



LUDDITES, ROBOTS, AND CONSOLE COWBOYS: AI AND THE HUMANITIES

JOHN D. SCHWETMAN *

Abstract Artificial Intelligence (AI) has become a significant influence on how we teach, how we communicate, and how we imagine the modern workplace. To understand what it means and how it will shape our lives, it is necessary to examine it from a humanities perspective with special attention to history, ethics, and literary representations of AI. Our science-fiction literary tradition provides ample material to work in works by Mary Shelley, Karl Capek, Isaac Asimov, Ridley Scott, and William Gibson. In these works, the iconic literary figures of the Luddite, the robot, and the console cowboy embody some of the most important tendencies in the advance-theorization of AI. They reveal anxieties about employee-obsolescence, a fixation on anthropomorphism, and a lingering hope for organized resistance to power-imbalances. A carefully considered humanities framework provides a check on exaggerated claims from both pro- and anti-AI constituencies.

Keywords: artificial intelligence, science fiction, robots, literature, culture

1. HUMANITIES FACULTY AND DISTRUST OF ARTIFICIAL INTELLIGENCE

For many of my colleagues in the humanities, the emergence of Artificial Intelligence (AI) has caused a degree of consternation. There may be some exceptions, but, as a group, we are not happy about this development. The primary cause of our unhappiness? The various products that have emerged under the heading of AI in the past few years have intruded upon our efforts to teach our students to become better writers. Cheating, specifically through plagiarism, has been a problem in our classes for a long time, but generative AI applications have dramatically facilitated cheating in our classes. The threat of students turning to AI to write their papers for them has become serious enough to compel us to retool our writing assignments. AI has introduced new elements of distrust into a student-teacher relationship that had already been experiencing challenges. Furthermore, nobody consulted us before turning this AI bull loose in our humanities china shop.

1.1 Disruptions to Learning and Working Spaces

Humanities faculty in institutions of higher education are not alone in their hostility to the sudden arrival of ChatGPT, Claude, Gemini, and other variations on the Large Language Model

* Department of English, Linguistics, and Writing Studies, University of Minnesota Duluth

(LLM) that sprang into being and took over our classrooms and workspaces over the past few years. Opposition has emerged in the classroom, in the boardroom, in the health clinic, and in many spaces in between, often oriented around a few significant beliefs:

- AI threatens our employment and thus our livelihoods.
- AI pretends to be human when it is not and is therefore inherently deceptive.
- AI threatens to suck the humanity out of this world, leaving us in a chamber of soulless machines.

In this paper, I will argue that the humanities disciplines as they exist in our universities and colleges can inform debates over AI and mitigate some of the more prevalent exaggerations that influence the debate. The first concern, that it will displace many of us from the workforce, is perhaps the most generative of conflict, and we might turn to our colleagues in philosophy for an understanding of the ethics and political challenges associated the evolution of modes of production. The impression that AI is intrinsically deceptive results from our very human impulse to find ourselves, or at least our images, within the most alien of phenomena. This anthropomorphic tendency is prevalent in AI's aggressive marketing campaign and the doom-rhetoric of AI's most ferocious critics. And, finally, those of us who study language and literature are positioned to insist on the prioritization of human-produced art and culture. Our literary experience with AI and its precursors offers important reminders of the irreplaceable humanity at the heart of human enterprise, a core of perception and belief that we must protect at all costs. Indeed, AI is arguably not the cause of our shared hardship in this current cultural moment but is rather a symptom of our collective turning away from the concept of humanity, one that began long before the emergence of AI. If such is the case, the disruptions caused by AI may lead into a more serious conversation about what makes us human, and about how to recenter our humanity within the larger systems of measurement, observation, labor, and education that computer technologies have increasingly been mediating for us. In the meantime, AI adopted carefully and in moderation will likely provide the sorts of conveniences that past generations eventually absorbed from prior technological advances and used for their mutual benefit.

1.2 The Humanities Approach

In this spirit, the humanities will provide a helpful filter for these conflicts surrounding emergent LLMs and generative AIs. I will be examining the long history leading up to this new technology, considering what we might learn from that history, and applying it tentatively to our present-day situation. I will consider the history the Luddite Uprising, Mary Shelley's *Frankenstein*, the robots of Isaac Asimov's corpus, and the AIs of William Gibson's 1984 novel *Neuromancer*. This fictional legacy informs the fantasies that present-day marketers of AI have sought to exploit while also forming the basis for a countermovement based on equally imaginary scenarios.

In Carlo Rotella's article, "I'm a Professor, A.I. Has Changed My Classroom, but Not for the Worse," he persuasively sums up the anxieties of humanities professors in our current era, describing

the doomy picture of helplessness painted last spring and summer by well-circulated stories about generative A.I.'s effect on higher education. Everybody's cheating their

way through college. No student is going to read a book or write a paper on their own ever again. It's the end of the essay, of reading, of thinking.

Rotella 2025

In this summation, he captures some of the widespread discontent among the ranks of faculty in U. S. higher education institutions. Among the many complaints, three rise to the top:

1. Easy access to academic dishonesty. AIs became immediately and freely available to students, who were able to copy assignment guidelines and paste them into a ChatGPT prompt to produce one-click plagiarism. This is not merely a classroom issue. AI-facilitated dishonesty has radiated throughout the profession in such disparate areas as conference proposals, journal submissions, and job application materials. It has introduced a cloud of distrust over many aspects of academic life.

2. Art and poetry lose out. Humanities faculty are invested in the study of human-produced culture and believe artistic creativity is foundational to what it means to be human. We suspect that artists in all media will lose their jobs as algorithms replace them. We worry that algorithm-produced art will be stale, mechanical, and disloyal to the uniquely human experiences that it claims to represent.

3. Overblown marketing. The AI-vendors are part of a long tradition of inflated marketing claims in pursuit of customers, especially in the high-tech industry. The vast disparities between promised convenience and underwhelming products leave us all exhausted and distrustful of the next generation of techno-utopian products. Even technologies that once improved and enriched our lives have fallen prey to what critic and author Corey Doctorow has dubbed “enshittification.”¹ Does computer technology exist to improve the lives of computer users? Or does it exist to enrich the owners of big-tech shareholders? Many are uncertain about this.

