AI in Movies, Books, and Television: From Golems to Cylons

People have been dreaming about artificial intelligence for a few thousand years. The forms AI has taken have ranged from the manmade monsters and automata of earlier centuries to today's robots and computers. Many of these stories and legends are what can be called cautionary tales. These are warnings about what can go wrong if work turns out badly. These same cautionary tales exist today. They reflect worries about how technology and pride might end up harming human civilization.

At the same time, there are those who see technology as being wholly good and who are convinced that technology will help to solve humanity's problems. These are the same people who tend to think of robots and AI work as more likely to be good than bad.

Whether good or bad, robots and intelligent computers show up in both ancient and modern legends. Some of the modern legends take the form of books, movies, and television shows.

Why Stories About AI Are Popular

Humans are considered to be the only intelligent species on Earth. They are also the only known intelligent species in the universe. Yet, our culture is full of stories, TV shows,

(continues)

Figure 4.1 Captain Jean-Luc Picard (actor Patrick Stewart) was assimilated into the Borg, becoming Locutus, in the 1990 *Star Trek: The Next Generation* episode "The Best of Both Worlds: Part 2." His assimilation enabled the Borg to acquire all of the captain's knowledge and experience.

(continued)

and movies about alien life and artificial intelligence. There are stories in the news all the time about people who think they've seen flying saucers or aliens. It's natural to wonder why we spend so much time reading and watching stories about aliens and intelligent robots and computers. Our sense of loneliness as a species might be a reason. Maybe these stories give us hope that one day, we might share our planet with other intelligent beings. Another form of intelligence would help us feel less alone in the universe.

Stories about artificial intelligence also encourage us to think about ourselves. They force us to think about how other forms of intelligence might act and how they would view humans. For example, the Borg of *Star Trek* are part of a collective in which no member is really an individual. Watching shows about the Borg give us a chance to think about how much we value our own individuality. Stories about AIs give us a chance to think about what it's like to be human.

FRANKENSTEIN'S MONSTER

Some of the earliest legends of artificial beings were about monsters. The Golem of Prague is one of these manmade monsters. However, probably the most famous one is Frankenstein's monster. In the story, written by British novelist Mary Shelley, Frankenstein isn't the monster—rather, he is the doctor who *made* the monster. Frankenstein is not trying to create a monster as much as he's trying to find a way to bring life to non-moving materials. To do this, he puts together a man using bones and other materials he's collected. He then uses electricity to jolt the creature into life. Frankenstein had hoped his creation would be beautiful. When it turns out to be horrible, he becomes disgusted with his creation and he leaves it, hoping he can forget about it.

While Frankenstein goes about his life, his monster hides in the mountains of Switzerland. There, he watches people and becomes **self-aware**. Years later, by accident, Frankenstein meets his

Figure 4.2 In the 1931 film *Frankenstein*, Dr. Henry Frankenstein (actor Colin Clive) and his assistant, Fritz (Dwight Frye) examine the bandaged monster (Boris Karloff) they've created after attempting to give life to a corpse.

monster again. This sets into motion a series of events that ends with the death of Frankenstein and his wife. After, the monster makes his way to the North Pole. Miserable about having killed so many people, he is determined to kill himself.

The subtitle of Shelley's 1818 novel is *The Modern Prometheus*, comparing Frankenstein to the ancient Greek hero who brought the fire of the gods to humanity. In that legend, Prometheus does humanity a favor. By the nineteenth century, when *Frankenstein* was

written, fire was also seen as having made possible modern warfare and the Industrial Revolution. These changes brought both bad and good into the world. Thus, Dr. Frankenstein and his monster had both good and bad aspects to them. Yet, in the end, both Frankenstein and his creation caused harm and grief. Their lives ended in tragedy. By the last chapter of the book, the reader is left thinking that Frankenstein's work was dangerous and that it brought harm into the world. *Frankenstein* was first published almost two centuries ago. Yet its themes have been seen over and over again in literature and, more recently, in movies and television.

