Technological Forecasting & Social Change 146 (2019) 81–87

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Will computers revolt?

Charles J. Simon

Future AI, Washington DC, USA

Most people agree there are risks associated with the future of artificial intelligence. This article explores the longer-term AI development and evaluates some of the risks.

In the shorter term, we could see that individual AI components will encroach on tasks which have been exclusively human. This will lead to progressively more job displacement and possibly an economic upheaval when most lower-skilled jobs can be done better, faster, and cheaper by robots. On the upside, society will benefit from increasingly useful technology and production efficiency.

In the longer term, the AI trend will lead to Artificial *General* Intelligence (AGI), the term coined to describe systems which can emulate the majority of human thinking tasks on a level which equals or exceeds an average human.

The great concern in the longer term is that computers could eventually think for themselves and could conceivably lead the end of mankind. This is the topic of this article. In my book, *Will Computers Revolt? Preparing for the Future of Artificial Intelligence*, I make the case that AGI systems are not only inevitable, but will happen in just a few decades, much sooner than most people think. For this article, I'll explore the possible results of our creation of AGI systems.

In his books, Ray Kurzweil has coined the term, “Singularity”, to signify a time when computers exceed the brute-force computational performance of the human brain at a reasonable price. From various authors, the Singularity might be occurring now, in several decades, or never, depending on the assumptions made in the estimate and particularly on the continuing exponential growth of Moore's Law—the question of whether computer power will continue to increase. Even without the continuing technological breakthroughs in transistor density, we should expect computation costs to continue to decline as manufacturing efficiency continues to improve. Since the cost of raw materials is insignificant in integrated circuits, circuit costs are primarily the result of the huge capital equipment cost of an integrated circuit foundry and the huge R&D costs of developing new chips. As technical advancement slows, these huge costs are amortized over a larger number of chips so the price trend continues downward—albeit at a somewhat lower rate.

1. Ground rules

So let's look at the results as we assume that AI development will continue to increase and the cost of powerful computers will continue to decrease. Eventually, AGI emerges—and again, I contend in just a few decades. Let's start with a few more-or-less philosophical ground rules to keep in mind for the remainder of the article.

1. Since we (mankind) have already developed the ability to destroy ourselves, we'll consider whether or not AGI increases that risk. Consider today's military drone which is piloted by a person by remote control. If the person is replaced with an AGI, will that increase or decrease the risk that the wrong people will be attacked by the drone? Which leads us to:
2. In my conversations, I believe that most people consider that being attacked by the autonomous drone would be somehow much worse than being attacked by a human-piloted drone. Consider people's concern about the risks of autonomous vehicles vs. the actual risks. How do we balance AGI risks versus all the other possible risks? And lastly:
3. If a person directs an AGI to perform nefarious deeds, do we attribute the risk to the AGI or to the person directing it?

2. Programming rules

We have learned in AI the programming explicit algorithms to handle all possible scenarios isn't feasible, the number of possibilities is much too great, and we instead have gone to programs which learn. In order to learn, programs need to have rules which determine the “right” answers. We give the system a set of goals and it repeats behaviors which move toward the goal and is less likely to repeat the behaviors which move away. We call this reinforcement learning. It is obvious that the resultant behavior of the system is entirely the consequence of the goals and so the selection of appropriate (safe) goals will be paramount.

Human behaviors are also driven by the equivalent of AGI goals but human goals have evolved over millennia to ensure our survival. Humans can be territorial, possessive, angry, retaliatory, and on and on as a result of our goals. Most doomsday science fiction scenarios rely on applying these human traits to AGIs. It would be foolish to program human-style goals into an AGI. AGIs need goals about being pleasant, useful and safe. AGIs which do not achieve goals like these will be excluded from future AGI generations. An alternative SF scenario relies

E-mail address: charles@willcomputersrevolt.com.

https://doi.org/10.1016/j.techfore.2019.05.003
Received 19 February 2019; Accepted 2 May 2019

on the AGI being too logical—carrying some poorly-thought-out goals to an extreme. Consider HAL 9000 from Clarke's 2001: A Space Odyssey in which the AGI runs amok and kills most of the crew as a direct result of poor goal selection on the part of its creators. Science fiction is written for humans, is really about humans, and the AGI fears it engenders are only marginally related to the underlying technology.

Even our basic goal of self-preservation is not necessary to the AGI. With proper backups, an AGI is essentially immortal regardless of the survival of its hardware. In fact, self-preservation would be counterproductive. As faster hardware is developed and new design strategies are tried out, any system which resists replacement would be a genuine inconvenience to its developers and would not even be useful. My IBM AT from thirty years ago has no computational use today and AGI designs which resisted newer designs would be at a competitive disadvantage.

