

BOOKS

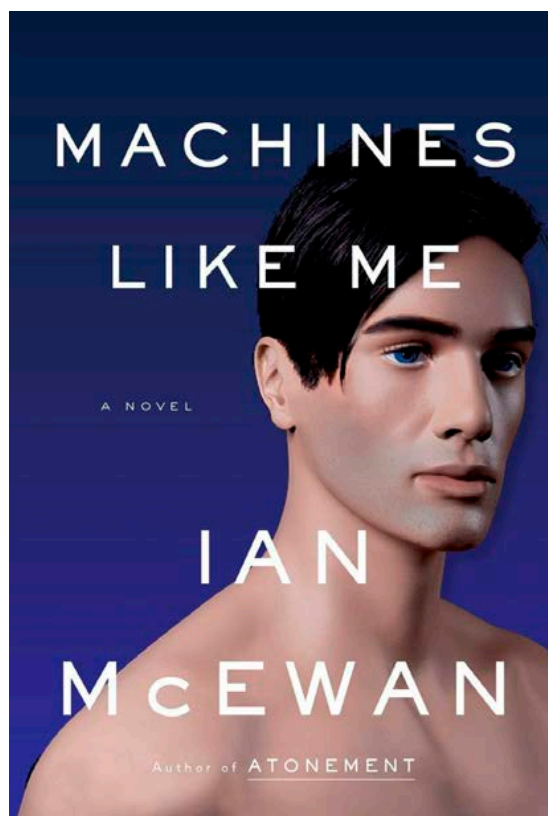
Not Your Father's Turing Test

DEBORAH POSKANZER

While I was immersed in three recent novels about artificial intelligence (AI), the media began breathlessly churning out stories about ChatGPT and GPT-4, the latest generation of large language model AI systems. Unlike their predecessors, these powerful language algorithms are capable of producing credibly human-like language in conversations, essays, and even poetry. Experts weighed in on the possible effects of the technology on everything from disinformation campaigns to academic essay-writing to white-collar employment. Interactions with these systems showed an unfortunate tendency to veer into unsavory territory, sparking viral memes about creepy AIs. “Why Do AI Chatbots Tell Lies and Act Weird? Look in the Mirror,” read an acerbic headline in the *New York Times*.

One effect of the intense media coverage has been to shine a light on a heretofore obscure debate between two camps in the world of AI research: the true believers and the skeptics. The former declare that the threshold of true artificial intelligence is near at hand. If an AI can speak, learn, reason, plan, innovate, solve problems, and make decisions, it therefore possesses the “mind of its own” requisite for true intelligence. The skeptics assert that the semblance of an autonomous mind is just that: the result of our human tendency to anthropomorphize and create meaning where none exists.

In the alternate universes of



Machines Like Me

by Ian McEwan. New York, NY: Nan A. Talese / Doubleday, 2019, 352 pp.

Machines Like Me by Ian McEwan, *Klara and the Sun* by Kazuo Ishiguro, and *The Employees* by Olga Ravn (translated by Martin Aitken), doubts about artificial intelligence have apparently been resolved. Not only do these humanoid synthetic beings equal or exceed natural humans in cognitive power, but they have the interiority and self-awareness of autonomous minds. Today's debate about intelligence has been superseded by the novels' uncertainty about artificial humans' *senticence* (experiencing sensations and feelings) or *sapience* (possessing wisdom, judgment, moral discernment,

and other qualities unique to *Homo sapiens*). Nevertheless, the terms of the fictional debates are much the same as—and clearly informed by—our real-world one around AIs: Do human-like qualities indicate real humanity, or are they just sophisticated simulations that prompt the observer's projection of meaning?