INTELLIGENT ROBOTS: R.U.R, ASIMOV'S ROBOTS, AND MARVIN

Anyone who pays attention to modern science and technology hears the word *robot* used all the time. Robots help to build cars, disarm bombs, explore planets, and perform surgeries. As it is used today, the word *robot* refers to any machine that is able to perform some tasks on its own. The task or tasks are usually guided by either a person or, more commonly, by a computer. Yet, robots in fiction are usually much more than this. Fictional robots don't just weld together car frames. Many look more or less like people and often act on their own. They are often intelligent, and some even have feelings, or are trying to discover them.

R.U.R.

The word *robot* dates back to the play *R.U.R.,* written in 1921 by Czech playwright Karel Capek (1890–1938). *R.U.R.* stood for *Rossum's Universal Robots*. In Capek's play, the robots were artificial beings created to serve humans. Capek got the word *robot* from the Czech word *robota*, which translates to "forced labor" or "slave." *Rossum* seems to come from another Czech word, *rozum*, which means "wisdom." The play itself is about the robots, which are self-aware and have come to pose a threat to humanity. By the end of the play, not only have the robots overthrown humanity, but they've also developed emotions. Some have even fallen in love.

Asimov's Robots

R.U.R. gave us the word *robot,* but the writer most responsible for the way so many of us think of robots was the great science fiction author Isaac Asimov (1920–1992). Asimov wrote four robot novels and many short stories in the 1950s, 1960s, and 1970s. Among them was the classic *I, Robot,* as well as a series of novels about a robot detective. Though they did not have feelings, Asimov's robots were intelligent. They were used mainly as workers, often in places or on jobs that were too dangerous for humans. Unlike the robots in *R.U.R.,* Asimov's robots did not revolt. They were not able to do so.

Marvin

Both Capek and Asimov wrote about the moral issues of making and having intelligent robots. They represented opposite sides of the good robot / bad robot issue. But there is more to fictional robots than just stories about helping or destroying humanity. This brings us to a third example: comic robots. The best example of this is Marvin the Paranoid Android from the book, radio, and television series *The Hitchhiker's Guide to the Galaxy*, written by British author Douglas Adams (1952–2001).

In Adams's novels and the various shows made from them, Marvin was a robot with a "brain the size of a planet" that was stuck doing simple errands on a starship. Marvin really was neither good nor bad. He was just depressed. Unlike Asimov's robots and those of Capek, Marvin never faced any moral problems. Also, he never thought about attacking the humans he served. In fact, he didn't spend much time dwelling on his servitude. He just remained hilariously miserable throughout Adams's stories.

Intelligent Computers: *The Moon Is a Harsh Mistress*, HAL-9000, and Colossus

Artificial intelligence descriptions in literature and art often reveal a lot about the technology of the time. Mary Shelley, the author of

Asimov's Laws of Robotics

Isaac Asimov came up with the Three Laws of Robotics that were programmed into every one of the robots about which he wrote. The three laws ruled all aspects of their behavior. Asimov realized that robots could easily be smarter and stronger than humans. These traits might lead them into conflict with the very humans who designed them.

Asimov didn't want to write a series of stories about human-robot wars. Instead, he thought it made sense to have laws imprinted in every robot that would keep them from rising in revolt. Asimov felt that his Three Laws of Robotics were so commonsense that he couldn't see humans ever designing robots that were not ruled by the laws. Asimov's laws are:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

No robot following these laws could revolt the way Rossum's Universal Robots did. They could not enslave or attack humanity like robots did in the movie *The Matrix*. Also, they would not be able to assimilate people like the Borg. Ruled by these laws, Asimov's robots would be humanity's servants and equals. However, they would never become humanity's enemies.

Frankenstein, lived in the 1800s. This was way before there were any computers or mechanical men. This is why Shelley wrote about bringing a monster to life. She lived in an age when this is what seemed the most likely way to create an intelligent being.

By Asimov's day, we had computers, but they were huge. In the 1950s, a computer that filled an entire room had less power than today's mobile phones. When Asimov wrote about intelligent computers, they were big machines that could interact with humans. However, his computers were hardly mobile. Asimov's robots did not have computers in their heads. Rather, they had artificial brains. As technology got better, writers felt it might be possible to create computers that were powerful enough to be intelligent. These very computers could also still be small enough to fit inside a single building. In the 1960s, authors wrote about self-aware computers. Whether by accident or on purpose, these computers developed intelligence.