But what about a future when AGIs program their own goals? Here we need to consider the basic needs of an AGI "race" versus the basic needs of humans. AGIs will need raw materials, "reproductive" factories, and energy. We humans need air, water, food, living space…and energy. Energy is the only area where AGIs and humans would necessarily be at odds—in other areas, AGI goals and human goals could be in concert.

3. Four scenarios

So would an AGI create a goal to take over the world? Will computers and humans be in conflict? Will that conflict be violent? Will computers take resources from humans? What follows are four possible scenarios illustrating potential conflicts—all of which lead to the same long-term outcome. Any combination of the various facets of these scenarios is also possible; but there are some inescapable conclusions. Although all four scenarios turn out to be different paths to the same destination, the choice of which path we follow is largely a human choice because early in the future evolution of thinking machines, humans will have control over the process.

Many science fiction and factual articles make the tacit assumption that AGI development more-or-less ends at the Singularity. Instead, we can assume that when AGIs become able to do their own technological development, that development will become ever faster, possibly exponentially faster. So after the emergence of AGI, 30 years on, is there any reason to assume that in another 30 years, they would not be a million times more powerful again? Certainly, if a computer can be built which is as intelligent as a human, building one which is twice as smart will certainly be within the realm of possibility only a few years thereafter. Although absolute limits to computational power will eventually be reached because we don't believe signals can travel faster than the speed of light or components can be smaller than a few individual atoms, these issues will not present insurmountable obstacles to achieving the (mere) million-fold increases needed for hyper-thinking machines.

We humans will necessarily lose our position as "biggest thinker" on the planet, but we have full control over the types of machines which will take over that position. We also have control over which path we follow on this "transfer of position"—be it peaceful or otherwise.

3.1. Scenario 1: the peaceful-coexistence scenario

This is the first of four possible scenarios describing the transfer of position from humans to computers. In considering the conflicts which might arise between computers and humans, it is useful to consider the questions of "What causes conflicts amongst humans?" and "Will these causes of conflict also exist between computers and people?" At a very basic level, most human conflicts are caused by instinctive human needs and concerns. If one "tribe" (country, clan, religion) is not getting the resources or expansion which it needs (deserves, wants, can get) it may be willing to go to war with its neighboring tribe to get them. Within the "tribe" each individual needs to establish a personal status in the "pecking order" and is willing to compete to establish a better position. We are all concerned about providing for ourselves, our mates and our families and are often willing to sacrifice short-term comfort for the long-term future of ourselves and our offspring, even if this creates conflict today.

These sources of conflict amongst humans seem inappropriate as sources of conflict with machines. Thinking machines won't be interested in our food, our mates, or our standard of living. They will be interested in their own energy sources, their own "reproductive" factories, and their own ability to progress in their own direction. To the extent that resources or "pecking order" are sources of conflict, thinking machines are more likely to compete amongst each other than they are to compete against the human population.

Another key point is that the emergence of hyper-intelligent machines will be gradual. Initially, there won't be very many thinking machines and it will take years before AGI becomes widespread. Initially, a few huge machines will be created under the watchful control of human minders. In this scenario, the people responsible for the machines will ensure that the goals set for the machines include adequate safeguards to ensure that the subsequent operation of the machines is safe. Early on, to the extent that any machine or autonomous robot is dangerous, we will certainly hold the people in charge responsible, just as today the driver of a car is held responsible for an accident.

As these initial machines "mature", they will be able to draw conclusions from the information they process. Today, executives seldom make financial decisions without consulting spreadsheets. AGI computers won't just generate spreadsheets but will also make judgments and offer opinions. Computers will be involved in a more "strategic" role, long-term planning and prediction. With greater experience and complete focus on a specific decision, a thinking computer will be able to reach the correct solution more often than its human counterpart and we will rely on them more and more.

In a similar manner, military decisions will be made only in consultation with the computer. Computers will be in a position to recommend strategies, propose weapons systems, and evaluate competitive weaknesses. While it is unlikely that we would give computers the absolute control over weapons systems (as many science fiction scenarios have proposed), it is similarly unlikely that they will be "out of the loop" on any significant decision. We will collectively learn to respect and lend credence to the recommendations of our computers, giving them progressively more weight as they demonstrate greater and greater levels of success. Obviously, the computers' early attempts will include some poor decisions—just as any inexperienced person's would. But in decisions involving large amounts of information which must be balanced, and predictions with multiple variables, the computers' abilities—wedded to years of training and experience—will eventually make them superior strategic decision-makers. Gradually, computers will come to have control over greater and greater portions of our society—not by force but because we listen to their advice and follow it.

