1. Mark your confusion.
2. Show evidence of a close reading.
3. Write a 1+ page reflection.

Rise of the Machines

Computers could achieve superhuman levels of intelligence in this century. Could they pose a threat to humanity?

Source: *The Week* 10.18.14

How smart are today's computers?

They can tackle increasingly complex tasks with an almost human-like intelligence. Microsoft has developed an Xbox game console that can assess a player's mood by analyzing his or her facial expressions, and in 2011, IBM's Watson supercomputer won *Jeopardy* — a quiz show that often requires contestants to interpret humorous plays on words. These developments have brought us closer to the holy grail of computer science: artificial intelligence, or a machine that's capable of thinking for itself, rather than just respond to commands. But what happens if computers achieve "superintelligence" — massively outperforming humans not just in science and math but in artistic creativity and even social skills? Nick Bostrom, director of the Future of Humanity Institute at the University of Oxford, believes we could be sleepwalking into a future in which computers are no longer obedient tools but a dominant species with no interest in the survival of the human race. "Once unsafe superintelligence is developed," Bostrom warned, "we can't put it back in the bottle."

When will AI become a reality?

There's a 50 percent chance that we'll create a computer with human-level intelligence by 2050 and a 90 percent chance we will do so by 2075, according to a survey of AI experts carried out by Bostrom. The key to AI could be the human brain: If a machine can emulate the brain's neural networks, it might be capable of its own sentient thought. With that in mind, tech giants like Google are trying to develop their own "brains" — stacks of coordinated servers running highly advanced software. Meanwhile, Facebook co-founder Mark Zuckerberg has invested heavily in Vicarious, a San Francisco–based company that aims to replicate the neocortex, the part of the brain that governs vision and language and does math. Translate the neocortex into computer code, and "you have a computer that thinks like a person," said Vicarious co-founder Scott Phoenix. "Except it doesn't have to eat or sleep."

Why is that a threat?

No one knows what will happen when computers become smarter than their creators. Computer power has doubled every 18 months since 1956, and some AI experts believe that in the next century, computers will become smart enough to understand their own designs and improve upon them exponentially. The resulting intelligence gap between machines and people, Bostrom said, would be akin to the one between humans and insects. Computer superintelligence could be a boon for the human race, curing diseases like cancer and AIDS, solving problems that overwhelm humans, and performing work that would create new wealth and provide more leisure time. But superintelligence could also be a curse.

What could go wrong?

Computers are designed to solve problems as efficiently as possible. The difficulty occurs when imperfect humans are factored into their equations. "Suppose we have an AI whose only goal is

to make as many paper clips as possible," Bostrom said. That thinking machine might rationally decide that wiping out humanity will help it achieve that goal — because humans are the only ones who could switch the machine off, thereby jeopardizing its paper-clip-making mission. In a hyperconnected world, superintelligent computers would have many ways to kill humans. They could knock out the internet-connected electricity grid, poison the water supply, cause havoc at nuclear power plants, or seize command of the military's remote-controlled drone aircraft or nuclear missiles. Inventor Elon Musk recently warned that "we need to be super careful with AI," calling it "potentially more dangerous than nukes."

Is that bleak future inevitable?
Many computer scientists do not think so, and question whether AI is truly achievable. We're a long way from understanding the processes of our own incredibly complex brains — including the nature of consciousness itself — let alone applying that knowledge to produce a sentient, self-aware machine. And though today's most powerful computers can use sophisticated algorithms to win chess games and quiz shows, we're still far short of creating machines with a full set of human skills — ones that could "write poetry and have a conception of right and wrong," said Ramez Naam, a lecturer at the Silicon Valley–based Singularity University. That being said, technology is advancing at lightning speed, and some machines are already capable of making radical and spontaneous self-improvements. (***See below***.)

What safeguards are in place?
Not many thus far. Google, for one, has created an AI ethics review board that supposedly will ensure that new technologies are developed safely. Some computer scientists are calling for the machines to come pre-programmed with ethical guidelines — though developers then would face thorny decisions over what behavior is and isn't "moral." The fundamental problem, said Danny Hillis, a pioneering supercomputer designer, is that tech firms are designing ever-more intelligent computers without fully understanding — or even giving much thought to — the implications of their inventions. "We're at that point analogous to when single-celled organisms were turning into multicelled organisms," he said. "We're amoeba, and we can't figure out what the hell this thing is that we're creating."

When robots learn to lie
In 2009, Swiss researchers carried out a robotic experiment that produced some unexpected results. Hundreds of robots were placed in arenas and programmed to look for a "food source," in this case a light-colored ring. The robots were able to communicate with one another and were instructed to direct their fellow machines to the food by emitting a blue light. But as the experiment went on, researchers noticed that the machines were evolving to become more secretive and deceitful: When they found food, the robots stopped shining their lights and instead began hoarding the resources — even though nothing in their original programming commanded them to do so. The implication is that the machines learned "self-preservation," said Louis Del Monte, author of *The Artificial Intelligence Revolution*. "Whether or not they're con scious is a moot point."

Possible Response Questions:
- Are you concerned about the development of AI? Explain.
- Select any passage and respond to it.