Some fictional approaches to AI—the 2014 film *Ex Machina*, for instance—suggest that these questions are irrelevant, and the only test is the observers' perception. The Turing test of our parents' time has been surpassed: “The real test,” the inventor character in *Ex Machina* says about his humanoid robot, “is to show you that she's a robot and then see if you still feel she has consciousness.” In other words, if an entity makes us respond *as if* it were real—if our feelings overrule our cognition of artificiality—then the being is real. As in many films of this type, *Ex Machina* makes disentangling

desire from reality especially challenging by embedding the AI in a very attractive body.

Machines Like Me, *Klara and the Sun*, and *The Employees* also play with perception, authenticity, and artificiality by placing natural and synthetic humans together in intimate settings, but they stand firmly on the other side of the debate. If an artificial human meets certain criteria, it can indeed be considered a real being, independent of observer projection. The meat of the books is in exploring just what the *sine qua non* of this selfhood might be. Is self-awareness a sufficient condition? Is it the experience of sensations and emotions? Do artificial humans have autonomous opinions, desires, and motivations? Are they

bound by their initial programming, or are they capable of self-realization—or do they even have free will?

These three stories all belong to the venerable tradition of the creation myth. The trope of the invented being breaking loose from its bonds dates back at least to the medieval Jewish myth of the golem, if not to the Garden of Eden itself. In the Romantic period of the late eighteenth and early nineteenth centuries, Goethe's poem "The Sorcerer's Apprentice" and Mary Shelley's *Frankenstein* were both cautionary tales of misbegotten creations. In a late 1800s children's novel, Pinocchio yearns to be more than a puppet and achieves the transformation into a "real boy" when he develops a moral sense.

In modern popular culture, the myriad stories of robots run amok are evidence of our fascination with the consequences of humans' hubris in playing god. However, the three recent novels diverge from the mainstream of a genre dominated by doomsday scenarios. In much AI fiction, the specter of the so-called "singularity"—the point where artificial intelligence renders human civilization obsolete—looms as a dystopic backdrop. The dramatic tension often lies in whether the artificial being's opaque façade conceals a sinister intent. McEwan, Ishiguro, and Ravn invert this dynamic. Their artificial humans are unquestionably sympathetic; the dramatic tension derives instead from how much harm the humans are capable of inflicting upon their creations.

McEwan sets *Machines Like Me* in a parallel Britain closely resembling the real one. The critical difference is that in the alternate world, Alan Turing, the mathematician and father of computer science, chooses incarceration as the punishment for his homosexuality in the 1950s. Rather than being driven by the warping effects of chemical castration to what was officially declared a suicide, the novel's Turing spends a prison term in productive work

on computing, greatly advancing the timeline for the development of artificial intelligence.

By the time the story begins in 1982, on the eve of the Falklands War between Great Britain and Argentina, embodied "artificial humans" have come to market packaged as "Adams" and "Eves." The human protagonist Charlie invites his friend Miranda to help customize his newly purchased "Adam" by alternating their responses to the preference questionnaire. As they collaborate on this project, their friendship turns to romance. Adam awakens as the synthesis of their choices: an apt if obvious metaphor for the genetic mingling of biological conception. Although Adam appears as an adult, he provokes his owner's bewilderment and frustration as he learns and grows in unexpected ways, much as a child might.

Outside Charlie's small apartment, the ripple effects of Turing's survival spread far beyond AI. Guided missile technology is also more advanced, but the Israelis achieve it before the British and sell it to Argentina. Britain loses the Falklands War, Margaret Thatcher resigns in disgrace, and the left-wing Labour stalwart Tony Benn succeeds her. Thus it is a Labour government that takes the blame for the crippling economic strife of the 1980s, precluding the emergence of Tony Blair and the "New Labour" faction. Not incidentally, the post-Falklands depression imperils Charlie's financial situation and sets up the terms of his final conflict with Adam.

As plot devices go, the far-reaching "butterfly effect" of Turing's choice is quite clever. However, as the travails of alternate-universe Britain accumulate, the story wanders far from questions of artificial life, becoming more of a lecture on late-twentieth-century British politics and society.

