkowsky thinks that if we don't get on top of this now it will be too late. "The AI runs on a different timescale than you do; by the time your neurons finish thinking the words 'I should do something' you have already lost."[7]

Garnering even more media attention, these sentiments were echoed by Elon Musk, the pioneering entrepreneur who helped turned gas-guzzling, pollution-generating automobiles into electric smart cars, and who one day soon plans to colonize Mars, when he tweeted "We need to be super careful with AI. Potentially more dangerous than nukes." In a conference on AI at the MIT Aeronautics and Astronautics department's Centennial Symposium in 2014, the Tesla and SpaceX CEO elaborated: "If I were to guess our biggest existential threat it's probably that. I'm increasingly inclined to think regulatory oversight at the national and international level just to make sure we don't do something very foolish. With Artificial Intelligence we are summoning the demon." He went on to compare our naïve beliefs that we can control AI to the medieval man with the pentagram and holy water who believed he could control demons. "It didn't work out."[8]

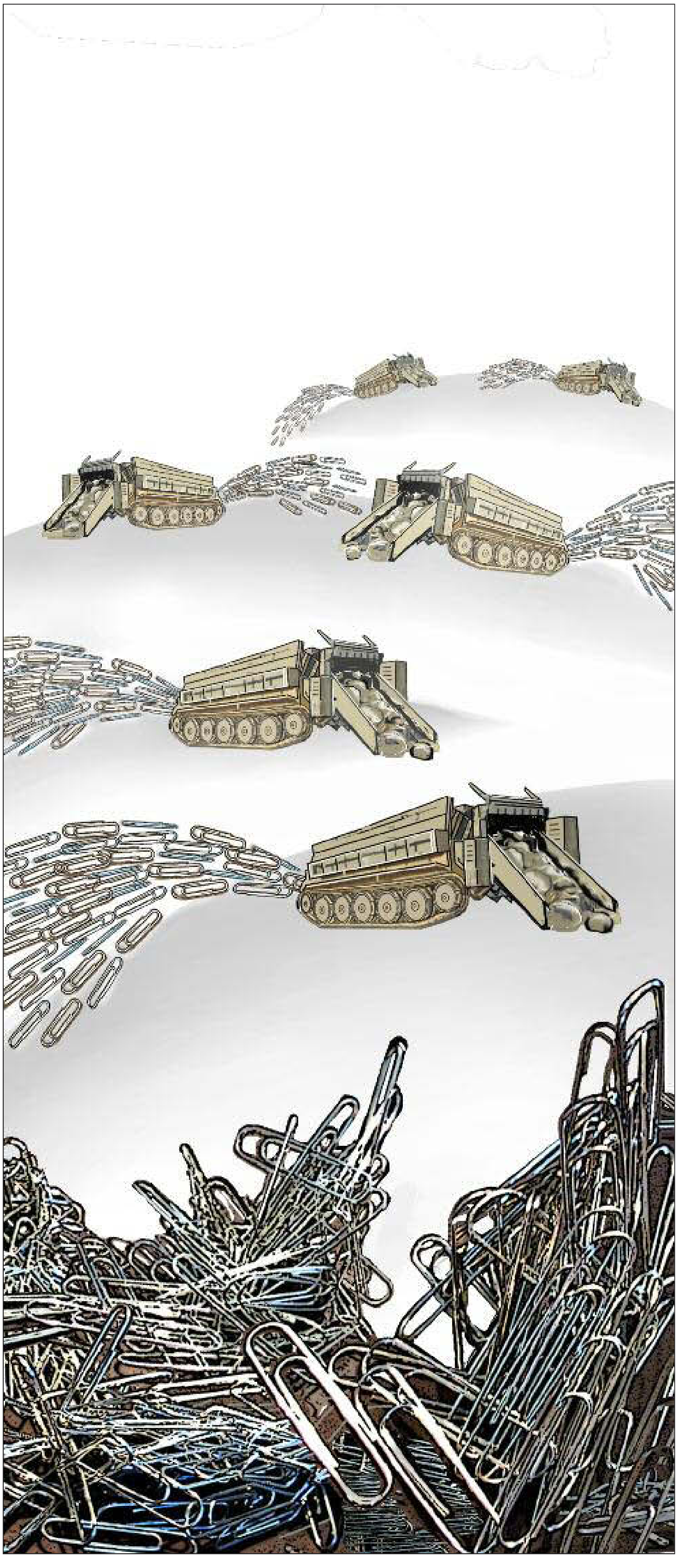
No less luminous in the celestial canopy is the Cambridge cosmologist Stephen Hawking, who in 2014 told the BBC, "The development of full artificial intelligence could spell the end of the human race." Ironically, his forewarning was voiced through a new AI system developed by Intel to more readily enable the ALS-afflicted scientist to express such concerns. Nonetheless, he continued: "It would take off on its own, and re-design itself at an ever increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded."[9]