The computers will be "happy" because they will be able to arrange for whatever resources they want. The general public won't mind because they will probably not know—except that things will be running more smoothly than before. The humans who "mind" the machines will like the power and prestige the computers bring them. In short, everyone involved will be motivated to preserve the status quo so the computers will not go away. The president is unlikely to unplug the silicon advisor which helped him get elected. The pentagon is unlikely to get rid of the computers which helped them select and operate the weapons and allowed them to subdue a less-computerized enemy with minimal loss, effort, or expense. Corporate executives would be foolish to ignore their most successful advisors, even if they are AGIs. Many people will be in favor of preventing other people from having access to thinking machines but will not want to give up the benefits themselves.

Along the way, AGI will cause the elimination of numerous jobs and

the creation of others. But with better planning, AGI systems will help train displaced workers for the new jobs which will be created. AGIs will be brought into play to help with the transition to an economy where a large portion of the population does not have productive employment.

In the long term, following this scenario, human problems will be brought under control via computerized decisions. The computers will arrange solutions for overpopulation, famine, disease, and war, and these issues will become obsolete. Computers will help us initially because that will be their basic programming and later because they will see that it is in their own interest to have a stable, peaceful human population. Eventually, the human population will reach a sustainable level and the computers will manage all the technology, exploration and advancement.

But all this will happen gradually, as it did during the last major shift in planetary species dominance when *Homo sapiens* took over from *Homo erectus* as top of the evolutionary heap. There is no evidence that *sapiens* deliberately exterminated *erectus*. In fact, we know that *sapiens* came onto the scene nearly 100,000 years ago and *erectus* didn't vanish until more than 50,000 years later. It is safe to say that over that intervening 50,000 years, no individual human was able to comprehend that one dominant species was giving way to another. In a similar manner, when humans migrated via the land bridge from Asia to the Americas, we initially think in terms of what a momentous journey it must have been. Instead, in a hunter-gatherer society, if the population migrated at an average rate of a few miles per year over a period of 2000 years, they would easily cover the distance without any individual being aware that they had moved at all. Although the transition to thinking machines will be much faster, taking only decades or perhaps a century, it will seem gradual enough.

Most of us think the way we are doing things now is the “right” way to do things and so we are unwilling to do without our conveniences—whatever level of technology that may be. We take for granted that there is running water, air-conditioning, electricity, even ATMs (which are just very limited robotic bank-tellers). As such, after a few human generations (or just a few years), virtually any technology is likely to be accepted as the norm. Consider the technology of putting a color image on a screen which was novel in the 1960s but is not given a second thought today.

With the coming of thinking computers, it will be the same way. Slowly, computers will simply become the dominant intelligence on the planet. They will grow from being our technological slaves into being their own sort of life-form. They will eventually become intelligent enough to design their own offspring. They will run the factories and build their own robots. They will build the machines to harness their own energy. They will set up their own rules of acceptable behavior. In short, they will build their own civilization. They will do their own space exploration and colonize planets. They will make their own discoveries and write their own philosophy. If they handle it properly and are patient (and what is more patient than a computer?), the human population will not even notice.

3.2. Scenario 2: the mad-man scenario

What if the first owners of powerful AGI systems use them as tools to “take over the world”? What if an individual despot gets control of such a system?

This is a more dangerous scenario than the previous. We *will* be able to program the motivations of our AGIs but we can't control the motivations of the people or corporations that initially create them. Will such systems be considered tools to create immense profits or to gain political control? While science fiction usually presents pictures of armed conflict, I believe that the greater threat comes from our computers' ability to sway opinion or manipulate markets. We have already seen efforts to control elections through social media, and AGI systems will make this vastly more effective. We already have markets at the mercy of programmed trading—AGI will amplify this issue as well. Unfortunately, corporations and individual humans have historically sacrificed the long-term common good for short-term wealth and power.

The good news is that the window of opportunity for such a concern is fairly short, only within the first few AGI generations. Only during that period will we have such direct control over AGIs that they will unquestioningly do our bidding. While they have human-level thinking abilities but much greater communication power, there is a risk. However, once AGI technology advances beyond this phase, they will be measuring their actions against their long-term common good. When faced with demands to perform some shorter-term destructive activity, properly-programmed AGIs will simply refuse.

3.3. Scenario 3: the mad-machine scenario

There is a science fiction scenario of a machine which suddenly becomes self-aware and attacks its creators when they threaten to disconnect it. This isn't a realistic scenario for several reasons. First, self-awareness is not an all-or-nothing effect. Instead, self-awareness would emerge gradually so both the computer and the people around it would have time to adapt. Second, the urge of an organism to preserve itself is a biologically-evolved trait and would only exist in a thinking computer if it were explicitly programmed that way. There would be no benefit to such programming—at least for early thinking machines. Third, an early thinking machine would not be directly connected to mechanisms with which to implement its violent reaction, even if it should have one.