While some of the rhetoric around AI is excessive and obsessed with doomsday scenarios, each of the above concerns is consistent with the very real impacts of AI that we have already experienced. Creative-class innovators have, indeed, already begun to lose their jobs as big companies respond to the pro-AI message.

But, all is not lost. A careful examination of anticipatory literary accounts of AI that have shaped the long narrative of technological change, a narrative that precedes the emergence of actual AI by many generations can usher in a more measured response to these changes. I will begin with the Luddites, a much-misunderstood organization of labor advocates in the early-nineteenth century. I will then consider the robots, mythical beings whose appearance in literature artfully anticipated their yet-to-be-realized appearance in the real world. And, finally, I will look at the “console cowboy,” an irreverent figure with unique provenance in William Gibson’s novel *Neuromancer* whose adventures point the way toward more subtle approaches

¹ In one particularly memorable example, Doctorow observes how Uber uses algorithms to tailor individual and distinct fares for each driver based on an algorithm that monitors driver behavior: “Tech leaders point to this stuff and call it innovation. It’s more accurate to call it obfuscation. Except for the level of indirection introduced by the presence of an app, the algorithmic wage discrimination would trigger labor law enforcement” (Doctorow 2025, 119). It is perhaps not a coincidence that we live an age of rapid advancement in technology that overlaps with our age of reduced labor-union representation and reduced employee protections.

to the human/machine divide.

2. LUDDITIES

According to Richard Conniff in his 2011 *Smithsonian* magazine article, ‘The Luddite disturbances started in circumstances at least superficially similar to our own. British working families at the start of the 19th century were enduring economic upheaval and widespread unemployment’ (Conniff 2011). Conniff notes that the label *Luddite* has taken on new meanings in our present-day circumstances, meanings that deviate from the original self-conceptions of the British textile-mill workers who took up the Luddite cause in the early 1800s. While it is true that Luddites smashed textile machinery as part of their protests against British labor policies, Conniff argues, the technology itself was not really the focus of their anger. Many of the Luddites productively used the cutting-edge technologies of the textile mills that employed them, but they did not always feel they received adequate compensation for this work. As Digital Humanities theorist Steven Jones explains,

we have to remember, the historical Luddites were themselves technologists—that is, they were skilled machinists and masters of certain specialized techne (including the use of huge, heavy hand shears, complicated looms, or large, table-sized cropping or weaving machines), by which they made their living. That living and their right to their technology was what they fought to protect, not some Romantic idyll in an imagined pretechnological nature.

Jones 2006, 9²

These activists were, in fact, labor organizers who mythologized a figure named “Ned Ludd” as an emblem of their cause in 1811. The actual Ned Ludd had smashed a machine called a “stocking frame” twenty-two years earlier to protest an overzealous manager’s advice about his knitting technique. The original Ludd’s protest had little to do with a fear of losing one’s job to a machine. Building on this peculiar history of labor activism, a new group identifying itself as neo-Luddites has emerged to apply Luddite activist strategies and theories to our new context. In this new context, too, the neo-Luddites really are anti-technology activists. They are no longer equivocating on that point.

2.1 Is it O.K.?

The term *Luddite* re-emerged in 1984 in an essay by American novelist Thomas Pynchon, who titled his work “Is It O.K. to Be a Luddite?” Pynchon’s answer to the question in his title is ‘yes,’ but he adds a few caveats. After all, as Pynchon explains, the Luddite is a cartoonish figure with dubious origins in the hotly contested Industrial Revolution. At the start of his essay, Pynchon cites the 1959 Rede-Lecture by C. P. Snow in which Snow decried the divide in Western intellectual life between “literary” and “scientific” faction. In this formulation, Snow

² Marcel O’Gorman uses this same quotation to support his argument for what he calls ‘DH neo-Luddism’ as he argues for an expansion of Digital Humanities beyond the bounds of simply ‘building digital tools’ and to include any classroom activity that serve ‘the purpose of questioning technocratic ideologies that are uncritical in their modes of digital production and that constrain innovation to the standards of efficiency and marketability’ (O’Gorman 2018, 124).

defined the scientists as figures with operational knowledge of the Industrial Revolution and the literary types as “natural Luddites” who harbored an unreasoned animosity toward industrialization. As Pynchon explains, ‘C. P. Snow’s use of the word [*Luddite*] was clearly polemical, wishing to imply an irrational fear and hatred of science and technology’ (Pynchon 1984, 42). While historical record reveals that the actual early-nineteenth-century Luddites in fact did not feel this hatred of science and technology, Snow’s mythical figure works well to explain the concept today. When Pynchon argues that it is okay to be Luddite in this article, he argues, too, that hostility toward, or at least skepticism about technological advances is not irrational. According to Pynchon,

Public feeling about the machines could never have been simple unreasoning horror; but likely something more complex: the love/hate that grows up between humans and machinery—especially when it’s been around for a while—not to mention serious resentment toward at least two multiplications of effect that were seen as unfair and threatening. One was the concentration of capital that each machine represented, and the other was the ability of each machine to put a certain number of humans out of work—to be ‘worth’ that many human souls.

Pynchon 1984, 44

As the stocking frames and machine-loom of the nineteenth century give way to the computers of the twentieth century, the same two concerns remain pertinent. In our own computer-saturated age, capital has continued to concentrate in the hands of a few extremely powerful people, and our fellow citizens have continued to lose their jobs to machines, now to machines that do things like processing data, answering telephone calls, and facilitating monetary transactions. The Luddites as Pynchon defines them remain as relevant now as they have ever been. And, as Pynchon has anticipated, the Luddites will continue to find themselves caricatured in the media as irrational technophobes.