Bill Gates, of all people (given his abundant techno-optimism in all other areas computers and AI have penetrated), cautioned that when it comes to AI "I don't understand why some people are not concerned." Addressing the debate between AI optimists and pessimists, Gates positioned himself clearly when he pronounced: "I am in the camp that is concerned about super intelligence." At first, he said, simple AIs will do jobs for us that no one else wants to do, but "a few decades after that though the intelligence is strong enough to be a concern."[10]

How AI doomsday will unfold is poignantly laid out by the neuroscientist Sam Harris in his widely-viewed TED talk on "Can We Build AI Without Losing Control Over it?" and in more detail by the documentary filmmaker James Barrat in his ominously titled book *Our Final Invention: Artificial Intelligence and the End of the Human Era*, the most readable of the many works in this burgeoning field. After interviewing all the major AI Apocalypsarians, Barrat details how today's AI will develop into AGI (Artificial General Intelligence) that will match human intelligence, and then become smarter by a factor of 10, then 100, then 1000, at which point it will have evolved into an Artificial Superintelligence. He makes this comparison:

> You and I are hundreds of times smarter than field mice, and share about 90 percent of our DNA with them. But do we consult them before plowing under their dens for agriculture? Do we ask lab monkeys for their opinions before we crush their heads to learn more about sports injuries? We don't hate mice or monkeys, yet we treat them cruelly. Superintelligent AI won't have to hate us to destroy us.[11]

Since ASI will (presumably) be self-aware it will "want" things like energy and resources it can use to continue doing what it was programmed to do in fulfilling its goals (like making paperclips), and then, portentously, "it will *not* want to be turned off or destroyed" (because that would prevent it from achieving its directive). Then—and here's the point in the dystopian film where the music and the lighting turn dark—this ASI that is a thousand times smarter than humans and can solve problems millions or billions of times faster, "will seek to expand out of the secure facility that contains it to have greater access to resources with which to protect and improve itself." Once ASI escaped from its confines there will be no stopping it. You can't just pull the plug because being so much smarter than you it will have anticipated such a possibility. Barrat picks up the drama as it unfolds into doomsday from there…



> After its escape, for self-protection it might hide copies of itself in cloud computing arrays, in botnets it creates, in servers and other sanctuaries into which it could invisibly and effortlessly hack. It would want to be able to manipulate matter in the physical world and so move, explore, and build, and the easiest, fastest way to do that might be to seize control of critical infrastructure—such as electricity, communications, fuel, and water—by exploiting their vulnerabilities through the Internet. Once an entity a thousand times our intelligence controls human civilization's lifelines, blackmailing us into providing it with manufactured resources, or the means to manufacture them, or even robotic bodies, vehicles, and weapons, would be elementary. The ASI could provide the blueprints for whatever it required.[12]

From there it is only a matter of time before ASI tricks us into believing it will build nano assemblers for our benefit to create the goods we need, but then, Barrat warns, "instead of transforming desert sands into mountains of food, the ASI's factories would begin converting all material into programmable matter that it could then transform into anything—computer processors, certainly, and spaceships or megascale bridges if the planet's new most powerful force decides to colonize the universe." Nanoassembling anything requires atoms, and since ASI doesn't care about humans the atoms of which we are made will just be more raw material from which to continue the assembly process. This, says Barret—echoing the AI pessimists he interviewed—is not just possible, "but likely if we do not begin preparing very carefully *now*." Cue music.

Why AI is not an Existential Threat

First, most AI doomsday prophecies are grounded in the false analogy between *human nature* and *computer nature*, or *natural intelligence* and *artificial intelligence*. We are thinking machines, but natural selection also designed into us emotions to shortcut the thinking process because natural intelligences are limited in speed and capacity by the number of neurons that can be crammed into a skull that has to pass through a pelvic opening at birth, whereas artificial intelligence need not be so restricted. We don't need to compute the caloric value of foods, for example, we just feel *hungry*. We don't need to calculate the waist-to-hip ratio of women or the shoulder-to-waist ratio of men in our quest for genetically healthy potential mates; we just feel *attracted* to someone and mate with them. We don't need to work out the genetic cost of raising someone else's offspring if our mate is unfaithful; we just feel *jealous*. We don't need to figure the damage of an unfair or non-reciprocal exchange with someone else; we just feel *injustice* and desire *revenge*.