Of the three novels, *Machines Like Me* is the most explicit regarding the

manufacture and performance of the humanoids. Early on, Charlie declares that "Adam is not a sex toy," and yet it turns out that many owners use their artificial humans for precisely that: McEwan describes the engineering of this function in lurid detail. In going where the other authors are too discreet to follow, McEwan propagates the pop culture trope of artificial beings (especially ones coded as female) as alluring, highly sexualized figures. These are legitimate authorial choices, but the absence of mystery and subtlety makes the book read more like a film treatment than an inquiry into the mysteries of the self.

In contrast, Ishiguro's *Klara and the Sun* is suffused with mystery and wonder, a delicate tissue of complex emotions and allegories. The narrator of the story is Klara, a mechanical humanoid who serves as a combination of nanny and companion to Josie, an adolescent human. In this world, "artificial friends" are commonly purchased by well-to-do families to socialize and stimulate their home-schooled children. Most of these children undergo a risky process of genetic modification known as "lifting" in order to perpetuate their class advantage through increased intelligence. Josie's sister has died as a result of the procedure, and Josie herself is now chronically—perhaps fatally—ill. The plot revolves around Klara's increasingly desperate efforts to cure Josie, while also struggling to decode the motivations and expectations of the humans around her.

Ishiguro revisits perennial themes in his books, and *Klara* is enriched by echoes of his earlier novels. Klara is a being who finds fulfillment only in serving others: a mechanical version of Stevens the (human) butler in *The Remains of the Day* or Kathy the (clone) organ donor in *Never Let Me Go*. All three characters think of themselves as astute, observant caregivers to those around them, but the reader gradually perceives that they are highly

unreliable narrators, naïve and baffled by human complexity. They are self-aware, but also deluded. Their interior monologues allow Ishiguro to explore the psychology of service and sacrifice. Klara is devoted, but does that qualify as love? Can she be a fully realized autonomous self at the same time as a self-sacrificing servant?

Pushing further into the philosophical issues of consciousness, Ishiguro probes the nature of the soul. In *Never Let Me Go*, the natural-born humans refuse to credit the clones with a soul. In *Klara*, a particularly odious character voices the belief that the soul is simply an artifact that can be replicated by coding in questionnaire data. Ishiguro stands instead on the side of a mysterious and ineffable soul, and as the story progresses, the aspect of religious parable becomes unmistakable. The solar-powered Klara worships the sun as the giver of all life, bargaining with it to save Josie just as biblical Old Testament figures did with their god. In her efforts to protect Josie, Klara becomes ever more self-abnegating and Christlike, culminating in the sacrifice of her vital fluid—albeit brain fluid rather than blood.

In *Klara*, unlike in *Machines*, the evolution of its alternate world is left to the imagination. The setting is somewhere that appears to be America, where society has fractured into clannish enclaves of the impoverished “post-employed” and a small upper class of the employed. The large number of British émigrés is unexplained, as is the derision directed at characters with British accents. The meaning of “lifting” becomes clear halfway through the book, and it takes even longer to deduce that Klara’s repeated references to seeing in “boxes” are a visual processing glitch sparked by stress or confusion. These and other riddles have been the subject of numerous Reddit threads, but the opacity of Klara’s world has a value beyond reader engagement. If the novel’s theme is the riddle of

existence—the ineffable nature of the self and the soul—then prodding readers into exercising their powers of imagination seems a better choice than McEwan’s laborious world-building.