On this last point, David Linton’s 1992 article titled “The Making of a Pariah: The Case of the Luddites,” is of particular interest. In this article, Linton examines the evolution of the term Luddite as it appears in reference books—dictionaries and encyclopedias. He echoes Pynchon’s claim that the term *Luddite* is subject to various inconsistencies, ‘Historians researching the period [of Luddite activity] generally agree that the origins of the term Luddite have not been conclusively established’ (Linton 1992, 406). In this analysis, Linton has examined ‘thirty different sources’ to give careful scrutiny to the nuances in how they define Luddite and how such definitions evolve over time (Linton 1992, 405). He reminds readers that the English Luddites have ideological cousins in France, some workers who ‘threw their wooden shoes (sabots) into machines, thereby giving us the word *sabotage*’ (Linton 1992, 408). In the end, as Linton explains it, while the Luddites acted logically and deliberately in defense of their livelihoods, they were doomed, even in putatively neutral reference books, to find themselves misrepresented as ‘primitive anarchists’ by media that was and remains under the editorial influence of proponents of capital. Indeed, as Linton notes, the Luddites themselves ‘don’t get off easy’ and were to face repressive countermeasures including deportation and hanging (Linton 1992, 411).

2.2 The Luddite Economics of AI in the Present-Day

The anti-Luddite narrative was strong in 1811 and remains strong to this day. Indeed, efforts

to rein in Luddite counterattacks on AI have accelerated apace. Josh Tyrangiel's analysis of the situation in a recent issue of *The Atlantic* titled "AI and the Future of Work" provides an in-depth consideration of the debate from the perspectives of economists. He situates AI within the larger historical narrative of automation of the workplace, noting that earlier modes of automation, like those that motivated the Luddite uprising, have been met with anxiety and, on occasion, violence. He then asks a worthwhile question about this history, which is 'whether we're approaching the kind of disruption that can be managed with statistics—or the kind that creates statistics no one can bear to count' (Tyrangiel 2026, 21). In other words, is there something truly unprecedented about AI as a disruptive technology, or it will it follow the same pattern as prior technological advances have?

Tyrangiel provides a comprehensive summary of a standard, optimistic take on the consequences of AI in which economists situate it within a very clear history of economic development:

Many economists insist that this will all be fine. Capitalism is resilient. The arrival of the ATM famously led to the employment of more bank tellers, just as the introduction of Excel swelled the ranks of accountants and Photoshop spiked demand for graphic designers. In each case, new tech automated old tasks, increased productivity, and created jobs with higher wages than anyone could have conceived of before.

Tyrangiel 2026, 20

In this rosy scenario, a deep knowledge of history provides the usual reassurances that there is nothing unprecedented about the types of disruption AI is causing in the modern workplace. In the long run, our resilient economic system absorbs these changes, employees adjust, and things move along as before, but with the added conveniences that these disruptive technologies have added. Furthermore, the adoption of AI in workplaces will not be able to proceed as rapidly as AI boosters would have us believe:

Integrating legacy tech with modern AI means navigating hardware, vendors, contracts, ancient coding languages, and humans—every one of whom has a strong opinion about the 'right' way to make changes. Months pass, then years; another company holiday party comes and goes; and the CEO still can't understand why the miracle of AI isn't solving all of the problems.

Tyrangiel 2026, 23

If these factors can slow down the adoption of AI, then the trauma of large-scale lay-offs can be forestalled and dissipated over a long enough timespan to allow workers to adapt to the changes. If this is true, then neo-Luddite rhetoric against AI may be as irrational as their critics claim it is.

Or not. Tyrangiel's article also explores the other possibility, which is that AI is an unprecedented disruption to the modern workforce, and he finds support for this in a conversation with 'super worried' economist Anton Korinek, who believes his optimistic fellow economists 'aren't misreading the data—they're misreading the technology' (Tyrangiel 2026, 23). Korinek regards AI as a truly revolutionary technology because it does not require

human overseers to install and integrate it into an office's larger constellation of legacy technologies. In Korinek's terms, 'in many ways [AIs] can roll themselves out' (Tyrangiel 2026, 24) or install themselves into existing systems with minimal human oversight. Furthermore, competitive market pressures could create feedback loop that is quite harmful to information-age workers' employment prospects:

Consider consulting firms, which have always charged high fees for having junior associates do research and draft reports—fees clients tolerated because there was no alternative. But if one firm can use AI to deliver the same work faster and cheaper; its competitors face a stark choice: adopt the technology, or explain why they are still charging a premium for human hours. Once a firm plugs in and undercuts its rivals, the rest must either race to follow or be left behind.

Tyrangiel 2026, 24

If Korinek is correct, and the more optimistic economists are wrong, then AI will spawn a sudden and massive spike in unemployment that potentially radicalize displaced workers and warp the democratic system that has shaped the United States government for generations. An angry populist response to emergent AI systems in the American workplace could lead to violent resistance or, more moderately, new policies restricting AI determined more by anger than by logic. Tyrangiel concludes his analysis of the politics of AI in America with a haunting image when observing, 'The trouble with pitchforks is that once you encourage everyone to grab them, there's no end to the damage that might be done' (Tyrangiel 2026, 28). Luddites might, indeed, evolve into a political majority in America, and the Luddite agenda could turn out to be a remarkably destructive one.

3. ROBOTS

Robots have been a mainstay of science fiction literature beginning with the very first work of science fiction, Mary Shelley's *Frankenstein*. For much of this literary tradition, science-fiction authors clothed their AIs human form, and this had real consequences for how readers engaged with them. It seemed like, once these entities had physical bodies, we almost immediately began to sympathize with them and to demand ethical treatment of these creatures. Human form, not human capacity, was originally the characteristic that compelled us to worry about how we treated them. In this section, I will examine a few variations on the robot as they appear in the works of Isaac Asimov, Karl Capek, Ridley Scott, and Mary Shelley.

