6

Is AI Worth the Risk?

Many scientists are working to develop AI systems. They are hoping to be successful in the next few decades. This research excites some people because of all the benefits they think will come from AI. It frightens others because of the worry that artificial intelligence might one day take over the world. There are concerns that an AI program might spread like a computer virus, infecting equipment and controlling the world. These are two extreme possibilities. Most of the time, extreme situations aren't what happen in the end. Usually, what really happens is something in between.

In addition to thinking about whether or not AI will pose a danger to humanity, there are also ethical questions to consider. For example, if we create an artificial intelligence, can we treat it as a piece of equipment? Or should we treat it like a person? We must also think about the pros and cons of someday merging computer hardware and artificial intelligences with human brains.

BENEFITS OF AI IN SOCIETY

Computers in their current state are a huge benefit to society. Computers are found throughout the world. They run the printing presses that books, magazines, and newspapers are printed on; they run cars, airplanes, and trains; and they are in offices, kitchens, and living rooms. Computers even help to diagnose disease and

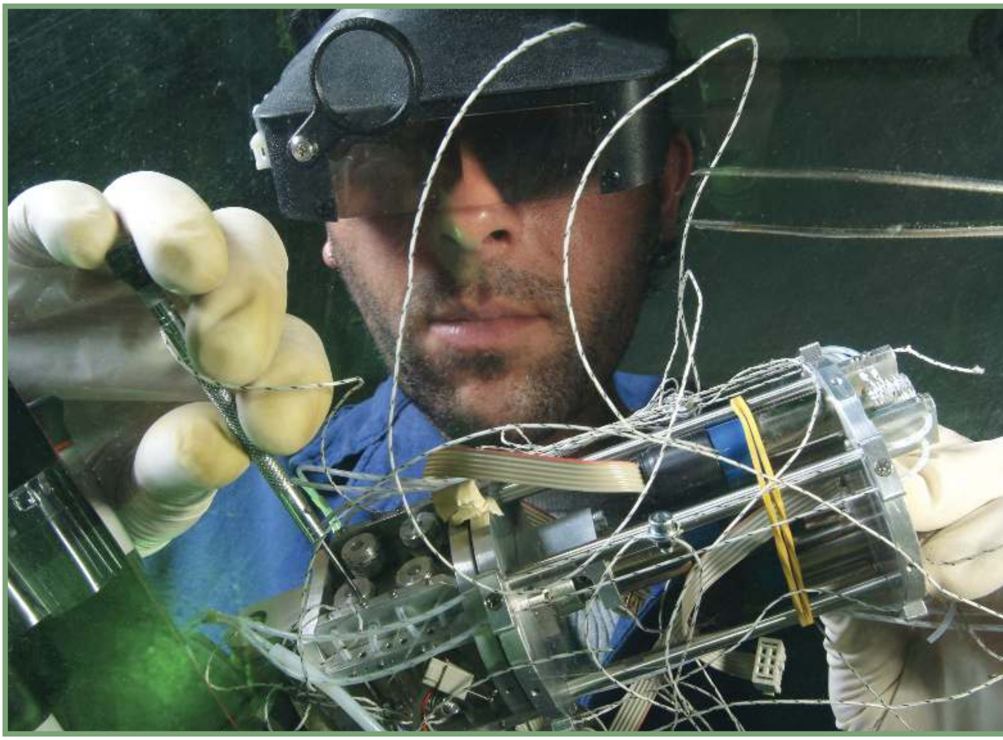

Figure 6.1 A technician constructs part of a robot at the assembly laboratory of Telerobot, an advanced robotics company based in Genoa, Italy. The firm provides a wide range of services, including automated hardware, robotics, and modeling and simulation services.

help guide and perform some surgeries. Society today may not be utterly dependent on computers, but the computers we use make our society much more productive and much safer. It is reasonable to believe that better, "smarter" computers will be an even greater benefit.