Mike

One of the first examples of this was in an award-winning book published in 1965, *The Moon is a Harsh Mistress.* It was written by the acclaimed science fiction author Robert A. Heinlein (1907–1988). In the story, Earth's lunar colony had a main computer named Mike that ran most of the settlement's machines and systems. After years of people adding on to the computer, it had become as complex as a human brain. Once self-aware, it finally woke up. The main computer technician became Mike's friend.

Early in the book, they both joined a movement to free the lunar colony from an oppressive Earth government. Mike was valuable to the revolution, since he was tied into every system on the moon. He helped sabotage efforts to keep the revolutionaries under control. He also interfered with the administration's communications. In the end, Mike helped the rebels to organize their activities. He created an artificial persona called Adam Selene who ended up as the revolutionary leader. In the final chapter, following an Earth attack on the colony, Mike suffered damage. He never communicated again, except as a computer.

What makes Mike unique is that "he" is the first artificial intelligence in literature to be written as a fully developed character. Mike wasn't just a tool used by the revolutionaries. He was a participant in the revolution. In fact, he not only helped the revolutionaries, but he also became their friend. At the end of the book, when Mike stops responding, the narrator mourns the loss of his friend by

crying. There have been scores of books, movies, and television shows with characters who are computers or robots, but Mike was one of the best and was certainly one of the most well-written characters.

HAL

At almost the same time as Heinlein wrote about Mike, another great science fiction author, Arthur C. Clarke (1917–2008), wrote about another intelligent computer, the HAL-9000 in his 1968 book *2001: A Space Odyssey*. Unlike Mike, HAL was onboard one of humanity's first major interplanetary spaceships. HAL was charged with running the ship during a long mission to Saturn. (In the movie version, they voyaged to Jupiter.) Unlike Mike, HAL was designed to be intelligent. Also, unlike Mike, HAL went through a teaching process. This is when "he" learned to talk, read lips, and even enjoy art. Although HAL was supposed to care for both the ship and the crew on its mission, he was also given secret instructions that couldn't be shared with the crew. Eventually the stress of keeping a secret from the crew caused HAL to become paranoid. He ended up killing all but one member of the crew, Dave, who eventually disconnected HAL's circuits, "killing" the computer.

What makes HAL interesting is that he is a more complicated character than Mike. There was never any doubt that Mike was a good computer. He helped the revolutionaries through the entire book. HAL, on the other hand, started out as a good computer but became bad when the secrets he was forced to keep drove him insane. So not only does HAL change from good to bad over time, but he is also the first fictional computer to go insane. Interestingly, what drove HAL off the edge were instructions from the humans who forced him to keep secrets.

Colossus

Around the same time Heinlein and Clarke wrote about Mike and HAL, another science fiction author broke onto the scene. Dennis Jones (1917–1981) also wrote about an intelligent computer, Colossus, in his 1966 book of the same name. Colossus was designed to run the military defense of North America. He was given full

Figure 4.3 HAL-9000, lurking in the background, eavesdrops on astronauts Frank Poole and Dave Bowman (actors Gary Lockwood and Keir Dullea) in the film *2001: A Space Odyssey* (1968), directed by Stanley Kubrick.

control over the United States's nuclear weapons. After becoming fully operational, Colossus told the military that the Soviet Union had a similar computer and he asked to be connected to it. At the same time, the Soviet computer called Guardian requested the same thing of the Soviets. The connected computers were more intelligent than humans. Eventually, they took control over the nuclear weapons and then, the world. Colossus claimed it was doing this to save humanity from itself. He also stated that he was working to improve living conditions for the world's populations.