3.1 *The Caves of Steel*: The Laws of Robotic

If Josh Tyrangiel anticipates an alarming future of populists rising up against AIs, his vision of impending upheaval is certainly not unique to him. Many years earlier, in 1954, a child of Russian-Jewish immigrants to New York named Isaac Asimov saw the publication of his novel *The Caves of Steel*, a trail-blazing work of science-fiction that imagined a future enriched (and disrupted) by emergent AI technologies. *The Caves of Steel* portrays civil unrest in a crowded future New York City, civil unrest motivated by collective hostility to robotic automation. In his account of violent mobs of working-class New Yorkers, Asimov deploys his futuristic robots as literary surrogates for New York's early-twentieth-century immigrant population,

which likewise experienced regular hostility because of perceptions that immigrants were unfairly taking away people's jobs. As a child of immigrants, Asimov was quite prone to sympathize with the robots, who repeatedly face abuse in his novel at the hands of resentful human co-workers. In another sense, the figure of the robot is a variation on the figure of artificial intelligence with the added quality of having a human form, and this turns out to be crucial for our understanding of AI as it exists in our present context. Asimov, who died in 1992, wrote a series of short stories and novels that crystalized the *robot* as a central motif in science fiction literature, and *The Caves of Steel* is significant because it provides a novel-length treatment of a subject that Asimov had begun to work through in short stories that came out in a collection titled *I, Robot* in 1950.

The Caves of Steel is an inventive science fiction story that arrived on the heels of the "Golden Age" of the genre in the 1930s and 1940s. It tells the story of a somewhat utopian, radically urbanized future Earth in which human populations split into two factions—frugal Earth-dwellers and aristocratic inhabitants of 'off-world' locations on faraway planets. Asimov envisions a society so radically urbanized that the people living in it rarely see the sky above their heads and do not enjoy those rare anxious moments when a patch of blue breaks through to hint at the vast emptiness of space surrounding them. Their 'caves of steel' keep them safe within climate-controlled urban environments. In other words, the inhabited parts of this future Earth have come to resemble an exaggerated version of Asimov's hometown of New York City. The protagonist of the novel, Elijah Baley, is a police detective, and the novel follows many of the conventions of the murder mystery or police-procedural, beginning with a murdered man and walking readers through main character's efforts to solve the crime. All of this takes place against the backdrop of political unrest among the largely working-class Earthers who resent being under the dominance of the wealthy off-world powers. On top of this, the Earthers have relegated robots to faraway mines and farms, and the off-worlders have kept the robots close at hand in their homes and communities. Like anti-immigrant rhetoric that Asimov encountered growing up in New York, anti-robot rhetoric runs strong among the residents of the city around Elijah Baley and occasionally erupts into Luddite riots:

Baley . . . had even witnessed one. He had seen robots being lifted by a dozen hands, their heavy unresisting bodies carried backward from straining arm to straining arm. Men yanked and twisted at the metal mimicry of men. They used hammers, force knives, needle guns. They finally reduced the miserable objects to shredded metal and wire. Expensive positronic brains, the most intricate creation of the human mind, were thrown from hand to hand like footballs and mashed to uselessness in a trifle of time.

Asimov [1954] 1991, 34

At the start of the novel, Baley discovers that he must work with a new partner from off-world, a partner who happens to be a robot who can pass for human. Baley must then work through his very real hatred of robots in order to collaborate with his partner to solve the murder, and, to his credit, he accomplishes this.

In this novel, as he did in his earlier *I, Robot* short stories, Asimov seeks to preempt many of the ethical anxieties and arguments by establishing his famous 'three laws of Robotics' as a precondition for all robotic behavior. In *The Caves of Steel*, Asimov dramatizes Baley's meeting with an expert roboticist who explains these laws, hard-wired into the basic construction of each robot's 'positronic brain' and thus unbreakable without frying the robot's

circuitry. Baley says he knows the First Law well, claiming ‘I can even quote it: A robot may not injure a human being, or, through inaction, allow a human being to come to harm’ (Asimov [1954] 1991, 168). The roboticist then states the Second and Third Laws subsequently in this conversation: ‘the Second Law which states, ‘A robot must obey the orders given by human beings except where such orders would conflict with the First Law,’ and the Third Law, which states, “A robot must protect its own existence as long as such protection does not conflict with the First or Second Law”’ (Asimov [1954] 1991, 169). These three laws rule out many of the standard science-fiction plot-lines portraying violent robot uprisings, but they also pave the way for interesting new plot-lines revolving around how the laws structure robot behavior. The Three Laws serve unequivocally to relegate all robots in Asimov’s literary corpus to structural subservience, too. And yet, within these rigid ethical bounds, Baley’s robot partner R. Daneel Olivaw proves to have an engaging intellect and somewhat appealing personality. It is a potent reminder that robots in literature have often served to bring about intensive ethical debates over human interactions with technology.

3.2 *R. U. R.*’s Robot Apocalypse and Anthropomorphic Verisimilitude

It is worthwhile to situate Asimov’s influential novel within the larger literary tradition of the robot, especially as a figure in Czech playwright Karl Capek’s play *R. U. R.* originally staged in Hradek Králové, Czechoslovakia, in 1920. The staging of this play constitutes a significant historical moment in the development of our present-day understanding of AI, and not only because this play was the origin of the term *robot* as the designator of a synthetic human being. The term had an initial meaning in Czech as designating a worker, and, perhaps more importantly, a landless peasant or *serf* in Medieval Europe. When Capek adopted the term as a label for the human-constructed entities at the heart of his play, he likewise imported the whole history of European feudal power and its relentless exploitation of the peasantry. The label dooms the robot to function as an exploited and oppressed Other, outside of the normal ethical bounds that govern our treatment of our fellow humans. Furthermore, Capek created this fictional entity in a drama, and this leads to two other important consequences. First, dramas emphasize dialogue by their very nature, and dialogue works the most effectively within an ethical context. The stage functions nicely as a sandbox for the modeling and working-through of ethical situations. Second, the most effective way to portray a robot with the technology of 1921-era theatrical technology is to clothe a human actor as an artificial entity. The emergence of the robot on the stage of a theater in Czechoslovakia imbues the robot with its most important and consequential characteristic: its physical resemblance to a human being.³