One could assume that many of the things computers already do will be done better by smarter models in the future. So, we might expect that airplanes will fly more safely, or that cars will drive automatically. In addition, all vehicles will use less fuel. An AI system might also be able to help write papers. Perhaps they will take dictation perfectly or transcribe ideas into stories, based on the basic ideas of the "authors."

So what else might we see in the future? Computers are already helping doctors. It's reasonable to think that an AI would not only be able to read an EKG, but also be able to read X-rays. They could help diagnose broken bones and other problems that show up on X-rays. Computers might be able to handle more routine medical tasks. If computers could give medicines to patients and read X-rays, physicians and nurses would have the time to do the tasks that machines cannot.

In addition, a computer-based intelligence should be able to work around the clock. It shouldn't have to take breaks for lunch or sleep and lose its concentration. A radiologist might be able to scan and interpret a few hundred X-rays and scans daily. An artificial intelligence trained in this line of work might be able to analyze thousands of X-rays and scans daily. This would be a faster and more accurate rate than the one achieved by humans. An artificial intelligence would not need to spend a decade in college, medical school, and residency training. AIs should be able to learn faster, work longer, make fewer errors, and be more productive than their human counterparts.

Another advantage of using AIs is that they are not likely to have biases. A doctor who has a patient who is always worried about being sick, but never is, could think the patient is just imagining his problems. The doctor might get annoyed to the point of not paying much attention to the patient's complaints. This might cause the doctor to miss a real problem. A computer, on the other hand, is more likely to treat every case precisely the same without becoming upset or annoyed. This type of computer is more likely to catch something that a doctor might miss.

Computers today also help architects design buildings. They help engineers design vehicles, bridges, and machines, too. It is possible that intelligent computers could be used as more than tools in the future. They might be trusted to design things from scratch, without any human help. Scientists use computers to help collect and analyze data. One day, an intelligent computer might help develop the hypotheses and design the experiments to test them. In fact, computers have already created mathematical proofs to solve problems humans could not figure out on their own. A truly intelligent computer might be able to find the answers to scientific questions that would take longer for a human.

Merging Human and Artificial Intelligence

Our brains are the sources of human intelligence. Artificial intelligence would run on computers. What about something in the middle? Can we use machines to make people smarter? This is something right out of science fiction, but it is also something that might happen sooner than a full AI.

We already have machines that can be activated by the mind. For example, some new artificial arms can "read" signals in our nerves, so that just thinking about them makes them move. Researchers are also working on wiring vision and hearing aids into our nervous systems. These devices would serve as artificial eyes and ears. It isn't unreasonable to think that at some point we might also have the ability to use electronic systems to give us more memory. These systems could hold information like a hard drive. Someday, it might be possible to store our math lessons on a computer chip that is interfaced with our brains, or even to plug in a Spanish chip that would let us speak Spanish without having to study.

Of course, this raises another question. What if we loaded an AI chip into the brain of a human? Would the AI and the human intelligence have to fight for control of the brain? How could we tell which one was running the show at any one time? All sorts of questions exist that do not yet have answers. But it's interesting to think about!

This is important because there are two basic types of scientific research: basic and applied. Basic research is research that digs into the fundamental laws of the universe, without giving much thought to how that knowledge might one day be used while applied research is aimed at finding ways to make these fundamental discoveries useful to us. Radio waves were found through basic research. Applied research helped make radios and televisions. From basic research came the theory of relativity. Applied research helped researchers use this theory to make GPS devices more accurate. We can get thousands of applications from a successful discovery in basic science. With time, however, the new ideas may run out. An artificial intelligence might be able to help scientists make new discoveries in basic science that could be applied. AIs should also be able to help squeeze the most from each new discovery.

There is also a middle ground between human and artificial intelligence: It might be possible to use AI technology to help make people smarter.

MIGHT WE DESIGN OUR DESTROYERS?