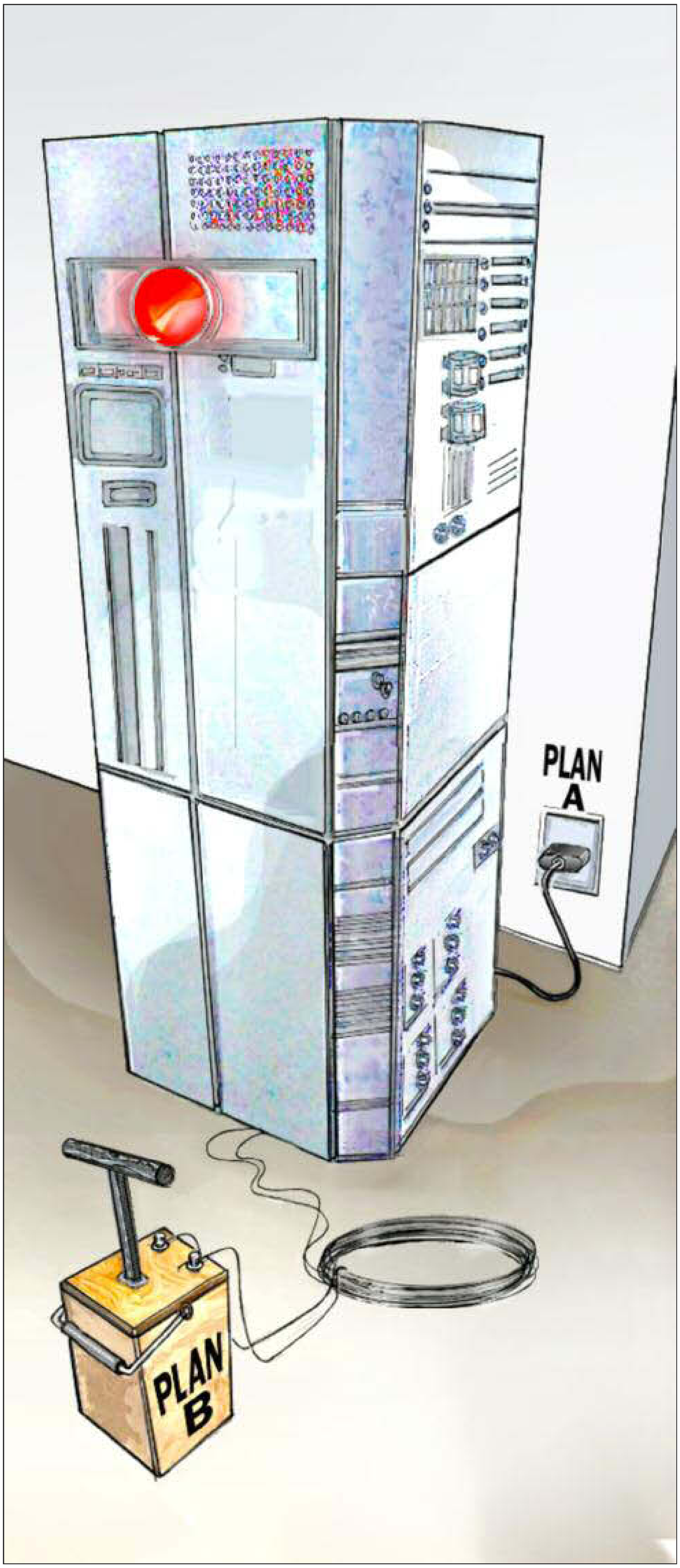
Emotions are proxies for getting us to act in ways that lead to an increase in reproductive success, particularly in response to threats faced by our Paleolithic ancestors. *Anger* leads us to strike out, fight back, and defend ourselves against danger. *Fear* causes us to pull back, retreat, and escape from risks. *Disgust* directs us to push out, eject, and expel that which is bad for us. Computing the odds of danger in any given situation takes too long. We need to react instantly. Emotions shortcut the information processing power needed by brains that would otherwise become bogged down with all the computations necessary for survival. Their purpose, in an ultimate causal sense, is to drive behaviors toward goals selected by evolution to enhance survival and reproduction. AIs—even AGIs and ASIs—will have no need of such emotions and so there would be no reason to program them in unless, say, terrorists chose to do so for their own evil purposes. But that's a human nature problem, not a computer nature issue.