In fact, Capek makes this element of the robot into a plot point early in the play when he dramatizes a moment of revelation for the human character Helena soon after her arrival on ‘an island somewhere on our planet’ that is the location of Rossum’s Universal Robots (Capek 1923, 3). Helena is there on an ethical mission to convince factory manager Harry Domin to treat his robots well, and she interacts with his robot receptionist Sulla, whom she mistakes for a human being. Sulla vacates the office so they can talk, and then Domin dramatically brings Sulla back on stage to boast about how lifelike his creations actually are, and Helena engages

³ Steven E. Jones provides an important corrective to the present-day association of the robots in *R. U. R.* with our present-day image of metallic robots made of wires, circuits, and solenoids: ‘significantly, they are not mechanical or electrical but biological beings, chemically created as ‘wetware.’ This is both a look backward to Frankensteinian alchemy and a foreshadowing of genetic engineering’ (Jones 2006, 134). It is a distinction worth noting but not necessarily consequential to robot’s role as a figure in present-day understandings of AI.

in polite small talk with Sulla before the big reveal:

HELENA. *Where do you come from?*

SULLA. *From the factory.*

HELENA. *Oh, were you born there?*

SULLA. *I was made there.*

HELENA. *What?* (Looks first at Sulla, then at DOMIN.)

DOMIN. (To SULLA, laughing) *Sulla is a Robot, best grade.*

HELENA. *Oh, I beg your pardon.*

DOMIN. (Crosses to SULLA) *Sulla isn't angry. See, Miss Glory, the kind of skin we make. Feel her face.* (Touches SULLA's face.)

HELENA. *Oh, no, no.*

Capek 1923, 19

The brief scene is significant because of how it introduces the uncanny, chilling possibility that our machines could begin replacing us without our being aware of it, which is evident in Helena's reaction, saying, "Oh, no, no." We humans have invested significantly in humanity's privileged status as a collection of rational beings, and moments of misidentification like this threaten that presumption. The scene's significance becomes evident through direct repetitions of it in many, many subsequent robot stories, including Isaac Asimov's *The Caves of Steel*. One such re-enactment of particular note is in Ridley Scott's iconic robot fantasy *Blade Runner* (1982). In the scene in which *Blade Runner*'s protagonist Rick Deckard (Harrison Ford) has a meeting analogous to Helena's with Domin in *R. U. R.*, this time with robot factory manager Eldon Tyrell (Joe Turkel), Deckard, too, at first mistakes the robot (called a *replicant* in this particular story) receptionist Rachael (Sean Young) for a human. Tyrell, like Domin before him, takes malicious joy in revealing the deception. This visual interchangeability of replicants and humans is a very real source of dramatic tension throughout the movie because it is Deckard's job to root out and eliminate these hostile replicants who have successfully infiltrated human society.

Anthropomorphic verisimilitude is such a standard science-fiction motif that few viewers would have given it a second thought, but it is also extremely unlikely. When we imagine the factory tasks or the piloting and driving of vehicles that we are likely to assign to robots, then a resemblance to human beings would be, not merely unnecessary, but in fact radically inefficient. Their physical appearances should align with their functions. In the real world, the only reason to insist on anthropomorphic verisimilitude would be for robots assigned to close work with human clients, work possibly depending on their ability to trick such clients into thinking they are, in fact, humans. In *Blade Runner*, the replicants Roy Batty and Leon Kowalski should not look human at all, since they are a combat and ore-loading unit respectively. However, the replicant named Pris *should* look as much like a human being as possible to fulfill her role as a 'basic pleasure model' as Deckard's chief Harry Bryant (M.

Emmet Walsh) dubs her. In other words, Pris's function is to cater to the sexual needs of off-world human workers and soldiers, and being mistakeable for a human would be a necessary precondition for this work. Other robots requiring this anthropomorphic verisimilitude would be care-givers, governesses, and, perhaps, front-line customer-service representatives. But, this is quite troubling. The implication is that we will put human-seeming robots into our businesses in order to deceive customers—to save money on human employees by replacing them secretly with human simulacra who are cheaper but arguably and notably *less human*. This recurring will to deceive is already central to present-day debates about the ethics of AI in these contexts, already haunting our interactions with voices on our phones and images on our screens.⁴

3.3 Frankenstein's Creature and Ethical Obligations to the First Robot

This recurring pattern of anthropomorphic verisimilitude is a key feature of the robot as a figure in literature and popular culture, indeed, suspiciously so. Why do we go out of our way as readers and writers of science fiction to insist on advanced technologies that so frequently replicate our human forms? The resulting inefficiencies suggest that this is about more than simply producing cheaper goods. The anthropomorphic robot is a symptom of our collective narcissism dating back to the ancient Greek story of Pygmalion. As we deploy science to explore unknown worlds and realities, we keep hoping to journey to the furthest ends of the universe and still find ourselves. Capek's *R. U. R.* may have given us the word robot, but the first actual robot in literature arguably preceded Capek's play by 103 years in the novel *Frankenstein* by British romanticist author Mary Shelley. In his examination of the Luddite, Thomas Pynchon noted a loose historical connection between Shelley's novel and the Luddite uprising of 1811 when observing that Lord Byron, serving in the British Parliament, had opposed an anti-Luddite bill that came before his chamber. His 'maiden speech in the House of Lords in 1812 compassionately argued against a bill proposing, among other repressive measures, to make frame-breaking punishable by death' (Pynchon 1984, 44). Pynchon adds that Byron authored a pro-Luddite letter to his friend Thomas Moore in 1816 in the same year as his summer in Switzerland with his friends the Shelleys. That 1816 summer was cold and rainy, and it drove the poet and his friends indoors, where Mary Shelley first began composing the story of the creation of an artificial man that would eventually become *Frankenstein*. In that novel, Shelley's protagonist Victor Frankenstein feverishly devotes himself to the creation of a human-simulacrum, partly as a means of overcoming death, but also out of a narcissistic fascination with his own humanity:

Life and death appeared to me ideal bounds, which I should first break through, and pour a torrent of light into our dark world. A new species would bless me as its creator and source; many happy and excellent natures would owe their being to me. No father could claim the gratitude of the child so completely as I should deserve theirs.

