MARCH 12, 2015

TEACHING ROBOTS TO BE MORAL

BY GARY MARCUS

A scene from “Chappie.”

PHOTOGRAPH COURTESY COLUMBIA PICTURES/EVERETT

“Chappie,” the highest-grossing movie in America last weekend, is, to put it mildly, not a great film; the critics have given it a twenty-nine on Rotten Tomatoes, and it is nowhere near as original as “District 9,” an earlier effort by the director, Neill Blomkamp. “Chappie” does not have the philosophical depth of “The Matrix” or the remade “Battlestar Galactica” series. Nor does it have the visual panache of “Interstellar (http://www.newyorker.com/magazine/2014/11/10/love-physics)” or “2001.” From its opening scene, the film comes across as little more than a warmed-over “RoboCop” remake, relocated to Johannesburg. There’s an evil company man, droids that menace the population, and a whole lot of blood, shooting, and broken glass. About the only thing that seems new is the intermittently adorable android protagonist, Chappie.

Meanwhile, most of the science fiction in the movie utterly fails as science. The plot, for example, revolves around a portable scanner that reads the contents of both human and robot brains. It makes sense that someone would be excited to study a human’s brain in a magic electronic skullcap. But if you were analyzing a robot, wouldn’t be it easier just to download the software? Spike Jonze’s “Her” (http://www.newyorker.com/culture/richard-brody/aint-got-no-body) presented a future that seemed utterly believable; Blomkamp’s film presents a mishmash. And yet, as bad as most of it is, I walked away with a newfound respect for the need for science fiction in a world of rapidly changing technologies. Even bad films can raise profound questions.

The interesting parts of the film stem almost entirely from Chappie. Whereas most science-fiction androids enter the scene as a fully functioning “adults,” Chappie enters as a virtual newborn. Initially, it doesn't know a word of English, nor even the most basic facts about the world.

"Chappie" takes its inspiration from a real (if small) field of artificial intelligence known as developmental robotics, in which simple robots like iCub (http://www.icub.org/) learn by doing. For now, that field has produced few results of any significance. Most robots today are largely preprogrammed; Sony's late AIBO robot came out of the factory walking, and the only thing that your top-of-the-line vacuuming robot ever learns is the layout of your home. The adorable commercial robot Baxter (http://www.rethinkrobotics.com/baxter/) does better. It can learn new tasks as it works, such as how to pick up objects of certain sizes and move them to particular places. But, at least for now, Baxter's learning abilities mainly consist of variations on themes; a certain basic set of skills is installed in the factory, and it seems doubtful that Baxter could ever learn on its own to do something novel, like knit or juggle.

Chappie is thus every roboticist's dream—a combination of hardware and software that can learn pretty much anything. In a single week, Chappie grows from infant to toddler to surly teen-ager. It masters everything from the English language to the fine art of throwing knives at moving targets. The film never tells us how Chappie learns so fast, but pretty much every software engineer and developmental psychologist I know would love a machine that could match Chappie's skills. Ultimately, though, the movie is about something different: not the cognitive scientist's question about how intelligent creatures manage to learn about the world but the educator's question about how human beings can raise moral robots. Chappie doesn't just learn a set of facts; Chappie learns a set of values.

As any parent knows, teaching values is hard. Early in the movie, Chappie's creator tries valiantly to teach the robot the difference between right and wrong, but even the most basic lessons fall flat; the robot understands that killing people is wrong, but is seduced by a rather less savory character into believing that merely cutting people—to "help them go to sleep"—is acceptable. Children often learn more from their peers than their parents, within minutes of his activation, Chappie, surrounded by a bad crowd, is unable to sort right from wrong.

None of us should want that. In a future world ever more populated with robots, we'll want them—whether driving cars or taking care of the elderly—to have some sort of moral compass. Elon Musk recently warned (http://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat) that artificial intelligence is the greatest existential threat to mankind. He may be overstating things, but he isn't wrong to be concerned that, in building robots and artificial intelligence, we could potentially unleash a demon. We do, as I argued here, in 2012, need to learn to build moral machines (http://www.newyorker.com/news/news-desk/moral-machines). Robots and advanced A.I. could truly transform the world for the better—helping to cure cancer, reduce hunger, slow climate change, and give all of us more leisure time. But they could also make things vastly worse, starting with the displacement of jobs and then growing into something closer to what we see in dystopian films. When we think about our future, it is vital that we try to understand how to make robots a force for good rather than evil.

"Chappie" isn't, of course, the first bit of science fiction to point that out. But the film raises the question in a different way. In Isaac Asimov's imagination, robots came straight from the factory programmed to obey three laws, starting with "A robot may not injure a human being or, through inaction, allow another human being to come to harm." In Blomkamp's imagination, robots don't come from the factory with any laws; they learn from their makers, companions, and what they see in the world around them.

Asimov's factory-installed laws seem fine for a short story, but inadequate for the real world. Even if they were the right laws, we certainly don't yet know how to turn them into computer code. How, for example, do you even translate the concept of harm into the language of zeroes and ones? But, and this is Blomkamp's point, learning morality is fraught with problems, too. Human children learn their values in at least two ways—through explicit instruction ("Stealing is wrong") and through observation ("What do my parents do? What are my friends getting away with?"). As the Yale psychologist Paul Bloom has argued, we also seem to be born with the innate underpinnings of a moral sense. Future robots will likely be similar. Their decisions will inevitably be guided by a mixture of what is preprogrammed and what they learn through observation. How will we make that work in a way that provides certainty about our own safety? Robots may someday serve as police officers and be pressed into deciding whom to protect and arrest; even in the home, dilemmas may arise. (What happens if an intruder breaks in and threatens the robot's owner?) In its best moments, “Chappie” can be seen an impassioned plea for moral education, not just for humans but for our future silicon-based companions.

How can we keep ourselves safe in a world in which we will be surrounded by autonomous steel contraptions that may someday be as smart as us, or even smarter? Blomkamp offers no real answers. That’s fine, and is his prerogative as a filmmaker. But society needs to think about these questions sooner rather than later. Contemporary robots are neither as dexterous as Chappie nor as quick to learn, but that’s just for now. No one knows what’s coming next.

Gary Marcus is a professor of cognitive science at N.Y.U. and the author of “Guitar Zero.”