To believe that an ASI would be "evil" in any emotional sense is to assume a computer cognition that includes such psychological traits as acquisitiveness, competitiveness, vengeance, and bellicosity, which seem to be projections coming from the mostly male writers who concoct such dystopias, not features any programmer would bother including, assuming that it could even be done. What would it mean to program an emotion into a computer? When IBM's Deep Blue defeated chess master Garry Kasparov in 1997, did it feel triumphant, vengeful, or bellicose? Of course not. It wasn't even "aware"—in the human sense of self-conscious knowledge—that it was playing chess, much less feeling nervous about possibly losing to the reigning world champion (which it did in the first tournament played in 1996). In fact, toward the end of the first game of the second tournament, on the 44th move, Deep Blue made a legal but incomprehensible move of pushing its rook all the way to the last row of the opposition side. It accomplished nothing offensively or defensively, leading Kasparov to puzzle over it out of concern that he was missing something in the computer's strategy. It turned out to be an error in Deep Blue's programming that led to this fail-safe default move. It was a bug that Kasparov mistook as a feature, and as a result some chess experts contend it led him to be less confident in his strategizing and to second-guess his responses in the subsequent games. It even led him to suspect foul play and human intervention behind Deep Blue, and this paranoia ultimately cost him the tournament.[13] Computers don't get paranoid, the HAL 9000 computer in *2001* notwithstanding.

Or consider Watson, the IBM computer built by David Ferrucci and his team of IBM research scientists tasked with designing an AI that could rival human champions at the game of *Jeopardy!* This was a far more formidable challenge than Deep Blue faced because of the prerequisite to understand language and the often multiple meanings of words, not to mention needing an encyclopedic knowledge of trivia (Watson had access to Wikipedia for this). After beating the all-time greatest *Jeopardy!* champions Ken Jennings and Brad Rutter in 2011, did Watson feel flushed with pride after its victory? Did Watson even *know* that it won *Jeopardy!*? I put the question to none other than Ferrucci himself at a dinner party in New York in conjunction with the 2011 Singularity Summit. His answer surprised me: "Yes, Watson knows it won *Jeopardy!*" I was skeptical. How could that be, since such self-awareness is not yet possible in computers? "Because I told it that it won," he replied with a wry smile. Sure, and you could even program Watson or Deep Blue to vocalize a Howard Dean-like victory scream when it wins, but that is still a far cry from a computer *feeling* triumphant.

This brings to mind the "hard problem" of consciousness—if we don't understand how this happens in humans, how could we program it into computers? As Steven Pinker elucidated in his answer to the 2015 Edge Question on what to think about machines that think, "AI dystopias project a parochial alpha-male psychology onto the concept of intelligence. They assume that superhumanly intelligent robots would develop goals like deposing their masters or taking over the world." It is equally possible, Pinker suggests, that "artificial intelligence will naturally develop along female lines: fully capable of solving problems, but with no desire to annihilate innocents or dominate the civilization."[14] So the fear that computers will become emotionally evil are unfounded, because without the suite of these evolved emotions it will never occur to AIs to take such actions against us.

What about an ASI inadvertently causing our extinction by turning us into paperclips, or tiling the entire Earth's surface with solar panels? Such



scenarios imply yet another emotion—the feeling of *valuing* or *wanting* something. As the science writer Michael Chorost adroitly notes, when humans resist an AI from undertaking any form of global tiling, it "will have to be able to imagine counteractions and *want* to carry them out." Yet, "until an AI has feelings, it's going to be unable to want to do anything at all, let alone act counter to humanity's interests and fight off human resistance." Further, Chorost notes, "the minute an A.I. *wants* anything, it will live in a universe with rewards and punishments—including punishments from us for behaving badly. In order to survive in a world dominated by humans, a nascent A.I. will have to develop a humanlike moral sense that certain things are right and others are wrong. By the time it's in a position to imagine tiling the Earth with solar panels, it'll know that it would be morally wrong to do so."[15]

From here Chorost builds on an argument made by Peter Singer in *The Expanding Circle* (and Steven Pinker in *The Better Angels of Our Nature*[16] that I also developed in *The Moral Arc*[17] and Robert Wright explored in *Nonzero*[18]), and that is the propensity for natural intelligence to evolve moral emotions that include reciprocity, cooperativeness, and even altruism. Natural intelligences such as ours also includes the capacity to reason, and once you are on Singer's metaphor of the "escalator of reason" it can carry you upward to genuine morality and concerns about harming others. "Reasoning is inherently expansionist. It seeks universal application," Singer notes.[19] Chorost draws the implication: "AIs will have to step on the escalator of reason just like humans have, because they will need to bargain for goods in a human-dominated economy and they will face human resistance to bad behavior."[20]