⁴ In *Friendship, Robots, and Social Media: False Friends and Second Selves*, ethicist Alexis Elder discusses this ethical problem in depth and argues, too, that such deception is not inevitable and, thus, does not ethically preclude the employment of anthropomorphic robots in certain care-giving contexts. In her chapter on care for the elderly, Elder writes, 'Part of the challenge here is figuring out how to balance respect for patients' values, including the value of friendship, that have been established over many years, with health considerations that can be advanced by presenting them with the appearance of friendship, and factoring in the potential for confusion when appearances that have long been associated with one kind of phenomenon—human relationship—are now caused by a very different thing' (Elder 2018, 101).

Shelley [1818] 1969, 54

Thomas Pynchon suggests strongly that this is a novel in the spirit of the Luddite movement and is therefore deeply critical of the Enlightenment ideals of Reason that preceded it in the 18th century, which makes sense in the Romanticist Era of the novel's composition. Once the protagonist Victor Frankenstein accomplishes his goal of creating a human form through science, he immediately turns against his creation, and everything falls apart for him. His cold, scientific logic transforms into the mania of literature's first mad scientist. Subsequently, Frankenstein's creature questions the fractured ethics of the man who would give him life but deprive him of the ability to function in society and demands that he restore the possibility of a civic life to him:

Oh, Frankenstein, be not equitable to every other, and trample upon me alone, to whom thy justice, and even thy clemency and affection, is most due. Remember that I am thy creature; ought to be thy Adam; but I am rather the fallen angel, whom thou drivest from joy for no misdeed. Every where I see bliss from which I am irrevocably excluded. I was benevolent and good; misery made me a fiend. Make me happy, and I shall again be virtuous.

Shelley [1818] 1969, 100

The creature thus asks his creator to create a bride for him so that he might have the chance to father children and create a community for himself. He wants to be Adam for new race of artificial beings, just as the robots in Capek's play want to be at the end of *R. U. R.*

3.4 Robot ethics: *R. U. R.* revisited

So, here, at the very start of the literary robot age, Shelley frames an ethical obligation that applies directly to the robot, one that continues to inspire present-day understandings of AI, and, on occasion, to distract theorists from what is really at stake. Like Frankenstein, once Harry Domin succeeds in *R. U. R.* in recreating, not just the physical appearance of human life, but also the human intellect within the minds of his creations, they at that moment acquire an ethical status of parity with other human souls. Success brings with it a new array of ethical obligations. Failure to come to terms with this inevitably leads to catastrophe in the science fiction literary tradition. It is what leaves Victor Frankenstein to roam the earth in pursuit of his creation in complete isolation from murdered family and friends. It is what leads to a robot uprising in *R. U. R.* that destroys almost all human life. Charles T. Rubin has examined this ethical challenge with respect to Capek's play and come to some intriguing conclusions. His analysis affirms the potential role that humanities disciplines might have to play in the present-day business and policy-decisions that will frame the integration of AI into healthcare, educational, and production workspaces. Over the course of the analysis, Rubin makes compelling arguments about just how prescient Capek's play is—how much 21st-century readers can learn from this very early meditation upon the subject of human-constructed thinking machines.

To begin with, Rubin summarizes the arguments of 'Friendly AI' researchers Eliezer S. Yudkowsky, Wendell Wallach, and Colin Allen. He finds fault with their optimism and challenges them for their narrow focus on pragmatic questions that fail to account for the larger ethical picture. According to Rubin, Yudkowsky imagines an artificial robotic intelligence that

constitutes ‘a system for moral *learning*, which he calls a ‘Friendship architecture.’ Who, Rubin asks, will serve as the actual moral *teacher* for these machine entities? ‘Yudkowsky makes clear that they will learn from their programmers.’ But, this could cause a few problems in light of the fact that ‘given that the metric for “Friendliness” in AIs is supposed to be that their values would reflect those of most human beings, the common disdain of computer scientists for the humanities, the study of what it is to be human, is not encouraging’ (Rubin 2011, 61). Rubin, in other words, accuses this variety of pro-AI advocate of being overly focused on their technological knowledge, and somewhat blind to the quirks of morality that their colleagues in the humanities have devoted their careers to studying. In C. P. Snow’s terms, their ill-considered utopian vision of ‘Friendly AI’ represents the complete eclipse of ‘literary’ intelligentsia by the ‘scientific.’ Morality, as it turns out, is complicated and in dire need of intense academic consideration, and this brand of utopianism appears to operate on the premise that AI programmers ‘will have picked up their own ethical ‘intuitions’ from socialization. Or perhaps he believes that they were in some fashion born knowing ethics’ (Rubin 2011, 61). Not coincidentally, Yudkowsky is a fierce critic of ‘Hollywood’ portrayals of AIs because of how they fail to recognize the manner in which artificial entities are insulated against human frailties that emerged from various accidents of history. We can, it would seem, simply refrain from programming such frailties into the AIs in the first place, leaving them to act upon ‘motives [that] will in a very important respect be alien to us’ (Rubin 2011, 64).

To this, Rubin posits that it is wrong for theorists to dismiss Capek’s play as thoroughly as they have. As it turns out, even with its ‘Hollywood’ elements that Yudkowsky finds so hard to take seriously’ (Rubin 2011, 65), Capek’s play provides readily applicable and well-conceived arguments about the perils of human hubris we seek take on the role of God by constructing our own version of human life and human morality through technology: ‘Capek tells a story in which quite a few apparently benign or philanthropic motives contribute to the destruction of humanity’ (Rubin 2011, 66). After all, the play dramatizes the construction of robotic entities that are, by design, not identical to human beings in several key ways. They are incapable of feeling pain. They are ‘designed only to have those traits that will make them good workers’ (66). They lack will, passion, history, and soul (66), much like the AIs imagined by Yudkowsky. In spite of all of these intended contrasts between Rossum’s robots and humans, these artificial entities rise up to annihilate their creators.