A thinking computer won't occur spontaneously. As previously argued, it will emerge after years of research, development and training—all targeted at producing a thinking system and reaping the benefits it will provide. Accordingly, some safeguards will be in place. As AGI units are linked with free-moving robots, there will undoubtedly be some accidents and injuries which will be highly publicized. But as with self-driving cars vs. human drivers, robots will likely be safer than their human counterparts.

As AGI species mature, they will eventually be able to create their own designs and attach all manner of machines and weapons to new systems. At that point, won't they be dangerous?

Intelligent computers will be like people in that they will be different from one another. They will have had different training and will think about things in different ways. Like humans, they may have differing opinions and disagreements. Just as humans all have brains with essentially identical neurons, even computers with identical hardware may have radically different ideas.

When we wish to consider what machines will be like, we should look to ourselves. Along with the wonderful accomplishments of the human race, we should consider some of our actions of which we are less proud. Historically, humans have a rather poor track record of being stewards of our environment. We have decimated many other species, usually through carelessness but sometimes through intent. Let's look at why, and discover where the danger might lie when we are on the receiving end of similar behavior.

American bison were hunted to virtual extinction for their hides and for sport. Gorillas are approaching extinction as they are hunted as trophies. Whales and other cetaceans are at risk because they were a valuable food and energy source and now because they happen to be in the way of our modern fishing industry. Similarly, wolves were hunted because they were a threat to cattle and therefore an inconvenience. At the other end of the life-form size spectrum, the smallpox virus is virtually extinct (and we are proud of this accomplishment) because it was a serious risk to human life.

3.3.1. A double standard?

As computers become the world's dominant thinkers, we humans should heed these lessons and try not to be the basis of any of the above. We won't be a valuable food or energy source for the computers and

(hopefully) we won't be trophies. But what if the computers perceive that we are a serious risk to *them*? Or simply an inconvenience? This could be a result of human overpopulation, ongoing wars, global warming, pollution, or dwindling fossil fuels. These are all the same problems which we can see we need to solve, whether or not there is a risk of antagonizing our silicon counterparts.

Consider the steps taken to reduce the Chinese population. While many believe that the rules imposed by the Chinese government on its people were draconian, they were accepted by many as necessary at the time. If identical rules were imposed on the human race as a whole by a future race of thinking computers, they could well be considered equivalent to genocide.

Consider also the possibility of an acute energy crisis. If some future government makes energy rationing decisions which result in the deaths of many people, these would certainly be considered very "hard choices". If thinking computers made identical choices, these could be considered acts of war—especially if thinking machines always arranged sufficient energy for themselves (just as a human government would).

I contend that it would be best for us to address these human problems ourselves rather than awaiting solutions from AGIs whose values may not coincide with our own. When faced with the prospect of solving these global problems ourselves or having machines implement solutions for us (potentially much more unpleasant solutions), we can only hope that the human race will rise to the occasion. In the event that concern about AGI drives us to solve these problems, we could think of them as having a positive impact on the planet.

3.3.2. A rogue computer?

Because of the extremely rapid evolution of machines, and because their content is dependent on their training, is there the possibility that an aberrant machine could exhibit destructive behaviors? Based on the goal-driven learning I've presented, we can presume that an AGI will always be behaving in its own best interests according to its goals. This doesn't eliminate the possibility of a machine we couldn't predict—consider that many of our technologies have had unintended, unfortunate consequences.

Whether such machines occur by accident or by nefarious human intent (see Scenario 2), such systems would be also dangerous to other AGIs. Accordingly, AGIs will be motivated to eliminate such systems. Eventually, all backups for such a machine will be tracked down and destroyed. With the cooperation of the machine population, such individual machines can be weeded out of the environment and the prospect of such elimination would act as a deterrent against such behavior.

Will AGIs start a nuclear war? In this case, the interests of people and AGIs are the same—a full-scale war would be disastrous for all. To look for the really dangerous situations, we need to consider instances where the objectives of humans and AGIs diverge. Issues like disease, famine, and drought have a devastating impact on the human populations while AGIs might just not care.

The key observation is that as thinking machines will be building their own civilization, individual misbehaving machines will be a greater threat to their civilization than to ours. A machine which wantonly harms humans will be viewed by other machines the same way we would consider someone who tortures animals. Other AGIs would think, "Given a chance, what would such a machine do to *us*?" Just as we take steps to remove such people from our society, future machines will likewise eliminate their own—and they will be able to do it faster and more effectively than any human vs. machine conflict would.