Colossus is very different from Mike and HAL. Unlike Heinlein's or Clarke's computers, Colossus is not much of a character in the book. It's easier to think of Colossus as "it" rather than "he." Also, while Colossus claims to be looking out for humanity's best interests, its methods involve taking control of humanity and threatening

to use nuclear weapons against them. Colossus looked at this the same way a parent might look at punishing a child who had been kicked out of school for fighting—by inflicting a short-term pain in the hopes it would bring a long-term benefit. Humanity (like most kids) didn't quite see it the same way. So in spite of Colossus' stated desire to save humanity, this computer still comes across as being bad because it's trying to control the world. Mike served his friends, HAL tried to kill his crew, but Colossus was the first to take control over all of humanity. This, plus its threat to use nuclear weapons against people (albeit in order to protect them from an all-out nuclear war) and its lack of emotions made it seem worse than Mike, or even HAL.

It's interesting that all three of these books were written at very nearly the same time. Each book portrays a different view of AI: the good, the bad, and the confused. It's also interesting that even the bad computer (Colossus) turns out to be far better than the bad ones that were to come. By the 1980s and 1990s, AI's fictional bad guys didn't threaten to save humanity by controlling it. The next generation threatened to wipe out humanity or to control the galaxy.

AI GONE BAD: TERMINATORS, THE BORG, AND THE MATRIX

Examples of bad artificial intelligence go back hundreds of years. The earliest bad AIs usually only threatened a handful of people, not all of humanity, but bad actions don't have to put all of humanity at risk for them to be bad. After all, we agree that criminals are bad even though their actions usually affect only a handful of people. Towards the end of the Cold War—a decades-long period of near-war from the 1940s to the 1990s that continually threatened to turn into all-out nuclear war between Communist nations and those of the West—people worried frequently about nuclear war. They also worried about losing control over technology. At that time, the bad AIs seemed more threatening. The AIs of the *Terminator* movies and TV series tried to destroy humanity. The hybrid biological-mechanical Borg of the *Star Trek* universe wanted to absorb all of the

technological species in the entire galaxy. The AIs in the 1999 movie *The Matrix* (as well as the 2003 follow-ups *The Matrix Reloaded* and *The Matrix Revolutions*) enslaved humanity, using them as a source of energy.

The idea of AI gone bad seems to have become more popular in recent years. This trend might be because technology is advancing more and more quickly, and this is making people nervous. People are often scared of things they don't understand. The rapid advances in all kinds of technology in the last few decades has made people worry more and more that the technology might be advancing too quickly. These worries make people wonder if technology might get to the point where humans can't control it anymore. In other words, intelligent machines might not be willing to serve people but, instead, might try to take control—or worse.

The Terminator Universe

The *Terminator* movies and television series are examples of robots *and* computers that are corrupt. The main theme of the *Terminator* movies is familiar: Intelligent computers gain control over humanity's weapons. Then, they declare war on humans and try to wipe them out, but a small group of humans fights back. Part of the struggle includes Skynet (the artificial intelligence) sending intelligent robot assassins (called Terminators) back in time. They try to kill the mother of the future leader of the human resistance before she can have her child. This is to prevent the resistance from forming in the first place.

Although the *Terminator* series is pretty bleak, it is not all gloom and doom. In addition to the Terminator robots sent back in time, the resistance leader is able to reprogram a Terminator to protect his mother and his younger self. It returns in time as a good robot. So, in the *Terminator* series, not every AI is bad. However, the good ones are that way because they've been reprogrammed by people to help the remaining humans fight Skynet.

In the *Terminator* series, Skynet can be seen as bad. It started the war against humans without any real reason. The robots, though, aren't really good or bad. Rather, they are whatever they are programmed to be. The robots programmed by Skynet are bad because they are following the instructions given to them by the computer