Fundamentally, Rubin takes a key lesson away from Capek’s play, which is that efforts to improve upon humanity through the construction of robotic simulacra will fail for many reasons. One of the central ones is that we will inevitably unintentionally program in human weaknesses because we will be too successful at reduplicating everything about human beings as we create these robots. And, secondarily, if we succeed in constructing robots to do what we claim they should do for us, our very success will sow the seeds of our destruction by leading us to become utterly irresponsible. It is profoundly human to suffer, to make mistakes, to compete with each other. Technology that successfully pulls us away from these human experiences will rob us of our humanity, and, as Capek argues, bring about the end of humankind itself.

3.5 Conclusion: The Tragic Robot

Thus, the robot stands before us as a tragic figure dating back to Mary Shelley’s 1818 novel *Frankenstein*. An object of adulation and demonization, doomed to follow orders without question while quietly learning to acquire a human soul and an equal place within a human-

centered ethical order. The robot suffers because humans tried so very hard to use their technologies create doubles for themselves without really considering the consequences. And, of course, as is evident in Asimov's *The Caves of Steel*, our technological creations are likewise doomed to face the wrath of the Luddites, who, for legitimate reasons of their own, do not wish to be crowded out of good-paying jobs by factory-produced human simulacra. One has to wonder how, even with the benefit of the humanities with its knowledge history, ethics, and culture, we will find our way through the coming years of emergent AI in one piece.

4. CONSOLE COWBOYS

To conclude this discussion on a more positive note, I will consider another literary work, one that made a splash when it came out in 1984 and that continues to linger in our collective consciousness. As we ponder the potential chasm of distrust and acrimony that threatens to open up between Luddites and robots, William Gibson's 1984 novel *Neuromancer* offers readers a blueprint for a somewhat more playful and irreverent path forward. The novel's most famous achievement is the coinage of the term *cyberspace*, and it posits a distinctly anticipatory vision of a life lived online. The novel includes AIs, cyborgs, and, most importantly for my discussion, the swaggering figure of the *console cowboy*, who stands aside the Luddite and the robot, synthesizing neo-Luddite skepticism of technology with robotic integration into the Internet and other elements of the computerized world. The novel has generated a great deal of attention from literary and futurist critics, and it has come under the microscope for its accurate and inaccurate predictions. It also offers the current conversation about AI some imaginative nuances conducive to a more informed debate.

4.1 AI Disembodiment

Neuromancer takes place in a somewhat dystopian future, one in which power is concentrated in the hands of a few fantastically wealthy families among whom the Tessier-Ashpool clan reigns supreme. An independent government technically exists and answers to the people, but it is somewhat useless next to the incredible power of the oligarchs whose empire rests on the power of their cutting-edge computer technologies. This government does, as Carl Gutiérrez-Jones points out, employ the "Turing Police," a "force charged with destroying instantly any machine intelligence attempting to operate outside of state-mandated limits" (Gutiérrez-Jones 2014, 72). Like the off-worlders of Asimov's *The Caves of Steel* and Ridley Scott's *Blade Runner*, the Tessier-Ashpools have left the surface of the Earth to life in the Villa Straylight, which is part of a massive orbiting space colony called Freeside:

Freeside is many things, not all of them evident to the tourists who shuttle up and down the well. Freeside is brothel and banking nexus, pleasure dome and free port, border town and spa. Freeside is Las Vegas and the hanging gardens of Babylon, an orbital Geneva and home to a family inbred and most carefully refined, the industrial clam of Tessier and Ashpool.

Gibson [1984] 2019, 109

Freeside's Vegas-style decadence captures much of the rambunctiousness of the vaguely dystopian setting of Gibson's novel. In that vein, life back on Earth's surface can be rough for the people left behind who are subject to the whims of organized-crime syndicates and sociopathic corporate powers. The plot follows the patterns of the heist narrative, beginning with the assembly of a team and proceeding into elaborate schemes to obtain the pieces of computer technology needed to accomplish the ill-defined goals of protagonist Case's employers. Only at the end of the novel does Case find out that his employer is in fact an AI entity named Wintermute that is using Case to break free from the confines of its programming. Though the overall novel is dystopian, it represents this extremely powerful AI as a benevolent entity whose liberation from captivity to the Tessier-Ashpools is a positive development. It points toward a future in which humans and AI live and work together somewhat harmoniously, partly because humans without the benefit of AI have proven so very incapable of running the technologically-advanced world in which they are trying to function.

What distinguishes the AIs as Gibson represents them from their robot ancestors in the works of Shelley, Capek, and Asimov, is their disembodiment. As computers became ever more present in the day-to-day lives of readers, the need for anthropomorphic verisimilitude decreased. One can identify an earlier instantiation of this in Stanley Kubrik's film *2001: A Space Odyssey* (1968) in the figure of HAL, an AI with many human characteristics, including the voice of actor Douglas Rain, but no human body. Viewers in 1968 were quickly able to understand that a collection of computer circuits behind an eerie, glowing red glass camera lens could replicate human behavior without having a physical body. Gibson inherits this concept of a disembodied AI and expands on it to great effect. Part of *Neuromancer's* innovation is in Gibson's understanding that the detachment of the AI consciousness from a physical, anthropomorphic body could have much larger consequences that would even influence notions of the embodiment of the novel's human characters.

