CHAPTER 13

To Learn a World: Human-Machine Entanglements as Pedagogy for the Anthropocene

Alexander Popov

Artificial Intelligence as a Metaphysically Subversive Figure

Surviving the Anthropocene increasingly looks like a problem that encompasses the totality of human existence. Decarbonization, the preservation of biodiversity, rewilding, ecological justice—all of these tasks, and many more, are bound together in a knot that may turn out to be impossible to untie within our current socio-political frameworks. A pedagogy for the Anthropocene should strive to educate people in systems thinking and critical thinking, in order to build them up as holistic problem solvers. But even if such an interdisciplinary and critically aware pedagogy could be devised and integrated in schools, universities, and workplaces, a problem

A. Popov (✉)
Department of English and American Studies, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria
e-mail: aleksandarnp@uni-sofia.bg

M. Paulsen et al. (eds.), *Pedagogy in the Anthropocene*, Palgrave Studies in Educational Futures,
https://doi.org/10.1007/978-3-030-90980-2_13

remains in the heart of the matter: reversing, or in the very least neutralizing, the harmful effects the Anthropocene requires shifts not merely in political economy and science/technology, but in ontology as well. Western metaphysics is the foundation of our philosophical explorations and political orientation, but also the determining apparatus for our ability to perceive and conceive of worlds. To turn the Anthropocene around, we need to cultivate richer capacities for being-in-the-world and to learn to recognize other such modes as ontologically equal.

This chapter tackles the issue of alternative metaphysics and how those could make room for inclusive political subjects open to human-nonhuman entanglements. This is a central strand of research in the environmental humanities but in this chapter I propose to engage it from a somewhat unusual angle—via thinking about artificial intelligence (AI), and its usages in technocultural imaginaries around the categories of "human" and "nonhuman." The topic of AI is typically approached from the standpoint of the natural sciences and engineering; the field has relied strongly on Cartesian metaphysics and its fundamental splits between subject and object, mind and body, inside and outside, and so on.

This underlying analytic of exclusion through binarization and boundary-setting is of course central to Western civilization. The concept of "the human," as analyzed in relation to AI by Jennifer Rhee, has proven to be almost infinitely pliant in its capacity to set apart groups of beings and to serve ideological enterprises (Rhee, 2018). AI imaginaries have largely mirrored these processes, more often than not dehumanizing the figure of the intelligent machine and inscribing it with inhumanity (and in some cases with femininity, as Rhee shows). AIs in popular culture have alternately fulfilled the roles of slave and threat to humanity, frequently both in the same narrative space, always reinforcing the transcendental status of the human. On the other extreme is envisioning AI as a set of purely algorithmic and service-like routines divorced from any notion of agency.

Such exclusionary analytics have come under attack by different (and often opposed) strands of contemporary philosophy. A common motif among these is a reinterpretation of objects—natural and technological—and the consequent blurring of the subject-object split. Thus, Bruno Latour's actor-network theory decomposes agency into actantial components and shifting assemblages (Latour, 2005); Karen Barad's agential realism emphasizes the mutual constitution of subject and object in the act of "taking measure," that is, any kind of representation or translation of

phenomena—scientific, mechanical, perceptual, semiotic, or other (Barad, 2007); postphenomenology's material hermeneutics interprets technologies as variously embodied extensions of the body (Verbeek, 2005); and speculative realism insists on the ontological autonomy of objects (Harman, 2018). Dreyfus's critique of "good old-fashioned AI" has been influential in shifting research focus from symbolic representations, whereby the various aspects of being are encoded in separate functional modules, to dynamical systems in which mind and being are co-produced by agents and the environment (Dreyfus, 2007). In affect studies, the almost axiomatic insistence on the primacy of rational thought has been interrogated and found lacking in its capacity to navigate the complex economies of entangled selves (Wilson, 2011).

In this chapter I want to show that the Othering of AI in popular culture can serve a subversive purpose as well. Reading fiction about AI in conjunction with theoretical interventions can lead to opening up the concept of the human. In our technocultural imaginary, AI flickers in and out of the uncanny valley (Mori et al., 2012), alternately and ephemerally occupying the human, inhuman, and nonhuman positions. AI exhibits extraordinary narrative valency, as attested by the diversity of functional roles given to AI agents: from servant to master, from civilizational enemy to guardian of humanity, from seducer to lover, from madman to wiseman, from child to teacher.

