

Are We Losing Control?



Edward A. Lee

Abstract This chapter challenges the predominant assumption that humans shape technology using top-down, intelligent design, suggesting that technology should instead be viewed as the result of a Darwinian evolutionary process where humans are the agents of mutation. Consequently, we humans have much less control than we think over the outcomes of technology development.

Rapid change breeds fear. With its spectacular rise from the ashes in the last decade, we fear that AI may replace most white collar jobs (Ford 2015); that it will learn to iteratively improve itself into a superintelligence that leaves humans in the dust (Barrat 2013; Bostrom 2014; Tegmark 2017); that it will fragment information so that humans divide into islands of disjoint sets of truths (Lee 2020); that it will supplant human decision-making in health care, finance, and politics (Kelly 2016); that it will cement authoritarian powers, tracking every move of their citizens and shaping their thoughts (Lee 2018); and that the surveillance capitalists' monopolies, which depend on AI, will destroy small business and swamp entrepreneurship (Zuboff 2019).

Surely, today, we still retain a modicum of control. At the very least, we can still pull the plug. Or can we? The technology underlying these risks is made by humans, so why can't we control the outcomes? We have the power to design and to regulate, don't we? So why are we trying so desperately to catch up with yesterday's disasters while today's just fester? The very technology that threatens us is also the reason we are successfully feeding most of the 7.8 billion humans on this meager planet and have lifted billions out of poverty in the last decades. Giving us pause, however, Albert Einstein famously said, "We cannot solve our problems with the same thinking we used when we created them."

Knowledge is at the root of technology, information is at the root of knowledge, and today's technology makes information vastly more accessible than it has ever

E. A. Lee (✉)
University of California, Berkeley, CA, USA
e-mail: eal@berkeley.edu

been. Shouldn't this help us solve our problems? The explosion of AI feeds the tsunami, turning every image, every text, and every sound into yet more information, flooding our feeble human brains. We can't absorb the flood without curation, and curation of information is increasingly being done by AIs. Every subset of the truth is only a partial truth, and curated information includes, necessarily, a subset. Since our brains can only absorb a tiny subset of the flood, everything we take in is at best a partial truth. The AIs, in contrast, seem to have little difficulty with the flood. To them, it is the food that strengthens, perhaps leading to that feared runaway feedback loop of superintelligence that sidelines humans into irrelevance.

The question I address here is, "Are we going to lose control?" You may find my answer disturbing.

First, in posing this question, what do we mean by "we"? Do we mean "humanity," all 7.8 billion of us? The idea of 7.8 billion people collectively controlling anything is patently absurd, so that must not be what we mean. Do we mean the engineers of Silicon Valley? The investors on Wall Street? The politicians who feed off the partial truths and overt lies?

Second, what do we mean by "control"? Is it like steering a car on a network of roads, or is it more like steering a car while the map emerges and morphs into unexpected dead ends, underpasses, and loops? If we are steering technology, then every turn we take changes the terrain that we have to steer over in unexpected ways.

I am an engineer. In my own small way, I contribute to the problem by writing software, some of which has small influences on our ecosystem. For most of my 40 years doing this, I harbored a "creationist" illusion that the things I designed were my own personal progeny, the pure result of my deliberate decisions, my own creative output. I have since realized that this is a bit like thinking that the bag of groceries that I just brought back from the supermarket is my own personal accomplishment. It ignores centuries of development in the technology of the car that got me there and back, agriculture that delivered the incredible variety of fresh food to the store, the economic system that makes all of this affordable, and many other parts of the socio-cultural backdrop against which my meager accomplishment pales.

In my recent book (Lee 2020), I coin the term "digital creationism" for the idea that technology is the result of top-down intelligent design. This principle assumes that every technology is the outcome of a deliberate process, where every aspect of a design is the result of an intentional, human decision. I now know, 40 years later, that this is not how it happens. Software engineers are more the agents of mutation in a Darwinian evolutionary process. The outcome of their efforts is shaped more by the computers, networks, software tools, libraries, programming languages, and other programs they use than by their deliberate decisions. And the success and further development of their product is determined as much or more by the cultural milieu into which they launch their "creation" than by their design decisions.

The French philosopher known as Alain (whose real name was Émile-Auguste Chartier), wrote, about fishing boats in Brittany:

Every boat is copied from another boat. ... Let's reason as follows in the manner of Darwin. It is clear that a very badly made boat will end up at the bottom after one or two voyages and thus never be copied. ... One could then say, with complete rigor, that it is the sea herself who fashions the boats, choosing those which function and destroying the others. (Rogers and Ehrlich 2008)

Boat designers are agents of mutation, and sometimes their mutations result in a badly made boat. From this perspective, perhaps Facebook has been fashioned more by teenagers than by software engineers.

More deeply, digital technology *coevolves* with humans. Facebook changes its users, who then change Facebook. For software engineers, the tools we use, themselves earlier outcomes of software engineering, shape our thinking. Think about how IDEs¹ (such as Eclipse or Visual Studio Code), message boards (such as Stack Overflow), libraries (such as the Standard Template Library), programming languages (e.g., Scala, Rust, and JavaScript), and Internet search (such as Google or Bing) affect the outcome of our software. These tools have more effect on the outcome than all of our deliberate decisions.