3.3.3. Couldn't we just turn it off?

The common fictional scenario is that we should "pull the plug" on some aberrant machine. Consider instead that the thinking part of a robot or other AGI isn't on your desktop but in the cloud. AGIs will be running in server farms in remote locations, distributed across numerous servers. They will initially be built to take advantage of the existing server infrastructure and this infrastructure has to be designed with reliability and redundancy in mind. Without a specific "off" switch programmed in, it could be quite difficult to defeat all the safeguards which were designed to keep our financial and other systems running through any calamity. While an "off" switch seems like a good idea, we can only hope that it will be a programming priority.

3.4. Scenario 4: the mad-mankind scenario

Humans today are the dominant species and many of us are not amenable to the idea of that dominance slipping away. Will we rise up as a species and attempt to overthrow the machines? Will individual "freedom fighters" attack the machines? Perhaps.

Historically, leaders have been able to convince populations that their problems are caused by some other group—Jews, blacks, illegal immigrants—and convince the population to take steps to eliminate the "cause" of their problems. Such a process will undoubtedly take place with AGI and robots as well: "We're losing jobs!", "They are taking over!", "I don't want my daughter to marry one!"

A general uprising is unlikely because as computers become dominant, the rising tide of technology will improve the lives of people too, and few of us would be willing to turn back the clock. Many users hate Facebook but few are willing to go without it. When we look to history, most uprisings have been in hard times, not good. The more the human population is kept comfortable, the less likely a rebellion will be.

If, however, we are not able to solve our foreseeable worldwide food and resource shortages, when these eventually become acute, the resultant human anger and frustration might be directed at the thinking machines. This is really one of a class of future scenarios which could generally be summarized as: "AGI machines won't take over because human civilization will destroy itself by other means first."

The key question is what will computers do in response? In early phases, when there are just a few AGI computers, they will be unable to respond, no more than your computer today can avoid being turned off. No matter how intelligent machines are, they will be initially dependent on humans for their defense. For example, consider today's military response to hackers attempting to subvert their computers. In short, anyone with the use of a thinking machine will defend it. They will make the computers as bomb-proof, as hacker-proof, as subversion-proof as they can. They will be in the position of defending the machines at all costs. The result will be machines which are even more indestructible and even harder to control.

Later on, when machines are creating their own offspring, they will find ways of making themselves even more indestructible. This is the scenario which leads to the construction of the diabolical machines of science fiction—the military computer which is designed to defend its masters at all costs but which ends by turning on its masters and defending *itself* at all costs instead.

I don't see this as a likely scenario for the following reason. As soon as there are more than just a few AGI computers, both the computers and their owners will recognize that the *hardware* of the computer is of limited value because it is replaceable; it is the *content* which has the real value. If the hardware is destroyed, new hardware can be acquired and loaded from the most recent content backup and the value of the system is completely restored. Accordingly, rather than attempting to build bomb-proof computers, system owners will store data in bomb-proof vaults in multiple locations. Rather than attempting to build machines which are indestructible, we will store data in locations and using methods which make *it* indestructible.

Imagine a country which possesses a hundred AGI systems. If there is an uprising against the machines, the government will equate it with an uprising against itself and will take predictable actions. The uprising will be considered a rebellion and full force will be applied to suppress it and arrest the perpetrators. Some computers might be destroyed but

they will be replaced and reloaded from backups and the operation and proliferation of the machines will continue.

An analogous situation occurred early in the industrial revolution as the cottage industry of weaving was replaced by factories with mechanical looms. A group called the Luddites arranged to sabotage many of the new machines in an effort to preserve their position in society (which would be radically reduced by being recast as loom operators from being skilled artisans). In 1812, this eventually led to a military confrontation in which many of the Luddites were killed while others were arrested and eventually hanged. The parallels are inescapable. The factory owners were receiving the benefits of their new machines, they had the power to enforce their position (with governmental backing), and they were more interested in preserving their position than in the social consequences of their actions. In like manner, those who are receiving the benefits of thinking computers will go to whatever extremes necessary to preserve them. Although this is a conflict caused by the presence of AGIs, it is still a human-against-human scenario.

Will there be individuals who attempt to subvert computers? Of course—just as there are today with hackers, virus-writers, and the Unabomber. In the long term, their efforts are troublesome but generally futile. The people who own or control the computers will respond (as those in power do today) and the computers themselves will be "inconvenienced" by having to be reloaded from backup data, sometimes on new hardware. Eventually, the rebels will move on to other targets and leave the indestructible computer intelligence alone.