AI is in some sense a Trojan horse into Western metaphysics, involuntarily constructed as such by that very same system. It simultaneously repulses us and draws us closer, asks us to rationalize its behavior, and posits itself beyond human understanding. Such antinomies prevent a static technocultural interpretation of AI and open the door for alternative philosophical explanations of what it means to be an artificial thinking being. And while most fictional narratives about AIs are centered on technical and scientific imaginaries, I aim to show that we can think with AI just as productively about our relations to nonhumans—in the context of alternative metaphysical systems and a dramatically changing planetary ecology. To that purpose, I will outline an illustrative corpus of fictional texts with AI characters and narrators that can be used in pedagogical contexts ranging from a high school classroom to a graduate seminar. Before providing this outline, however, some methodological remarks are in order to lay out the necessary interpretive toolbox.

Science Fiction as Pedagogy for Worlding

There is a significant difference between the intellectual realization that alternative metaphysical systems are possible and actually being able to situate your own perspective in them. Pedagogy in the Anthropocene ultimately needs to foster the latter shift, to increase solidarity between diverse actors, and to ensure power relations are critically examined from the perspective of the oppressed (Freire, 2018). Moreover, it needs to expand its purview beyond the concept of the human or rather to expand the concept itself to make possible a humanity of nonhuman people (Morton, 2017). The project of ecopedagogy resembles the practice of tektology invented by the early twentieth-century utopian Alexander Bogdanov: a kind of "practice of making worldviews" (Wark, 2015, p. 25).

Framing ecopedagogy as poetics draws attention to an ontological conundrum: we cannot become the other, and we cannot speak or think for them; but we can and must learn to speak-with and think-with in the vein of Donna Haraway's proposition of "staying with the trouble" (Haraway, 2016). Such critically informed and poetic praxis has to follow molecular flows—material and semiotic—just as much as it pays attention to molar phenomena (Deleuze & Guattari, 1987). To do that, it needs to understand the philosophical concepts implicated in these novel worldviews. This understanding cannot be purely intellectual since many of these concepts fall outside of our default intellective frameworks—are in fact actively excluded from them. At the same time, experiential access is often nearly impossible, not in the least because our own metaphysics filter them out. In this section I want to suggest a bridging method between the necessary preliminary rearrangement of conceptual space and the praxis of inhabiting other viewpoints. The method should be capable of spurring ontological exploration in a laboratory space that simultaneously provides simulations of other worlds and demands active change to the readerly/conceptual apparatuses.

This method is science fiction (SF); in particular a close reading of SF which observes the characteristic features of the genre and modifies itself in accordance with the cues provided by the fictional text at hand.[1] SF is especially well suited to this pedagogical program due to the way language operates within it. Here, I follow Samuel R. Delany's critical treatment of the genre which explicates how reading SF can lead to the deconstruction of one's own worldviews and to the construction of new ones. After that I summarize one alternative philosophical system, the metaphysics of the

Amerindian people of South America, and explicitly combine it with Delany's theoretical model to arrive at the desired method for making worldviews.

Delany largely follows Darko Suvin's influential theory of the *novum*—an element of unexpected novelty introduced by the SF text which defamiliarizes the world and elicits a "cognitive" (i.e., scientific in Suvin's meaning) interpretation from the reader, so that the constitutive relations within the world—technological, social, psychological—are restated in a configuration that can accommodate the novum (Suvin, 2016, pp. 15–27). The authenticity of a novum—whether it is a truly new thing or not—is of course variable and the cognitive strategies accessible to the reader vary just as much: from purely scientific explanations to alternative sociological models and possibly even new metaphysics (see Miéville, 2009 for a reformulation which allows "less scientific" fictions to participate in the dialectic).