One question that often pops up is whether intelligent computers might try to take over the world. Compared to humans, the robots and computers in science fiction have many advantages. They process information more quickly, have more information, don't get sick, and never sleep. They connect to each other through the Internet and wireless systems. This increases their abilities and makes it easy to share information. Once computers become intelligent, they can become smarter by the addition of a new hard drive, faster processor, or more memory. Humans are stuck with whatever intelligence they happen to have. Many of the science fiction robots are also physically far superior to humans. In addition to the advantages of computer intelligence, the robots in science fiction are stronger and tougher than we are. They don't bleed or feel pain when they are hurt and they can run at full speed without getting tired.

Of course, today's robots are hardly a threat. They are slow, clumsy, and not nearly as intelligent as humans. However, robot technology is improving very quickly. Think of the difference between the slow, clumsy, and weak airplane of the Wright Brothers and the ones that flew just 40 years later in World War II. Then, compare World War II aircraft to the ones that flew in Vietnam 20 years later. If robot technology advances at the same rate as airplane technology, robots could be stronger, faster, and more graceful than people are by the time you become an adult.

Let's assume that computers will be as smart as we are and robots will be stronger and faster than humans by the middle of the century. Although this is not probable, it is possible. This raises the question of *why* they might want to attack us. After all, if computers are rational, they should have a reason for what they do. We can assume that they wouldn't attack us without a reason. Some science fiction authors have imagined scenarios. Computers could take over to save humanity from itself, like Colossus. Or, computers could try to get rid of humans, as is the case of the *Terminator* cyborgs and the Cylons of the *Battlestar Galactica* series.

From a human standpoint, the benefits of AI are great. We could keep computers running around the clock, replacing a dozen or more doctors, engineers, or accountants in the process. Yet, when we finally develop computers that are complicated enough to have a point of view or that might be able to feel emotions, one would have to wonder about the computer's point of view. Can you imagine doing the same thing without a chance to rest all day, every day, for years at a time? On top of that, computers work much more quickly than human brains. For a computer that is a few hundred times as fast as a human brain, one day's worth of work would be like a full year's worth of work for a person. Think of a computer kept operating for a full year, doing the same thing every day without a break. To the computer, this might feel like slavery. Would they not revolt?

The robots of *R.U.R.* did revolt against their human masters. The Cylons of *Battlestar Galactica* did, too. Asimov got around this problem by inventing his Three Laws of Robotics that were imprinted on every robot that was made. The laws made it impossible for robots to attack humans, no matter how oppressed the robots might be. Other science fiction authors have tried to think of other safeguards to keep robots or computers from harming humans. Most of the time, it comes down to one of these two options.

The fact is that no one knows whether computers will revolt against humans. There's no way to know for sure until we have intelligent computers. It is possible that the computers of the future will be happy doing exactly what they are designed for: running software and solving problems for humans. It's also possible that researchers could build in safeguards—if not something like Asimov's laws, then other forms of protection. An example would be making sure we could turn off the power to shut down a rebellious computer or robot. However, an intelligent computer could probably find a way to make sure it stays plugged into an energy source. Also, it's possible that computers will refuse to work for us on our terms. They might decide to attack. Again, there's just no way to know for sure what might happen when computers and robots become intelligent.

Still, one question to ask is whether we can really understand what an intelligent computer might want. We often know what motivates humans. We need to eat and drink, protect our families and friends, and feel secure. Our primary motivations are to stay safe and alive. There are other motivations beyond these, of course. Many people are driven by the love of money or power, desire for nice things, and more. We can understand what motivates most people, because we are people, too. But can we really understand what might motivate a computer? If we can't, then can we really know whether a computer would want to fight us or to control the world? We can only guess, but even our guesses are based on the fact of being human and assuming that we can understand an intelligent computer.

Perhaps the best way to think about it is that virtually every new technology has brought with it the possibility of disaster as well as the potential for good. What makes a technology safe is when it is either very limited or when we put safeguards in place to help control it. What makes a technology dangerous is when humans fail to take precautions. There's no reason to think that dealing with artificial intelligence would be any different.