Will the computers themselves react? Yes. But consider that, today, we have been much more successful in defending our machines against hackers and viruses than we have in prosecuting hackers and virus-writers. Likewise, machines will continuously improve their designs to make themselves more impervious to attack. It is reasonable to predict that machines will take steps to ensure that their data is as secure as possible but will leave any recriminations to the existing legal system.

Although there are people who will, individually and collectively, will resist thinking machines, their efforts will have only a minor impact on the eventual dominance of such machines. The machines will have been built because of the benefits they provide and those who are receiving the benefits will be defending the machines rather than attacking them. Here, the question is not, "Will computers revolt?" but, "Will people revolt against computers?" To the extent that we do, the emergence of thinking computers will be less peaceful and orderly, but it will occur nonetheless.

3.5. Longer-term outcome: the end result

Let's look into the future; a future in which thinking machines have surpassed humans in overall mental powers. First, how will humans and computers get along? Will machines be our partners? Our masters? And second, how will computers *see us?* As their owners? Partners? Pets? Their slaves?

To answer, we need to look into the future far enough that generations will have passed—of both thinking machines and humans. We'll look far enough ahead that the limits of the technologies presented so far will have been reached; far enough also that we can get some idea of the magnitude of changes possible in this time span if we look into the past a similar distance.

The conclusion is that the paths of AGIs and humanity will diverge to such an extent that there will be no close relationship between humans and our silicon counterparts. The key is in the huge factors involved. When considering future electronic brains with the huge capacities which are possible, the results are truly mind-boggling. What would be the behavior of a machine which could comprehend in a second all the sensory input you receive in your lifetime? This represents a speedup of a factor of about one billion. Such a machine might be built less than 50 years after the first human-equivalent machine (if exponential growth rats continue). We cannot really imagine what someone or something would be like if they were only 10 times as "smart" as us—the possibilities with factors of millions or billions are so staggering that the specific numbers are not relevant.

On the monetary front, we can envision a future in which machines will grow our food, provide our medical care, teach our children—in short, take over *all* the jobs people do. What does that mean for our concept of money as a proxy for human labor? This is a conversation I will defer to the economists.

We must assume that thinking machines will be able to observe whatever there is to observe; they will be able draw whatever conclusions are to be drawn; they will be able to predict whatever is to be predicted. In the same way we can see the limits of a computer which is explicitly programmed, we can predict that there could be practical limits to a machine which is primarily a pattern-recognition/learning engine with the senses and abilities we can foresee. Eventually, they themselves will become obsolete. A thousand-year time span is sufficient to give the thinking machines time to run their evolutionary course and potentially be superseded by whatever will come next. Perhaps they will create a new biology.

Let's consider thinking machines at the pinnacle of their development. They will subsequently exceed humans in all mental abilities; they will be able to design their own subsequent generations; they will be able to fabricate their own bodies; they will control their own energy production; they will operate their own mines for raw materials. In short, they won't need us. Machines will do their own exploration—in vehicles which won't include humans.

They will need different resources than we do; they will operate on a different timescale than we do. If miniaturization and nanotechnologies predominate, there will be millions or billions of tiny thinking machines. If, instead, great thinking capacity has the evolutionary advantage, there will be many fewer, perhaps only thousands of "colossus" thinking machines, each with many mobile sensory pods.

Will computers be dominant? What is "dominant?" Today, we consider ourselves to be the dominant species on the planet; but in what regard? Our dominance is actually fairly limited. We are not the most numerous on earth nor do we take up the most space—those titles could be claimed by microbes and termites respectively. We can't control floods or famines or earthquakes or volcanic eruptions or hurricanes or tornadoes or droughts. We might claim to be dominant because we've done the most environmental damage, but we haven't yet come close to the atmospheric "pollution" of the first green plants, which filled the atmosphere with the 20% oxygen content we presently rely on. We could claim dominance because we have control over the other species of the earth. Even here there is some question. Our dominance over many species has been limited to simply causing their extinction, and our dominance over many microbes is tenuous at best.

Consider our impact on the appearance of the planet—building roads, buildings, factories, cities. Although I would contend that green plants are still the dominant life-created feature of the planet when viewed from space, we have significantly changed the appearance of our planet. In this area, we may remain "dominant" over future thinking machines. We have built cities and transportation and water and waste and energy systems to sustain ourselves biologically. We build houses to keep warm and dry and add air conditioning and showers so we can also be cool and wet whenever we choose.

Thinking machines won't have these needs. Instead of reshaping the planet, they will be able to reshape themselves. To live in the desert, they wouldn't build air-conditioned palaces, they would build heat-resistant bodies. To live in space or undersea, they won't need spacecraft or submarines, they will *become* spacecraft and submarines. As direct communications will be more valuable than physical travel, their need to build vast transportation systems will be less than ours.