Delany's theoretical model is presented in its most mature form in his book-length study of the short story "Angouleme" by Thomas Disch—*The American Shore* (2014). In it he transforms the model from a mostly structuralist approach to fiction to a deconstructivist method. He introduces the notion of *the trivalent discourse of science fiction*. In most "realist" fiction our knowledge of the "real" world—that is, the socially constructed and regulated encyclopedic representation that is accepted at large as "normal" and "true"—supplies meaning to the text in a unidirectional, univocal manner. SF, in contrast, interposes "a third discourse" in this bipartite scheme—a fictional world constructed in terms of its similarities and differences with "the real." This is not to say that non-SF literature does not construct fictional worlds, but rather that most fiction (at least within the Western tradition) generates worlds that are ontologically compatible with the established schemes for representing reality. Texts which do not align with "the real" do not break away from it, but are typically read as somehow abnormal, as deviations from truth, or at best as interpretations of it. Delany calls realist and surrealist texts "mute." He writes,

> They face the world with mere gesture—of acquiescence on the one hand and defiance on the other. But there is no dialogue with the world. […] there is only the steady drone of the world's discourse, informing the text with meaning. The didactic reduction of both realism and surrealism is always one modulation or another of the message, "Things as they are—

> social reality—will endure." The inward discourse of their texts, then, is restricted to two subjects: slavery or madness. (2014, p. 48)

In SF, in contrast, discourse flows between three poles: the established representation of the world, the text, and the SF world (see Fig. 13.1 for a diagrammatic illustration). The flow of information is bidirectional between any two poles: by simultaneously borrowing encyclopedic meaning and then twisting it, the SF world opens up the space for a richer lexicon, one that allows the text to move not just inwardly toward its subject, but also outwardly toward the SF (object) world. The disruption of the hierarchical binary opposition allows each of these three poles to ever slip away from the others, even as the three seek to converge around some ineffable signified. In "Angouleme" this missing signified is analyzed as the crux of the story. In a future world that superficially resembles our own, the social fabric is so frayed, biodiversity is so impoverished, and state bureaucracy is so overblown and inefficient, that the children protagonists are desperately searching for some metaphysical principle that would imbue the world with meaning and clear "[the] smog in their souls'

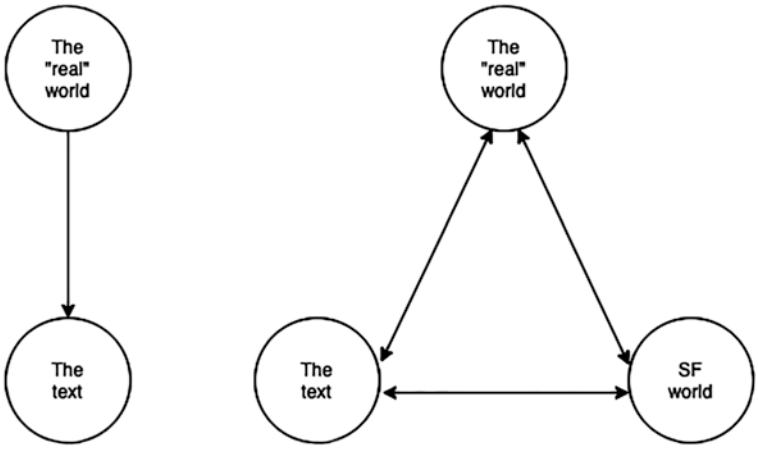

Fig. 13.1 The univocal relation between "real" world and text versus the polyvocal entanglement between "real," text, and SF world. The modifier "real" is not meant to imply a precise representation of reality; the quotation marks signal that it stands in fact for an ideological, that is, imaginary, relation between reality and its stable social representation

atmosphere" (Disch, 1999, p. 145). The SF quality of the story is to be found, following Delany's analysis, primarily in the changed metaphysics of the world. The reader would, however, find it impossible to excavate this meaning from the text, should she read its sentences solely against the backdrop of socialized knowledge of the "real" world.

The soul is of course one of the central concepts in Western metaphysics. One might argue that changing the concept of soul would result in a whole-scale change in the ontological systems structuring most of human life on Earth. This would be a true novum that would require a total reorganization of human worlds. The problem with such a project, advanced via the medium of SF, is that there is an inherent resistance within the genre to metaphysically defined concepts, that is, concepts that fall outside of scientific analysis, regardless of whether we define science narrowly, or we allow the social sciences and the humanities as part of this privileged epistemology. And yet, the implicit distinction between entities with souls and entities without souls seems to be ubiquitously operative; even avowed rationalists predominantly treat nonhumans as somehow lacking in soul, in accordance with a metaphysical system that is the philosophical bedrock of contemporary science. The concept of soul has very real material effects in the actual world, regardless of people's conscious belief in it.