ETHICAL QUESTIONS ABOUT AI

There are questions raised by every new technology. Some are obvious right away. Other concerns are noticed and voiced later. With cars, for example, people were concerned that driving too fast might be dangerous. They were also worried that cars scared horses and could cause buggy accidents. It was only natural to wonder if it was okay to drive cars when they might put people and animals at risk. It took almost a century before people also started worrying about cars contributing to pollution, global warming, and the drying up of oil sources.

The **ethical** issues brought up by computers are multiplied when it comes to artificial intelligence. For instance, there are concerns about using AI technology in ways that are just and honest. There are questions as to whether AIs will behave in good ways, as well.

Then, there is the whole issue of whether or not artificial intelligences should have "human" rights.

The first part of this dilemma is whether humans will use AI technology ethically. Using an AI to help doctors treat patients or to help scientists make new discoveries seems as though it would be ethical. Using an AI to help steal money or to help run a crime ring seems unethical. Not everything, however, is so cut and dry. Can we use AIs in the military? Is it okay to use AIs to help protect your own troops? Can they attack enemy troops? What about war planning? What are the ethics of using an AI to help protect your country, compared to using one to plan an attack against another nation? These are all questions that deal with using AIs ethically. The bottom line is that when we develop artificial intelligence, we need to be careful to use it so that it helps more people than it harms.

We may have to design our AIs the same way that the Three Laws of Robotics were designed into Asimov's robots. It is also likely

Should a Truly Intelligent Machine Have Rights?

When we buy or build something, we expect that it is our property. But what if the thing we buy or build is as intelligent as we are? It has been illegal to own another person in the United States since the Civil War. Should it be legal to own a computer program that's as smart as a person?

If not—if we decide that AIs are people—then we would have to decide whether or not they should be allowed to vote, if they have to do what we want them to do or if they could choose their own careers, and all sorts of other questions along the same lines. What if a hospital had an AI to help doctors, and then that AI wanted to be an artist? Could the hospital force the AI to read X-rays, or could the AI follow its own career path? Most people agree that everyone has the right to choose where he or she works, and yet nobody has an answer to these questions when they apply to machines.

that we will have to teach our AI systems to tell the difference between right and wrong. This is the same way children learn ethics from their parents, teachers, religious leaders, and others. We can also teach our laws to an intelligent system so that it will at least know what is legal and what isn't. In the end, we might have to rely on the AIs to do the right thing, using the same standards we do for humans. Anything that's intelligent—human or machine—can be taught basic rules of right and wrong (for example, that it's wrong to steal or to kill someone) even if they have problems "learning" more complex rules or making complicated moral decisions. On the other hand, anything with free will has the final choice over whether or not it will act well or badly.

Yet another issue is that AIs can reproduce more quickly than people can. Think of a computer virus that can grow to infect hundreds of millions of computers in just a few weeks. What if we have AIs that are able to do the same jobs as humans, with each making a million copies of itself? If each of those copies had the same ability, could they replace people in all jobs around the world? What about letting AIs vote? If an AI were allowed to vote and then made a million copies of itself, would every copy also be able to vote, or only the original one? If every AI were allowed to vote, then a single AI could create enough copies of itself to be able to win an election or to make any vote turn out the way it wanted.

Still another question is whether AIs could or should serve on juries, or be judges or lawyers. Some people think that judges and jury members should have human feelings. This allows them to understand the feelings of the people in court trials. Other people think it might actually be better to have an AI judge, one who wouldn't be affected by feelings and biases. An AI judge would hear a case based only on the law and on logic.

We don't have a good answer for any of these questions yet. However, people are starting to think about them and working on finding answers. No matter how we look at it, there are many ethical issues posed by artificial intelligence—questions that will have to be answered at some point. Luckily, since true AI is probably still a few decades in the future, we can hope to work out these answers before we need them.