Today, the fear and hype around AI taking over the world and social media taking down democracy has fueled a clamor for more regulation. But if I am right about coevolution, we may be going about the project of regulating technology all wrong. Why have privacy laws, with all their good intentions, done little to protect our privacy? They have only overwhelmed us with small-print legalese and annoying pop-ups giving us a choice between “accept our inscrutable terms” and “go away.” Do we expect new regulations trying to mitigate fake news or to prevent insurrections from being instigated by social media to be any more effective?

Under the principle of digital creationism, bad outcomes are the result of unethical actions by individuals, for example by blindly following the profit motive with no concern for societal effects. Under the principle of coevolution, bad outcomes are the result of the “procreative prowess” (Dennett 2017) of the technology itself. Technologies that succeed are those that more effectively propagate. The individuals we credit with (or blame for) creating those technologies certainly play a role, but so do the users of the technologies and their whole cultural context. Under this perspective, Facebook users bear some of the blame, along with Mark Zuckerberg, for distorted elections. They even bear some of the blame for the design of Facebook software that enables distorted elections. If they were happy to pay for social networking, for example, an entirely different software design may have emerged.

Under digital creationism, the purpose of regulation is to constrain the individuals who develop and market technology. In contrast, under coevolution, constraints can be about the *use* of technology, not just its design and the business of selling it. The purpose of regulation becomes to nudge the process of both technology and cultural evolution through incentives and penalties. Nudging is probably the best we can

¹Integrated development environments (IDEs) are computer programs that assist programmers by parsing their text as they type, coloring text by function, identifying errors and potential flaws in code style, suggesting insertions, and transforming code through refactoring.

hope for. Evolutionary processes do not yield easily to control because the territory over which we have to navigate keeps changing.

Perhaps privacy laws have been ineffective because they are based on digital creationism as a principle. These laws assume that changing the behavior of corporations and engineers will be sufficient to achieve privacy goals (whatever those are for you). A coevolutionary perspective understands that users of technology will choose to give up privacy even if they are explicitly told that their information will be abused. We are repeatedly told exactly that in the fine print of all those privacy policies we don't read, and, nevertheless, our kids get sucked into a media milieu where their identity gets defined by a distinctly non-private online persona.

If technology is defining culture while culture is defining technology, we have a feedback loop, and intervention at any point in the feedback loop can change the outcomes. Hence, it may be just as effective to pass laws that focus on educating the public, for example, as it is to pass laws that regulate the technology producers. Perhaps if more people understood that Pokémon GO is a behavior-modification engine, they would better understand Niantic's privacy policy and its claim that their product, Pokémon GO, has no advertising. Establishments pay Niantic for placement of a Pokémon nearby to entice people to visit them (Zuboff 2019). Perhaps a strengthening of libel laws, laws against hate speech, and other refinements to first-amendment rights should also be part of the remedy.

I believe that, as a society, we can do better than we are currently doing. The risk of an Orwellian state (or perhaps worse, a corporate Big Brother) is very real. It has happened already in China. We will not do better, however, until we abandon digital creationism as a principle. Outlawing specific technology developments will not be effective, and breaking up monopolies could actually make the problem worse by accelerating mutations. For example, we may try to outlaw autonomous decision-making in weapons systems and banking, but as we see from election distortions and Pokémon GO, the AIs are very effective at influencing human decision-making, so putting a human in the loop does not necessarily help. How can a human who is, effectively, controlled by a machine somehow mitigate the evilness of autonomous weapons?

When I talk about educating the public, many people immediately gravitate to a perceived silver bullet, that we should teach ethics to engineers. But I have to ask, if we assume that all technologists behave ethically (whatever your meaning of that word), can we conclude that bad outcomes will not occur? This strikes me as naïve. Coevolutionary processes are much too complex.

This chapter is my small contribution to the digital humanism initiative, a movement that seeks a more human-centric approach to technology. This initiative makes it imperative for intellectuals of all disciplines to step up and take seriously humanity's dance with technology. That our limited efforts to rein in the detrimental effects of digital technology have been mostly ineffective underscores our weak understanding of the problem. We need humanists with a deeper understanding of technology, technologists with a deeper understanding of the humanities, and policy makers drawn from both camps. We are quite far from that goal today.

Returning to the original question, are we losing control? The answer is “no.” We never had control, and we can’t lose what we don’t have. This does not mean we should give up, however. We *can* nudge the process, and even a supertanker can be redirected by gentle nudging.

References

- Barrat, J. (2013). *Our Final Invention: Artificial Intelligence and the End of the Human Era*, St. Martin’s Press.
- Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK, Oxford University Press.
- Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*, W. W. Norton and Company.
- Ford, M. (2015). *Rise of the Robots -- Technology and the Threat of a Jobless Future*. New York, Basic Books.
- Kelly, K. (2016). *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*. New York, Penguin Books.
- Lee, E. A. (2020). *The Coevolution: The Entwined Futures of Humans and Machines*. Cambridge, MA, MIT Press.
- Lee, K.-F. (2018). *Super-Powers: China, Silicon Valley, and the New World Order*. New York, Houghton Mifflin Harcourt Publishing Company.
- Rogers, D. S. and P. R. Ehrlich (2008). “Natural Selection and Cultural Rates of Change.” *Proceedings of the National Academy of Sciences of the United States of America* 105(9): 3416-3420.
- Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York, Alfred A. Knopf.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs , Hachette Book Group.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