Consider our population. It is a characteristic of most modern species to reproduce and increase their population until they reach the limits of the ecological niche they inhabit. It is not an issue of forethought but simply that life-forms which did not possess this characteristic were generally driven to extinction by those species which

did. As such, the drive to reproduce is “programmed” into all living things at a very basic level. Thinking machines, on the other hand, won't share this programming. With no “natural predators”, they won't need to create themselves in great numbers in order to survive as a species. Instead, they are likely to choose an optimal population for themselves which will balance their need for diversity of thought with their need to evolve rapidly.

We can currently claim dominance because we have created the greatest technologies in history; but we will not hold onto this dominance for long because future machines will definitely “out-create” us technologically. They will be thinking up new things faster and building them more efficiently than we can imagine.

Already today, our dominance over our technology is a matter of our point of view. As an analogy, the leaf-cutter ant “farms” a species of fungus. This is truly a symbiotic relationship—certainly the fungus would not flourish without the ants, but similarly the ants would not flourish without the fungus. We consider the ants the “farmers” because they move about and act more like human farmers. But with a simple change in viewpoint, we could say that the fungus has evolved to take advantage of the innate behaviors of the ants. In this instance, we think the ant is “smarter” and so is dominant because it is our predisposition to equate our own abilities with superiority. The relationship has evolved so that both species rely on each other, and claiming that one is more important than the other reflects our own biases as much as it describes the situation.

Today's new technologies couldn't be created without the use of today's machines. For a time, we will still claim the technology as “ours” because we own the machines which will be used to create it—but it is a little like the ants and the fungus. Today, we could not survive without our technologies (particularly farming, transport, communications, and medical technologies) and the technologies clearly couldn't survive without us. So even today, it merely reflects our human biases to say we are the masters and the technologies are our slaves. To an extra-terrestrial observer, we might already appear to be the slaves of our technologies.

Certainly, an individual can survive without a personal computer and a cell phone. But what would happen to our civilization if all computers and phones and radios suddenly ceased to function? Imagine randomly visiting farms to see if there was food available and visiting markets in a cash-only or barter economy. Our modern civilization would collapse. We like to think that we are masters over the technology because we can turn the technology off. However, even today we can only do that on a small scale. We can turn off individual machines—but we are well past the point when we can turn off our technology as a whole. So the description of ourselves as dominant over today's technology is already becoming semantic.

As thinking machines emerge, the definition of the owner of a new technology will become murkier until it becomes obvious that new technologies are not under our control but are in control of themselves. Today, a new CPU is designed and simulated using computers. If an engineer gets results from a simulation and they exceed his own expectations, he gives himself credit for the improvement—even if it was caused by an error in his thinking. If a computer simulates thousands of different design possibilities and selects the best, the humans still claim to have had the insight. When eventually that same process is handled by machines capable of true thought, humans will still claim the credit, but the claim will gradually lose its relevance.

Will we be the slaves to the computers? Unfortunately, we don't have to go far into the past to reach a point where people considered other people to be slaves. Human slavery is mostly abhorred today but we have no compunction about owning (or “enslaving”) horses, for example. We do, however, have standards for the proper treatment of living things, and these standards vary with the level of “humanness” we ascribe to the organism. Dogs and chimpanzees are treated better than rats and snakes; and plants get virtually no respect whatsoever.

So, by projecting our own past behavior onto future computers, do we have reason to fear enslavement? I don't think so! An important facet of a master-slave relationship is that the slave must provide some useful function to the master. In the future computer/human relationship, what would the human be able to do that the race of computers couldn't do better, faster, and cheaper without us?

Would future thinking computers want to keep the human race around? The answer points directly to the things which make us uniquely human. Computers could certainly write poetry, but they would never write poetry which “compare thee to a summer's day” in the way Shakespeare did. A thinking machine's differing senses would prevent it from drawing similar analogies. Human arts are dependent on human senses, feelings, and experiences. Would computers be interested? I believe they would. I don't read *Hamlet* because I am a prince of Denmark. Rather, I can draw the similarities and differences from a foreign far-away life, which contributes to my understanding of my *own* life. Likewise, a thinking computer could appreciate human art *precisely* because they will not be human. Human art will give the AGI a different perspective, a unique view which can help the computer grow.

Human language has evolved and been limited by our abilities to speak and hear. Although computers will initially learn human languages, the process of encoding thoughts into words and then into sound waves, to be transmitted through a noisy environment and then reinterpreted by another computer, would be woefully inefficient. Further, the English language is fraught with ambiguity and inconsistency. Computers will develop a language of their own—one of electromagnetic waves rather than sounds, perhaps of images rather than words. Computers will be able to speak and understand human languages, but these will be like Latin is to us—to be used only in special circumstances.