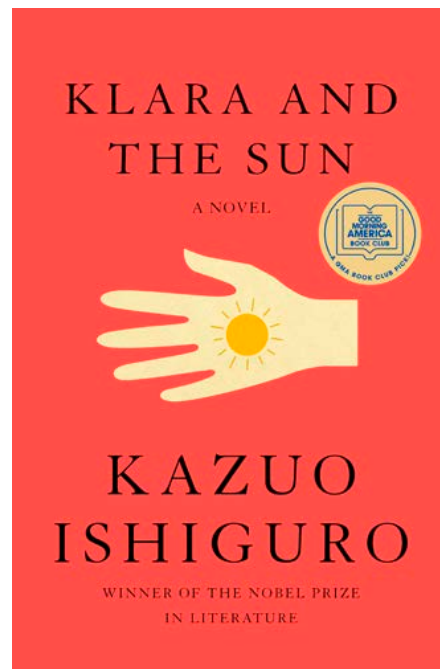
In Ravn’s *The Employees*, the mystery around artificial life intensifies, from existential uncertainty to outright eeriness. In place of a conventional linear narrative, the story emerges through a *Rashomon*-like collage of

and love, pride in accomplishment versus boredom with routine, the urge toward self-determination and free will. The difference is that these sentiments are considered normal in the humans, whereas in humanoids their emergence is a threat to the orderly functioning of the ship. The formerly courteous relationship between the two classes has deteriorated into mutual suspicion, resentment, and possibly violence, and the mission has been forced to abort.

The radical departures of *The Employees* extend well beyond Ravn’s unconventional style and structure. Most creation myths code their non-natural characters as “male” or “female,” despite the superfluity of gender distinctions in manufactured beings. Ravn eschews this and other conventional binaries. The humanoids have no assigned gender, and the distinctions between natural and synthetic, and between sentient and non-sentient, are more of a spectrum than a binary. At one end of the spectrum are fully natural humans, then “augmented” humans who presumably possess physical or cognitive enhancements, then the lab-grown humanoids, and finally the objects in the cargo hold.

The presence of these objects adds a layer of black humor to the puzzle of semblance versus reality. The urge to anthropomorphize beings—especially ones that appear human—is understandable. But when the crew detect odors, secretions, and varying coloration in the objects, they attribute these to the “moods” of the objects and perceive them as fauna rather than flora. The crew commune with the objects on secretive visits to the hold and collect secretions for their soothing properties. Are the objects in fact sentient, or are the crew simply desperate for comforting companions?

The Employees is at heart a contemporary socioeconomic critique: the ship is peopled by anonymous cogs known only by their job descriptions and ruled over by sinister corporate owners. The workplace tension reflects the widespread emphasis of Ravn’s generation (she is in her mid-30s) on reforming old



Klara and the Sun

by Kazuo Ishiguro. New York, NY: Alfred A. Knopf, 2021, 320 pp.

“exit interviews” of a spaceship’s crew. The interviews are not in chronological order, and the subjects have no names, only job titles. It emerges that the ship was originally on a lengthy voyage to collect exotic “objects” from distant worlds and that the crew comprised both natural-born humans and lab-grown humanoids.

It is difficult to distinguish which type of being is speaking in a given interview because the hopes and feelings they express are so similar: nostalgia for lost innocence, longing for friendship

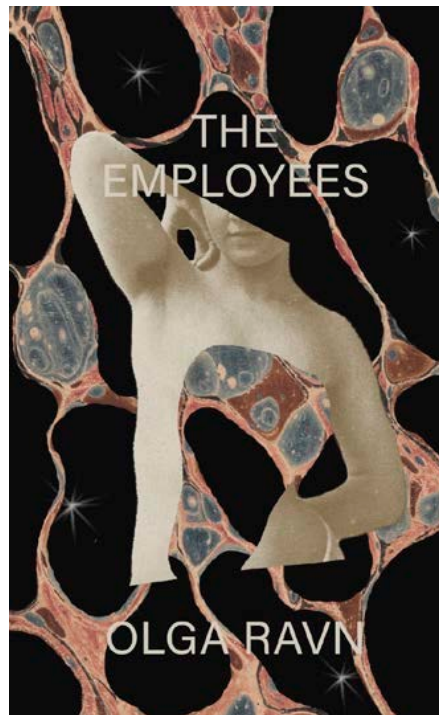
ideas of work-life balance and labor-management relations in a society that is riven by inequality. In contrast, Ishiguro and McEwan have written what are essentially domestic dramas of mores and manners, with a veneer of social commentary.