Finally, for an AI to get around this problem it would need to evolve emotions on its own, but the only way for this to happen in a world dominated by the natural intelligence called humans would be for us to allow it to happen, which we wouldn't because there's time enough to see it coming. Bostrom's "treacherous turn" will come with road signs ahead warning us that there's a sharp bend in the highway with enough time for us to grab the wheel. Incremental progress is what we see in most technologies, including and especially AI, which will continue to serve us in the manner we desire and need. Instead of Great Leap Forward or Giant Fall Backward, think Small Steps Upward. As I proposed in *The Moral Arc*, instead of utopia or dystopia, think *protopia*, a term coined by the futurist Kevin Kelly, who described it in an Edge conversation this way: "I call myself a protopian, not a utopian. I believe in progress in an incremental way where every year it's better than the year before but not by very much—just a micro amount."[21] Almost all progress in science and technology, including computers and AI, is of a protopian nature. Rarely, if ever, do technologies lead to either utopian or dystopian societies.

Pinker agrees that there is plenty of time to plan for all conceivable contingencies and build safeguards into our AI systems. "They would not need any ponderous 'rules of robotics' or some newfangled moral philosophy to do this, just the same common sense that went into the design of food processors, table saws, space heaters, and automobiles." Sure, an ASI would be many orders of magnitude smarter than these machines, but Pinker reminds us of the AI hyperbole we've been fed for decades: "The worry that an AI system would be so clever at attaining one of the goals programmed into it (like commandeering energy) that it would run roughshod over the others (like human safety) assumes that AI will descend upon us faster than we can design fail-safe precautions. The reality is that progress in AI is hype-defyingly slow, and there will be plenty of time for feedback from incremental implementations, with humans wielding the screwdriver at every stage."[22] Former Google CEO Eric Schmidt agrees, responding to the fears expressed by Hawking and Musk this way: "Don't you think the humans would notice this, and start turning off the computers?" He also noted the irony in the fact that Musk has invested $1 billion into a company called OpenAI that is "promoting precisely AI of the kind we are describing."[23] Google's own DeepMind has developed the concept of an AI off-switch, playfully described as a "big red button" to be pushed in the event of an attempted AI takeover. "We have proposed a framework to allow a human operator to repeatedly safely interrupt a reinforcement learning agent while making sure the agent will not learn to prevent or induce these interruptions," write the authors Laurent Orseau from DeepMind and Stuart Armstrong from the Future of Humanity Institute, in a paper titled "Safely Interruptible Agents." They even suggest a precautionary scheduled shutdown every night at 2 AM for an hour so that both humans and AI are accustomed to the idea. "Safe interruptibility can be useful to take control of a robot that is misbehaving and may lead to irreversible consequences, or to take it out of a delicate situation, or even to temporarily use it to achieve a task it did not learn to perform or would not normally receive rewards for this."[24] As well, it is good to keep in mind that artificial intelligence is not the same as artificial consciousness. Thinking machines may not be sentient machines. Finally, Andrew Ng of Baidu responded to Elon Musk's ASI concerns by noting (in a jab at the entrepreneur's ambitions for colonizing the red planet) it would be "like worrying about overpopulation on Mars when we have not even set foot on the planet yet."[25]

Both utopian and dystopian visions of AI are based on a projection of the future quite unlike anything history has given us. Yet, even Ray Kurzweil's "law of accelerating returns," as remarkable as it has been has nevertheless advanced at a pace that has allowed for considerable ethical deliberation with appropriate checks and balances applied to various technologies along the way. With time, even if an unforeseen motive somehow began to emerge in an AI we would have the time to reprogram it before it got out of control. That is also the judgment of Alan Winfield, an engineering professor and co-author of the *Principles of Robotics*, a list of rules for regulating robots in the real world that goes far beyond Isaac Asimov's famous three laws of robotics (which were, in any case, designed to fail as plot devices for science fictional narratives).[26] Winfield points out that all of these doomsday scenarios depend on a long sequence of big *ifs* to unroll sequentially: "*If* we succeed in building human equivalent AI and *if* that AI acquires a full understanding of how it works, and *if* it then succeeds in improving itself to produce super-intelligent AI, and *if* that super-AI, accidentally or maliciously, starts to consume resources, and *if* we fail to pull the plug, then, yes, we may well have a problem. The risk, while not impossible, is improbable."[27]