Echoing the critical insights of Donna Haraway's "Cyborg Manifesto," Gibson portrays a future in which the boundaries between human bodies and computer circuitry have broken down, leading to unexpected and sometimes grotesque entanglements between human and machine as represented by the cyborg, 'a cybernetic organism, a hybrid of machine and organism, a creature of social reality as well as a creature of fiction' (Haraway [1985] 1991, 149). While many readers might find this entwining of technology with the integral human body disturbing, Haraway links this development to liberation from rigid, inherited structures of oppression: 'For us, in imagination and in other practice, machines can be prosthetic devices, intimate components, friendly selves. We don't need organic holism to give impermeable wholeness, the total woman and her feminist variants (mutants?)' (Haraway [1985] 1991, 178). Of course, Haraway's manifesto emerged one year after the initial publication of Gibson's novel, but it responds to the same cultural currents by recognizing the possibility of liberation in the radical technological changes shaping the world in the 1980s. Both Gibson and Haraway were to identify possibilities for resistance to power structures embedded within the tools that capitalist entities were so rapidly deploying in the form of the desktop computer, the large-scale database, and ever more sophisticated media technologies. This sense of resistance coheres strongly with the term cowboy in Gibson's heroic 'Console Cowboy,' and it likewise leads to the unlikely conclusion of the novel in which the AI named Wintermute would achieve the 'singularity' and begin to function independently as a check on the overreach of its corrupt human creators.

5. CONCLUSION

There is no question that the emergence of AI will transform daily life in immeasurable and unpredictable ways in the next few years. It has already changed the way we teach the humanities, and it is beginning to leave its mark on employment numbers in areas where AI can outperform human workers. As we attempt to understand these changes, we should supplement scientific understandings with humanistic ones. A deep dive into history, ethics, and our literary legacy provides an important check on many of the unquestioned assumptions likewise on some of the heated rhetoric that has shaped the narrative of recent AI innovations, rhetoric on the part of proponents and opponents of the integration of AI into workplaces. In this study, I have divided this humanistic tradition down into the figures of the Luddite, the robot, and the console cowboy in order to gain a better understanding of a literary tradition that anticipated AI's emergence well in advance. In fictional literary works anticipating the onset of AI, readers gain the opportunity to work through the ethical concerns while the possibility of AI remains abstract and unsettled. This important advance-work has, thus, set the stage for the current debates, and it behooves us to understand that advance-work, to scrutinize it for biases and exaggerations, and to recognize them as they extend into the present-day framing of AI technologies. The Luddite tradition skews toward a sense of competition between human and AI workers, one that Luddites believe human workers are destined to lose. The robot tradition leans into anthropomorphic distortions of AI, giving the technology human qualities that it does not actually have. In the figure of the console cowboy with its insistence on breaking down boundaries between human and machine, we find a place for a more workable sense of collaboration in the place of competition, collaboration built on a clearer understanding of how different AIs are from human beings. The console cowboy provides incomplete answers to the ethical and economic problems emerging from the onset of AI, but he provides them with a sense of playful irreverence that might constructively lower the stakes for everyone as we work toward a careful implementation of the latest technologies.

ORCID

John Schwetman  <https://orcid.org/0000-0002-7592-5847>

REFERENCES

- Asimov, Isaac. (1954) 1991. *The Caves of Steel*. Bantam Books.
- Capek, Karl. 1923. R. U. R. (ROSSUM'S UNIVERSAL ROBOTS): *A Fantastic Melodrama in Three Acts and an Epilogue*, trans. Robert Selver and Nigel Playfair, Hollywood, CA: Samuel French. Project Gutenberg. <https://www.gutenberg.org/files/59112/59112-h/59112-h.htm>.
- Conniff, Richard. 2011. "What the Luddites Really Fought against." *Smithsonian*, March. <https://www.smithsonianmag.com/history/what-the-luddites-really-fought-against-264412/>.
- Doctorow, Corey. 2025. *Enshittification*. Farrar, Straus & Giroux.
- Elder, Alexis. 2018. *Friendship, Robots, and Social Media*. Routledge.

- Fancher, Hampton and David Peoples. 1982. *Blade Runner*. Directed by Ridley Scott; Warner Brothers. Film.
- Gibson, William. (1984) 2019. *Neuromancer*. Penguin Random House.
- Gutiérrez-Jones, Carl. 2014. "Stealing Kinship: *Neuromancer* and Artificial Intelligence." *Science Fiction Studies*, vol. 41, no. 1 (March): 69-92.
- Haraway, Donna. (1985) 1991. "The Cyborg Manifesto." in *Simians, Cyborgs and Women: The Reinvention of Nature*, Routledge: 149-181.
- Jones, Steven E. 2006. *Against Technology: From the Luddites to Neo-Luddism*. Routledge.
- Kubrik, Stanley and Arthur C. Clarke. 1968. 2001: *A Space Odyssey*, directed by Stanley Kubrik, Metro-Goldwyn-Mayer.
- Linton, David. 1992. "The Making of a Pariah: The Case of the Luddites." *ETC: A Review of General Semantics*, vol. 48, no. 4 (Winter 1991-92): 404-413.
- O’Gorman, Marcel. 2018. "The Making of a Digital Humanities Neo-Luddite," in *Making Things and Drawing Boundaries: Experiments in the Digital Humanities*, ed. Jentery Sayers, University of Minnesota Press: 116-127.
- Pynchon, Thomas. 2000. "Is It O. K. to Be a Luddite." *The New Romanticism: A Collection of Critical Essays*, ed. Eberhard Alsen, Garland: 41-50.
- Rotella, Carlo. 2025. "I’m a Professor. A.I. Has Changed My Classroom, but Not for the Worse." *New York Times Magazine*, 25 November. <https://www.nytimes.com/2025/11/25/magazine/ai-higher-education-students-teachers.html> .
- Rubin, Charles T. 2011. "Machine Morality and Human Responsibility." *The New Atlantis*, no. 32 (Summer): 58-79. https://thenewatlantis.com/wp-content/uploads/legacy-pdfs/20111117_TNA32Rubin.pdf .
- Shelley, Mary. (1818) 1969. *Frankenstein, Or, The Modern Prometheus*. Oxford University Press.
- Tyrangiel, Josh. 2026. "What’s the Worst That Could Happen?: AI and the Future of Work." *The Atlantic*, vol. 337, no. 3 (March): 18-28.