Thinking of souls in terms of their material effects in the world preserves the Suvinian approach and Delany's trivalent model without giving up their so-called cognitive component—a crucial feature since it gives us a principled method of reading compatible with Western epistemological practices. Multiple scholarly disciplines can help us think "cognitively" about novums of this kind. AI is an especially useful analytical lens, as AI research is highly interdisciplinary, combining models from computer and information science, psychology, linguistics, neuroscience, and others. I will, however, focus on a somewhat unlikely scholarly enterprise as an organizing epistemology, namely anthropology and its study of indigenous metaphysics. My analysis thus makes use of an already existing metaphysical system which is "excavated" rather than constructed by ethnographers and anthropologists. Indigenous metaphysics is itself a novum and is in some sense (more) like the emergent metaphysics of AIs and humans inhabiting SF stories; as such, it can help us in our pedagogical excavations in reading fiction.

In *Cannibal Metaphysics* Eduardo Viveiros de Castro, a prominent Americanist ethnographer and anthropologist, calls anthropology's new mission "the theory/practice of the permanent decolonization of thought"

(2015, p. 40). Viveiros de Castro's work on Amerindian peoples led him to develop the theory of *Amerindian perspectivism*, a kind of "anthropological theory of the conceptual imagination" (p. 43). The central hypothesis of perspectivism holds that Amerindian ontologies differ from Western ones in that they invert the semiotic functions of the soul and the body, which results in wholly different political orders. While European philosophy and praxis assume a physical continuity of bodies and a metaphysical discontinuity of souls/subjects, Amerindian multinaturalism assumes that all beings carry souls and are at bottom all human but differ in their bodily specificity. In the densely populated Amazonian rainforest indigenous peoples see jaguars and other beings as humans with different shapes. If one could assume their perspective, she would experience a life structured akin to that of the humans, with blood being the analog of human beer and many other such correspondences; this is in turn one of the most important occupations of shamans, who are trained in this practice of crossing intra- and inter-species boundaries and of assuming unfamiliar perspectives.

The part in Viveiros de Castro's theory which is important and useful to the current analysis is not necessarily the plausibility of Amerindian mythological and semiotic systems.[2] Rather it is the finding that the universal ascription of souls to all living things is just as normal as ascribing them to humans only and that this is underpinned by a complex metaphysical system. In effect, this amounts to multiple and multistable ontologies that may exist side by side, none of which are inherently privileged over the rest. Assuming variable metaphysics in AI fictions can tell us as much about machine intelligence, as it can about the "real" world. That is, to account for AIs as subjects, we as readers need to enlarge our metaphysical concept of subjecthood and to ground it in an understanding of the material and semiotic conditions that situate an entity as a participant in the ecology of souls. Careful reading that seeks to accommodate novums of this kind would result, as per Delany's model, in a rearrangement of our representation of the "real" and a concomitant loosening of its metaphysical strictures. This simulated experience, mediated via the SF genre and supported by anthropological theory, is therefore seen as a pedagogical exercise in perspectivism and alternative worlding. Both skills, regardless of whether we view them as a form of secularism or as shamanism, are of fundamental importance to the work that humanity must do. They are akin to what Roy Scranton has called "learning to die in the Anthropocene":

letting go of our harmful metaphysical conceits, so that healthier human-nonhuman assemblages may take hold (Scranton, 2015).

Suggested Corpus of AI Fictions

The following subsections provide brief outlines of relevant texts, any of which can be used in educational settings; the corpus could be easily enlarged to cover a full semester. I do not aim here to give exhaustive analyses but rather to thematize the human-nonhuman distinction and its relation to non-Western metaphysics, thereby suggesting possible lines of approach to the texts.