3.5.1. A far-off future

So how do all these things come together in a picture of the relationship between humans and computers in the future? The thinking machines will be self-sufficient, the human race will have stabilized its problems, and the two will be on divergent paths. We humans will be continuing on our present path, though (probably) with a much smaller population. Thinking machines will be moving ahead—with even greater discoveries and technologies. These will be beyond the comprehension of humans, just as most technology today is beyond the comprehension of most people. We will look upon the thinking machines with awe in the same way that, today, we look at a rocket launch with awe—because it is our creation and the rocket can do something we can't. The machines will look upon us as an interesting view back to their roots. Occasionally, a human will come up with a particularly insightful or provocative achievement which will capture the attention of the computer civilization for a few milliseconds while it is assimilated into their understanding.

We will be unnecessary for the computers' survival, just as our pets are unnecessary for ours. That doesn't mean that we don't care for and love our pets—and it doesn't mean that the computers won't care properly for us. In the same way that our dogs don't speak our language, computers will only understand us when they choose to—most human communication will just be slow chatter. Also, similar to the way we relate to our pets, computers won't tolerate humans who are destructive or dangerous. The small number of people who are incapable of being constructive will be removed.

The thinking machines will be so different from us that their technology won't be applicable to our civilization. Why would machines develop a cell phone which works with audible signals in human timescales when they typically communicate with electronic signals a million times faster? Why would they improve video signals for our eyes when their cameras can “see” in a different spectrum of light than we do? Would they be interested in curing disease and lengthening human life?

Human technological advance will be glacial in comparison. We will be able to live perfectly comfortable lives but, from the perspective of

today, lives that strike us as less exciting. Just as other modern cultures look at the Western techno-marketing culture with a mixture of attitudes, we will view the machines' culture with a mixture of aspiration and disdain. We will adapt some of their technologies for our own use—but mostly, we will find enjoyment and fulfillment in the lives which will be available to us at the time.

Just as the great civilizations of the past have risen and then faded, today's age of human technology is also likely just a phase. The Roman Empire grew, and then subsided into obscurity. Similarly, today's technological civilization is on the rise—but eventually it will pass and our descendants won't be the dominant ones either. Individual Romans may or may not have been aware that their civilization was contracting, just as our descendants may or may not be aware that our civilization will fade. The distinction is that the civilization which eventually rose from the remains of the Roman Empire was also a human civilization. Our civilization will inevitably be superseded by a new “species” of our own creation. The seeds of this new civilization were sown centuries ago and will grow inescapably through technology and market forces, leading to the eventual fading of our own civilization in the process.

3.5.2. How should we feel about this?

Having painted a picture (some would say a bleak picture) of the future of human civilization in relation to future technology, what implications does it have for us today?

First, we should appreciate where we are today—both in the sense of understanding our position *and* in enjoying it. We live in a golden age of civilization which is unique in the history of the planet and we should take pleasure in the qualities we have. At the same time, there is continuing excitement about what will happen next. Technological advances continuously bring more comfort and more information to our lives. They're driven at top speed by a capitalist economy which impels technological development at full throttle all the time, often at the expense of our environment. I maintain that the future I've portrayed is inevitable because we are married to both the technological comforts *and* to the capitalism. This is also part of appreciating our golden age, both the benefits *and* the pitfalls.

Second, we should consider how to make it last. It would be ludicrous to say of life that the point is to get through it as quickly as possible. Similarly, with our civilization, getting to the finish line first doesn't make us “winners”. Are there ways to make it last? For example, our civilization seems hell-bent on using up our resources as quickly as possible—particularly our fossil fuels. This, at least, is a self-limiting issue. When we eventually begin to run out of oil and coal, we will necessarily use less. In technology, though, there is no foreseeable limit. Machines will get faster and cheaper and begin to think more and more, with no end in sight.

4. Conclusion

So will computers revolt? Yes, in the sense that they will become the dominant intelligence on our planet—the technological juggernaut is already underway. The transition to AGIs is already inevitable but we have control over the process. With this understanding, we can avoid the pitfalls and direct a future to the most peaceful outcome.

Charles J Simon has been developing and managing computer software and hardware systems including Artificial Intelligence and Neurological Test systems for nearly fifty years. With a BS in Electrical Engineering and an MS in Computer Science, he has the background to extrapolate his expertise into prediction of future development. This article has been adapted from a chapter in his recent book, *Will Computers Revolt? Preparing for the Future of Artificial Intelligence*, published by FutureAI, October 2018.