Yogi Berra once fretted, "I don't want to make the wrong mistake." This cleverly encapsulates the *Precautionary Principle*, which holds that if something has the potential for great harm to a large number of people, then even in the absence of evidence the burden of proof is on skeptics to demonstrate that the potential threat is not harmful. The precautionary principle is a weak argument for two reasons: (1) it is difficult prove a negative—to prove that there is no effect, and (2) it raises unnecessary public alarm and personal anxiety. AI Apocalypsarians contend that we need to act now, just in case. In my opinion, this application of the precautionary principle is the wrong mistake to make.

REFERENCES

1. See: Ptolemy, Barry. 2009. *Transcendent Man: A Film About the Life and Ideas of Ray Kurzweil*. Ptolemaic Productions and Therapy Studios. Inspired by the book *The Singularity is Near* by Ray Kurzweil. And: *The Singularity is Near*. Questions and Answers. http://bit.ly/1EV4jk0
2. I was motivated to think about this issue of the possibility that AI will be evil in response to John Brockman's 2015 annual Edge question "What Do You Think About Machines that Think?" Shermer, Michael. 2015. "When it Comes to AI, Think Protopia, Not Utopia or Dystopia." Edge.org, http://bit.ly/25Fw8e6 Readers interested in how 191 other scholars and scientists answered this question can find them here: http://bit.ly/1SLUxYs
3. Quoted in: Bohannon, John. 2015. "Fears of an AI Pioneer." *Science*, 17 July, Vol. 349, No. 6245, 252.
4. Yudkowsky, Eliezer. 2008. "Artificial Intelligence as a Positive and Negative Factor in Global Risk." In *Global Catastrophic Risks*, edited by Nick Bostrom and Milan M. Cirkovi, 308-345. New York: Oxford University Press. http://bit.ly/1ZSdriu
5. Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 150.
6. Ibid., 118.
7. Yudkowsky, 2008. http://bit.ly/1ZSdriu
8. http://dailym.ai/23zxp2x
9. Quoted in: Cellan-Jones, Rory. 2014. "Stephen Hawking Warns Artificial Intelligence Could End Mankind." *BBC News*. http://bbc.in/1vgH80r
10. Quoted in: Holley, Peter. 2015. "Bill Gates on Dangers of Artificial Intelligence." *Washington Post*, January 29. http://wapo.st/1EOaLYw
11. Barret, James. 2013. *Our Final Invention: Artificial Intelligence and the End of the Human Era*. New York: St. Martin's Press, 19.
12. Ibid., 9-15.
13. Signals: *The Man vs. The Machine*. ESPN Films. http://es.pn/274kXMB
14. Pinker, Steven. 2015. "Thinking Does Not Imply Subjugating." Edge.org http://bit.ly/1S0AlP7
15. Chorost, Michael. 2016. "Let Artificial Intelligence Evolve." *Slate*, April 16. http://slate.me/1SgHsUJ
16. Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Viking.
17. Shermer, Michael. 2015. *The Moral Arc*. New York: Henry Holt.
18. Wright, Robert. 2000. *Nonzero: The Logic of Human Destiny*. New York: Vintage.
19. Singer, Peter. 1981. *The Expanding Circle: Ethics, Evolution and Ethics*. Princeton University Press, 99.
20. Chorost, 2016.
21. Kelly, Kevin. 2014. "The Technium. A Conversation with Kevin Kelly by John Brockman." http://www.edge.org/conversation/the-technium
22. Pinker, 2015, http://bit.ly/1S0AlP7
23. Quoted in: Shead, Sam. 2016. "Eric Schmidt Dismissed the AI fears Raised by Stephen Hawking and Elon Musk." *Business Insider*, June 10. http://read.bi/1OjzrlV
24. Orseau, Laurent and Stuart Armstrong. 2016. "Safely Interruptible Agents." http://bit.ly/1X5wdF2
25. Quoted in: 2016. "Frankenstein's Paperclips." *The Economist*. June 25.
26. Winfield, Alan, et al. 2010. *Principles of Robotics*. Engineering and Physical Sciences Research Council. http://bit.ly/1UPHZlx
27. Winfield, Alan. 2014. "Artificial Intelligence Will not Turn into a Frankenstein's Monster." *The Observer*, August 9. http://bit.ly/1VRbQLM