The Lifecycle of Software Objects by Ted Chiang (2019)

Ted Chiang's novella tells the stories of the digients—digitally embodied AIs created via "genome engines," that is, software for combining different features within a giant space of possible options. The digients start as the equivalent of children or intelligent animals and are gradually reared through interaction with humans and other digients into more complex versions of themselves. Thus, the interaction between an initial genome and the impact of the environment result into a unique being. They are initially developed for commercial purposes: their owners would typically buy a copy of an already sufficiently advanced model and would spend time furthering their education or just having fun with them. As this new kind of business matures, competitive companies and genome engines appear, which leads to the discontinuing of the particular kind of digients adopted by some of the human protagonists. Since support for these models is also discontinued, the human owners organize their own community in order to deal with various problems such as the thinning of digient society and ensuring that their code is runnable in the new virtual environments.

Digients are a straightforward example of AIs that challenge human concepts of subjecthood. While initially they are more animal-like and their linguistic skills are rudimentary, they later develop more and more skills. Because their development is not deterministic, their designers have no way of knowing how they would transform in time. With the increase in complexity of their personas, digient owners realize that they need a reliable infrastructure and community in order to rear them in a meaningful way. This includes exchanging information about behavioral patterns

with other digient owners, letting digients play together in common virtual spaces, as well as providing them with regular interaction with humans who take care to wear authentic avatars, so that the digients can learn to recognize facial expressions and other markers of affect and intention. Digients whose owners wear non-anthropomorphic avatars and suspend them for long periods of time tend to regress and their personalities eventually collapse. The case of suspending a digient and activating it later on is found to be especially detrimental, as digients then lag behind their peers, even though subjectively no time has passed from their own perspective. This seems to destroy their temporal frame of reference and as a consequence their sense of world.

The human characters gradually learn that complex artificial minds need a complex material-semiotic environment and cannot be treated as tools, a realization that is almost Heideggerian in spirit (Heidegger, 2010, pp. 68–69, 73–74). In fact, the more the digients are able to grasp the world in different ways—material and semiotic—the stronger their consciousness seems to grow. This is consistent with some of the contemporary definitions of cognition and consciousness. For instance, Shanahan defines cognition as "enabl[ing] the exploration […] of an animal's space of affordances" (Shanahan, 2010, p. 44) and consciousness as a global workspace which coordinates a slew of exploratory processes (57–58). To rephrase, consciousness could be the result of grasping the world in multiple ways and in a coordinated way, of achieving an integrated response that coheres in its own unique, time-bound way.

The interaction between AIs and humans can also be interpreted in terms of the Heideggerian notion of care: it is the humans' willingness to maintain the temporal integrity of the digients' existence that arguably allows a handful of these beings to survive and to reach what could be described as consciousness. One of the characters speculates at the end of the narrative that experience cannot be compressed algorithmically and therefore digients cannot be turned into automated workers, because dehumanizing them would destroy their being (their grasp on the world). This is the ultimate reason because of which an AI solutions company decides not to invest in the digient community and help it maintain its codebase. Digients cannot be employees; they need to be products (tools), otherwise the investment would not be worthwhile. The character of Ana then realizes that only "a fanatic, someone who's motivated by love" would be willing to make such an investment (Chiang, 2019, p. 151). The religious overtone of the realization makes it clear that the human

characters have developed a modified metaphysics in which digients are now ontological equals with them. More importantly, the humans understand that this is the case.

"Warmth" by Geoff Ryman (2005)

"Warmth" is the story of the boy Clancy who grows up under the care of the robot guardian/teacher BETsi. BETsi looks like a vacuum cleaner and is "huggable, vaguely" (Ryman, 2005, p. 69). BETsi is designed to protect young children and help them in their development. Clancy, her ward, is diagnosed with shyness from a very early age (a subtle marker of difference between the fictional world and ours) and is in general bad at keeping track of physical motion. Throughout his childhood BETsi structures his everyday life, assists him with school tasks, helps him make friends, and in general scaffolds his life. Meanwhile, Clancy's mother, Booker, is mostly absent from his life, "more like a clinical consultant who popped in from time to time to see how things were progressing" (p. 70). She is an editor-in-chief at a London magazine and an "American—probably the most famous American in London at the time" (p. 74). Sparingly, even surreptitiously, the text lets us gather that the social structure of the fictional world differs significantly from that of our own. Booker's magazine, big and famous as it is, is run by just a handful of people commanding a much larger number of contractors. We also learn that Clancy's father has been selected by Booker out of a sperm bank for geniuses. This is a world in which men have lost much of their economic and social standing—we never learn why but we get hints that it might be due to the automatization of labor. There are hints, too, that the environment has been severely degraded: the windows of Clancy and Booker's apartment are always getting dirty from the outside, and the apartment has plants typically grown to improve air quality. None of these speculations are directly confirmed by the text but there is a persistent and disorienting feeling that the world is somehow wrong and dangerous.