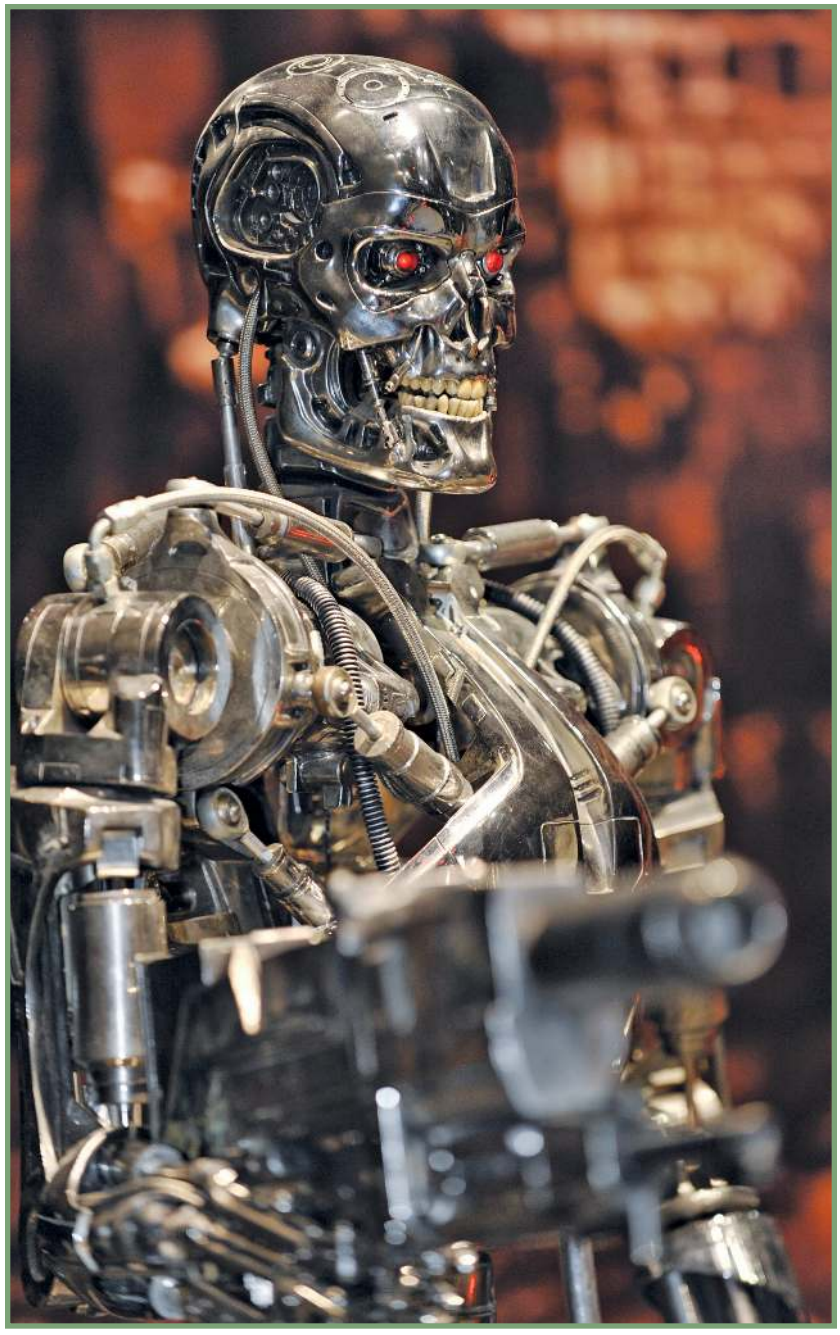

Figure 4.4 Terminator robot T-800 was featured in the movie *Terminator 2: Judgment Day* (1991).

system. The robots reprogrammed by humans are good because they were programmed to help. Even though the people watching the movies are tempted to see all of the Terminator robots as being bad, the reality is that they are only as bad or good as they're programmed to be. It's Skynet that's bad—not the robots.

The Borg

There are many bad AIs represented in movies and television. However, the most frightening one might be *Star Trek*'s Borg. The Borg are **cyborgs**, living beings that have merged with machines. The Borg joined to form a giant artificial intelligence they call the Collective. When the Borg meet another species, one of two things can happen: They can assimilate them and make their victims lose their identities. Or, they can wipe them out. The Borg don't allow for the possibility of leaving other species alone. Everyone must be either assimilated or eliminated.

What the Borg do to those who are assimilated is make them part of the Borg Collective. The entire Borg race is linked together into a single intelligence. The members of this intelligence are either artificial or natural in origin. In some ways, the Borg are more frightening and evil than either their *Terminator* or *Matrix* counterparts. At least those computers let humans keep their individuality even as they enslave or kill them.