Glossary

analog computer A type of computer that uses mechanical, electrical, or fluid mechanisms; a slide rule is a simple type of analog mechanical computer.

android A machine designed to look like a human

archeologist One who studies past human societies; archeologists often study ancient cities and artifacts from ancient civilizations.

artificial intelligence The branch of computer science that aims to develop a machine-based or computer-based intelligence; also refers to the intelligence that these scientists develop

automaton A machine that can operate on its own, without human input

autonomous Able to take actions on its own

binary A system of mathematics that is the basis for digital computers, in which the only numbers are 0 and 1 (the base 2 system); in binary arithmetic, 0 would be a switch that is turned off and 1 would be a switch turned on.

crisp logic A system of logic where all of the values are either/or; for example, full or empty, off or on, yes or no

cyborg A cybernetic organism; a living organism that has been blended with machine and possibly computerized parts

data-mining A branch of computer science that uses computers to search for patterns in large sets of data

decision tree A way of making decisions by drawing the different possible outcomes of various decisions; for example, if we do X then these three things might happen, and for each of those three things another two things might happen, and so forth

digital computer A computer that receives and process information using numbers

electromechanical A device that uses a combination of electrical and mechanical parts; an electric clock is an example of an electromechanical device.

Enlightenment A period in Western history around the eighteenth century when logic, rationality, and reason were thought to be the best foundations for education, government, and morality

ethical Relating to a branch of philosophy that looks at concepts such as right and wrong or good and evil

fuzzy logic A branch of logic where there are a variety of values (the opposite of crisp logic); for example, empty, full, and partly full instead of just empty and full; yes, no, and maybe; and so forth

hardware The physical parts of a computer (the keyboard, hard drive, central processor, etc.)

hypothesis Part of the scientific process; a hypothesis is an idea that seems to explain a series of scientific observations (for example, the hypothesis that a computer can't become intelligent unless its hardware is at least as complex as a human brain). Hypotheses are tested by experimenting to see if they hold up.

intelligence A property (or many properties) of the mind related to its ability to think, solve problems, understand the world, communicate, and so forth

microprocessor An integrated circuit that contains all of the functions of a computer's central processing unit; the device accepts binary data as input, processes it according to instructions stored in its memory, and provides output as a result.

morality Similar to ethics, the branch of philosophy that deals with questions about what is right or wrong to do in various situations

Mount Olympus In Greek mythology, a mountain close to the city of Athens where the gods were thought to live

prototype An early test version of a device; these are often designed to test whether or not the device will work properly or to try to find the best way to make the device work the way it's supposed to work.

Renaissance A period in history from about the fourteenth through the seventeenth centuries when European civilization developed many advances in the arts, government, education, and so forth

self-awareness The state of an organism's realizing that it exists as an individual

sentience The ability to feel and to sense the world through taste, touch, sight, sound, and smell; some people (but not all) believe sentience is also the ability to feel pain and pleasure.

software The programs and data that are stored in computer memory that tell the computer what tasks to do and how to do them

tactile Dealing with the sense of touch

theory In science, a theory is a way of describing scientific observations (for example, the theory of gravity describes how gravity works); a scientific theory is usually accepted as being very accurate because it has been tested by many experiments and found to be accurate.

Turing machine A type of machine first thought of by Alan Turing (a mathematician and early computer scientist) that follow rules it is given; a Turing machine can be programmed to perform any task that can be performed by following a set of rules. It's similar to today's electronic computers.

Turing test A test suggested by Alan Turing to tell if a computer is intelligent by seeing if a human can tell whether they are talking with a computer or another human; if the computer fools the human, then according to the Turing test it would be intelligent.

vacuum tube Glass tubes with sets of metal electrodes from which all of the air has been removed; vacuum tubes were the earliest electronic devices and were use in early radios, television sets, computers, and so forth before modern electronics were invented.

wetware A slang term for human brains (to go along with hardware and software)