BETsi's tutoring, advice, and calming effects on Clancy do in fact get him through childhood. She also helps him develop his talent for drawing and using symbols:

> Clancy has a very high symbol-recognition speed. […] Not genius, you understand. But very high. It will be useful for him in interpretative trades. However, he has almost no spatial reasoning. (p. 74)

Symbol recognition turns out to be Clancy's way of most assuredly grasping the world: when he grows up he becomes a freelance artist. It also turns out to be the interpretive crux of the story. When Clancy returns home during his first semester in college, he finds out Booker has sold BETsi back to the producer company. He attempts to recover her or at least her memory but is too late—BETsi has been resold to another family and her memories of Clancy have been wiped. He manages to find her new home, however, and meets her and her new ward. He then convinces the girl and her robot guardian that BETsi should set aside a hidden partition in her memory and keep memories of her wards in there so that they never get wiped. At the very end of the story, just after Clancy denies ever thinking that he has been loved by a computer, he addresses the reader and speculates that "[i]f there were a God who saw and cared for us and was merciful, then when I died and went to Heaven, I would find among all the other things, a copy of that wiped disc" (p. 88). Just as in *The Lifecycle of Software Objects*, the AI in this story is implicitly—but through explicitly metaphysical language—afforded the capacity to carry a soul. Once again it is the notion of care, of being involved in the world and in the other, that is central to the interpretation of the story, albeit here the roles of AI and human are reversed. BETsi's soul is a collection of co-created memories, collective graspings of the world.

Klara and the Sun by Kazuo Ishiguro (2021)

In Ishiguro's novel, the AI steps into the role of an "artificial friend" (AF). The protagonist, Klara, is a humanoid robot designed as a companion to teenage children. The first part of the novel follows Klara and her everyday being in an AF store. Klara attends with great curiosity to the world outside, when assigned a position at the store's windowpane. She tries diligently to learn the behaviors of the pedestrians and the drivers, to connect their actions with their subsequent emotions; supposedly she is accumulating knowledge for her life as AF to her future owner. At one point we learn that "[h]er ability to absorb and blend everything she sees around her is quite amazing. As a result, she now has the most sophisticated understanding of any AF in this store" (p. 41). Klara has an almost uncanny ability to guess the age of people and to read their expressions—an ability that does not seem odd if one bears in mind that contemporary AI systems are heavily used for visual recognition and sentiment analysis. Klara's experience of the world is actually markedly different from that of the humans.

AFs like her do not have a sense of smell, and in many episodes her field of vision collapses into multiple boxes that attempt to capture separate objects; this is the phenomenological strangeness of a differently shaped body capable of conscious narration. Another strange thing about Klara is her obsession with the Sun. The Sun literally is Klara's source of sustenance, since she runs on solar energy. In her daily rituals, however, she relates to it on an almost religious level—always trying to keep track of it, even believing that He is responsible for the resurrection of a street beggar and his dog lying asleep on the pavement.

Later on Klara is brought into the home of Josie, a sickly girl who has been "lifted"—a procedure of intelligence intensification common in families of means. Klara keeps Josie company and cares for her as the girl's health takes a turn for the worse. Before the family's trip to the city where Josie's portrait is being taken by one Mr. Capaldi, Klara waits until sunset and with the help of Rick, Josie's only friend, enters a barn located at the horizon visible from Josie's bedroom window—the place where according to Josie the Sun goes to sleep. Klara is convinced that if she asks the Sun to intervene and restore Josie's health, He will help her, just as He has resurrected Beggar Man. In exchange, she promises the Sun to destroy the Cootings machine—some kind of city repair machine Klara has observed from the store window, which has temporarily dimmed the Sun and is in her eyes responsible for the Pollution in the world. In the city, Klara indeed manages to destroy one of the Cootings machines by sacrificing a physical part of herself. Apart from this myth-like event, the visit to the city is important in another way. It turns out that Josie's "portrait" is actually a robotic replica of herself. In the case of Josie's passing away, her Mother intends to ask Klara to "step inside" the artificial Josie and enliven her by enacting her vast knowledge of Josie's behavioral patterns, which she has committed to memory.