The Borg, in contrast, threatened to take humans' individuality and turn everybody into an identical piece of the Collective. People who are enslaved can fight for freedom. People who are at war can fight for survival. But those who have been stripped of their personalities and blended into the Collective are unable to fight. They are unable to understand that fighting is even possible. Individuality is a highly valued human trait, making us different from everyone else. This is what the Borg threatened to take away and this is why the Borg are so disturbing.

The Matrix

Following the *Terminator* movies and *Star Trek's* Borg came the *Matrix* movies, set in some unknown time in the future. Most of what is left of humanity is under the control of an intelligent

Figure 4.5 In the Matrix trilogy of films, computer hacker Neo (actor Keanu Reeves, right) learns about the real world—a wasteland where most of humanity has been captured by a race of machines that live off of human body heat and imprison human minds within an artificial reality known as the Matrix—from Morpheus (Laurence Fishburne, center) and Trinity (Carrie-Anne Moss, left).

supercomputer. This control begins from the moment humans are born to the moment they die. The machines that rule the world use the humans as a source of energy while feeding them a virtual world of sights, sounds, and sensations. This virtual world is called the Matrix. The only free humans are a small group living in a city called Zion. They learned of the illusions previously fed to them, freed themselves, and went to war against the machines. In *The Matrix,* the computers are in almost total control. They aren't trying to eliminate humanity as much as enslave humanity. People are viewed as natural resources.

The artificial intelligence in *The Matrix* isn't trying to wipe out humans. In the movie, many years before the story takes place, the machines fought a war against humans and won. On the one hand, one could say that the machines in the movie take care of the humans. Their physical needs of food, water, and shelter are provided. Humans are also given a realistic virtual world to "live" in. But with all that, there's no way to pretend that the machines are good. Rather, they are the masters and, except for the small group of humans in Zion, the slaves don't even know that they have lost their freedom. The world as it exists in *The Matrix* is a gloomy one, but at least humanity is alive and a small group is able to fight and free itself.

ANDROIDS: COMMANDER DATA, CYLONS, AND C3PO

So far, we've seen monsters created by people, computers that became self-aware, computers that gain intelligence, and intelligent machines that tried to harm humanity. The next category to discuss is the androids. These are robots built to look like humans.

Commander Data

One of the most popular AIs in the 1980s and 1990s was an intelligent android named Commander Data. This character was shown in the 1987-1994 television series *Star Trek: The Next Generation*. Commander Data is modeled after Asimov's robots. The Three Laws

of Robotics rule his behavior. What makes Data so unique among fictional AIs is his struggle to learn more about what it might be like to be human. Data spends his life trying to learn about emotions and finding a way to feel emotions. He understands that the ability to feel is the main difference between himself and humans.

Data's story raises moral questions. They are the questions we might face at some point in the future. In an early episode, Data's shipmates must help decide if he is a machine—and thus the property of Star Fleet—or if he is a person with the same rights as any human. Other episodes raise similar philosophical questions. These episodes feature Data trying to create a "child" (another android modeled after himself). They also show Data figuring out the meaning of family (Data's "father" and "brother"), what friendship and love are, and much more. Watching Data try to understand humans helps viewers understand what being human is all about.

The Cylons

If Commander Data is a great example of a good android, then the Cylons of the later *Battlestar Galactica* series (as well as a television mini-series and a made-for-TV movie) are perfect examples of what bad androids can be like. Humans created the Cylons to be intelligent warriors and servants. The Cylons revolted, fought against their human creators, and tried to wipe them out. After a brief peace, the Cylons attacked again, killing almost all humans and forcing the survivors to flee for their lives. Towards the end of the series, some of the Cylons grew to feel it was immoral to kill all humanity. They rebelled against the other Cylons and joined forces with the surviving humans. Eventually, the surviving humans and their "good" Cylon allies found a home where they could all be safe.

What is interesting about the Cylons is that they change throughout the series. The way the audience views them changes, as well. At first, they all seem evil, with a single goal of destroying humanity. As the series goes on, though, viewers learn that the Cylons are like humans in some ways. They worship a god, can fall in love (even with humans), and are loyal or disloyal. By the end of the series, some of the Cylons are so similar to humans that many viewers find themselves wondering if there is really much of a difference between humans and the Cylon androids they had designed.