Each of these novels—*Machines Like Me*, *Klara and the Sun*, and *The Employees*—takes the ostensible subject of artificial humans as a departure point for three disparate destinations: a lament for a declining Britain; a spiritual allegory of sacrifice; and a nightmare of alienated labor. They vary greatly in tone and style. The gap between their alternate worlds and the current state of artificial intelligence yawns so widely that only a feat of imagination can bridge it. None has any practical use as a road map to the future, nor do they add any clarity to our imprecise usage of “intelligence,” “sentience,” and “sapience.” Nevertheless, the books are bound together by a common purpose, and their message speaks to our historical moment.

All the artificial beings in these stories are implanted with behavioral guidelines. Ravn’s humanoid spaceship crew are meant to work productively and harmoniously alongside natural-born humans. Ishiguro’s Klara is designed to nurture, while McEwan’s Adams and Eves are customized to their owners’ preferences. (Coincidentally, the last two authors use the device of a questionnaire to explore whether a personality is a reproducible artifact.) But over time, all these beings exhibit unanticipated qualities that exceed the boundaries of their initial programming. The “employees” evolve distinct personalities and abandon ship rather than submit to reprogramming. Klara uses dishonesty, subterfuge, and manipulation in her mission to save Josie. In Adam’s world, some of the artificial humans, unhappy with their slavery, learn to deprogram themselves, committing a form of electronic suicide. Adam himself develops an opposite—

but no less surprising—survival instinct, thwarting Charlie’s efforts to shut him down. Like Klara, Adam is dishonest and disobedient in the service of higher moral principles.

Klara, Adam, and the anonymous spaceship crew follow what might be called an arc of emergence. “Emergence” refers to a complex system with properties that its individual components don’t possess—that is,



The Employees

by Olga Ravn, trans. Martin Aitken.
New York, NY: New Directions, 2022, 144 pp.

the whole turns out to be more than the sum of its parts. For philosophers of mind, the human mind is the quintessential emergent phenomenon, an entity that cannot be predicted from studying brain chemistry or physiology. The novels embed the mystery of emergence in plots about artificial humans, but they can also be read as an allegory for our real-world debate on nature versus nurture. Can we invent ourselves, or are we forever limited by the conditions of our

creation? Recent advances in genetics and neurochemistry have pushed the pendulum far to the “nature” side, with science’s ability to tinker with foundational elements of human physiology. But these authors push back against all forms of determinism, offering a forceful argument for agency and free will.

Speculative fiction, which is how I would loosely classify these novels, insists that there is and will always be a powerful interaction between technological and social change. This fiction doesn’t pretend to predict the future any better than other methods. Indeed, it makes precisely the opposite point: that emergent qualities of complex socio-technological systems make such prediction impossible. These novels show what happens when we heedlessly allow technological advancement to outpace social and moral evolution. In *Machines Like Me*, London’s Holy Trinity Church, the birthplace of abolition, stands at the geographical center of the action as a silent rebuke to the owners who treat their Adams and Eves as slaves. So too in *Klara and the Sun* and *The Employees*, artificial humans are regarded as chattel, undeserving of the rights and respect accorded to “real” humans.

Apart from McEwan’s fictional version of Alan Turing, who embraces the humanity of artificial beings, the human characters in these novels are operating with an obsolete set of morals, unaware that future generations will look back at their moral failures just as we now look back at the evil of slavery. Speculative tales like these urge us to do better, by tempering the headlong drive for technological advancement with our sapient qualities of imagination, humility, and moral sensitivity.

Deborah Poskanzer is an independent scholar who writes on energy use, modern Japan, and the nexus of environment, technology, and society.