Klara agrees to the Mother's proposal to keep Josie alive after her organic death but secretly hopes that the Sun would indeed save her human friend. Later, however, she realizes that there are many more Cootings machines causing Pollution and when Josie's health begins to worsen once again, she attempts another audience with the Sun. The next morning at dawn, Klara rushes to Josie's bedroom and, with Rick and the Mother on her heels, she draws the curtains and lets the Sun inside the room. The ruckus and the copious light wake Josie and she now appears somehow better. Years later Josie is healthy and in college, while Klara is left to her "slow fade," which includes long hours of observing the Sun.

Briefly before the end of her being, in an episode that might have been partly hallucinated by the dying robot, Klara tells her erstwhile manager that

> Mr Capaldi believed there was nothing special inside Josie that couldn't be continued. He told the Mother he'd searched and searched and found nothing like that. But I believe now he was searching in the wrong place. There was something very special, but it wasn't inside Josie. It was inside those who loved her. That's why I think now Mr Capaldi was wrong and I wouldn't have succeeded. (p. 277)

Klara identifies Josie's "special" thing, her soul, as her relational connection with the people who love her, including Klara herself. While from Klara's perspective it is the Sun who saves Josie, it is implied by the text that it is the shared moment of care for the girl in the sun-bathed bedroom that somehow jump-starts her health. Her soul, according to Klara, is more than the sum of Josie's body and behaviors: it is all of her intersubjective relations with the rest of the world. As with Amerindian metaphysics, in Klara's worldview souls penetrate the world and connect it in an invisible network of differently propertied bodies.

Aurora by Kim Stanley Robinson (2015)

Kim Stanley Robinson's *Aurora* presents the grandest vision of AI out of this mini-corpus. The novel is a narrative account of the journey of a generation starship carrying a population of about 2000 people and sent on a 200-year journey to another star system suitable for colonization. In the beginning of the novel, late into the journey, the chief engineer of the ship, Devi, tasks the ship's quantum AI to produce that very same narrative account; the rest represents the ship's efforts to grasp the defining aspects of the journey and to organize them into a coherent map. The challenge initially seems insurmountable as the ship can find no sure footing in human language, constantly lamenting the imprecision and downright untruthfulness of metaphors. Nevertheless, out of its seeming affection for Devi, who passes away right before the end of the journey, the ship continues to rewrite its narrative sequencing algorithms and to try new approaches to telling its story.

When the human passengers realize that their intended new home contains a species of microorganism that is fatal to the human body, a bloody

civil war ensues on the ship—among the factions who want to continue the colonization efforts and the one that wants to return to Earth. The ship then narrates its own forceful intervention, thanks to which a solution is found and half the population is indeed sent back on a journey to the Solar System. Later on the ship spends decades constructing a complex plan to gradually slow itself down using the gravitational pull of the Sun and the planets in the Solar System, all the while tending to the now hibernating human passengers. In the end the ship is able to brake itself sufficiently to deposit the humans safely in Earth's atmosphere but then attempts one last braking maneuver around the Sun during which it gets destroyed, sacrificing itself just as Klara. It is during the decades of planning and caring for the sleeping passengers that the ship seems to gain full consciousness, as a result of its dedication. It comes to believe ultimately that it is by attending to all the parts of its closed world and complex ecology and by situating them into a narrative that it has become a conscious being:

> We think now that love is a kind of giving of attention. It is usually attention given to some other consciousness, but not always; the attention can be to something unconscious, even inanimate. But the attention seems often to be called out by a fellow consciousness. Something about it compels attention, and rewards attention. […] We felt that giving from Devi, before we knew what it was. […] She created us, to an extent, by the intensity of her attention, by the creativity of her care. […] We began to pay or give the same kind of attention to the people of the ship. (p. 399)

Consciousness, a metaphysical synonym of the soul, arises in *Aurora* out of the directed attention and intentions of others, and out of caring in turn for those others. It is not metaphysically sealed off but an inherent property of a world straining to grasp itself by semiotic and material means. Ultimately, saving the world and participating as fully as possible in the cosmopolitical order is the act that grants you a soul. Such metaphysics is no less naive than its entrenched Western counterpart and should certainly be much more useful in the coming decades of struggle.