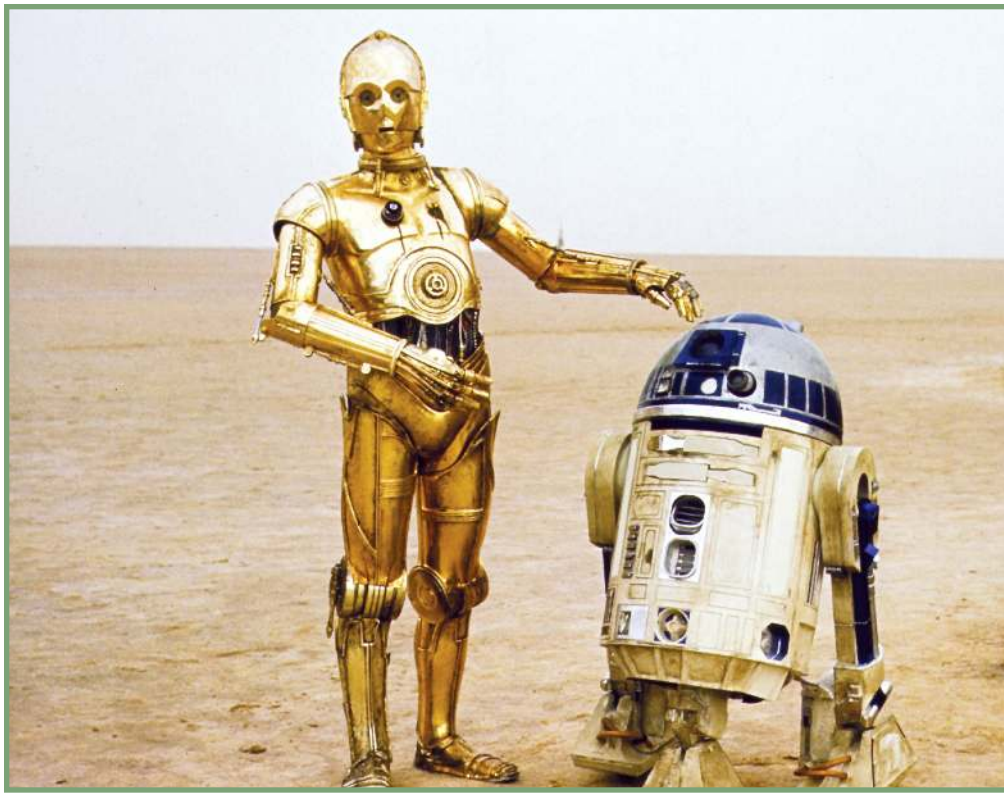

Figure 4.6 C-3PO and R2-D2 (actors Anthony Daniels and Kenny Baker), the expressive robots in director George Lucas's 1977 film *Star Wars*, are examples of human-friendly robots.

C3PO

If Data is one of the most famous of the good androids, then the *Star Wars* android C3PO is almost certainly the most loved. This is interesting, as he is also the least humanlike of the three discussed here. C3PO had a metal exterior and his machinery could be seen. Although C3PO looked less human than either Data or the Cylons, he acted the most human. He had a sense of humor, felt the whole variety of human emotions, and genuinely cared for his human and robot companions.

What we learn from C3PO is that we are only truly able to feel comfortable with other humans. Data was an android that the

viewers could like and respect, but it's hard to love him because he was so obviously not a human. The Cylons weren't humans, either. Viewers could fear some and could admire those who threw their lot in with humanity. However, viewers could never really feel comfortable with the Cylons. This is because, no matter how human they looked, their past actions always left doubt that they might go back to their old ways. Both Data and the Cylons looked and behaved humanlike, but never enough to be completely human.

The other trait neither Data nor the Cylons had was a sense of humor. This was something that C3PO had in abundance. Despite the fact that he looked less human than either the Cylons or Data, C3PO acted more human than either one of them. This included his ability to make jokes. Critics note that this human quality endears him to many viewers.