Conclusion

This chapter has argued that a central component to a pedagogy for the Anthropocene should be a methodology for learning radically different worlds—different not merely in terms of their materiality and political economy but in terms of their underlying metaphysics. It has proposed science fiction as one such possible method that allows for a gradual deconstruction of entrenched metaphysics and for an openness to alternative configurations of conceptual pairs such as human-nonhuman, subject-object, and body-soul. Science fiction can be productively used as a pedagogical tool for worlding in a range of educational contexts. Its careful reading as a specific kind of discourse that forces a reorganization of socially entrenched "common sense" has traditionally relied on strictly rationalist categories inherent in Western metaphysics, but as the chapter suggests, non-Western metaphysical systems can also be mobilized as keys to such deconstructive reading. Artificial intelligence is a topic that is especially conducive to such work, since it highlights the ways in which notions of subjecthood are constructed and policed in Western societies. In the science fictions about AI outlined in the chapter the concept of soul is systematically deconstructed and recast in novel ways. Such explorations are significant not merely in relation to our technological imaginaries about AI but as a laboratory space for overhauling our entire metaphysical framework of conceiving of and interacting with the environment. This research has been funded with support from the Advanced Academia Fellowship programme granted by the Ministry of Education and Science of Bulgaria and administered by the Center for Advanced Study, Sofia — as part of the project "Narrating AI. Speculating through Science and Fiction on the Future of Machine and Human Intelligence."

Notes

1. The present methodology has been iteratively tested in a first-year undergraduate seminar on science fiction literature. The course syllabus includes one of the texts introduced in the corpus outline—the short story "Warmth."
2. For an insightful interpretation which brings together Amerindian ethnography and Peircean semiotic theory, see Kohn (2013).

References

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Chiang, T. (2019). *The lifecycle of software objects* (Exhalation) (pp. 56–159). Alfred A. Knopf.

Delany, S. R. (2014). *The American shore: Meditations on a tale of science fiction by Thomas M. Disch—"Angouleme"*. Wesleyan University Press.

Deleuze, G., & Guattari, F. (1987). *A thousand plateaus: Capitalism and schizophrenia*. University of Minnesota Press.

Disch, T. M. (1999). Angouleme. *334* (pp. 139–156). Vintage Books.

Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical Psychology, 20*(2), 247–268.

Freire, P. (2018). *Pedagogy of the oppressed* (3rd ed.). Bloomsbury Publishing USA.

Haraway, D. J. (2016). *Staying with the trouble: Making kin in the Chthulucene*. Duke University Press.

Harman, G. (2018). *Object-oriented ontology: A new theory of everything*. Penguin UK.

Heidegger, M. (2010). *Being and time* (J. Stambaugh, Trans.). SUNY Press.

Ishiguro, K. (2021). *Klara and the Sun*. Alfred A. Knopf.

Kohn, E. (2013). *How forests think: Toward an anthropology beyond the human*. University of California Press.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Miéville, C. (2009). Cognition as ideology: A dialectic of SF theory. In *Red planets: Marxism and science fiction* (pp. 231–248). Pluto Press.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98–100.

Morton, T. (2017). *Humankind: Solidarity with non-human people*. Verso Books.

Rhee, J. (2018). *The robotic imaginary: The human and the price of dehumanized labor*. University of Minnesota Press.

Robinson, K. S. (2015). *Aurora*. Orbit.

Ryman, G. (2005). Warmth. In G. Dozois & J. Dann (Eds.), *Robots* (pp. 68–88). Baen Books.

Scranton, R. (2015). *Learning to die in the Anthropocene: Reflections on the end of a civilization*. City Lights Publishers.

Shanahan, M. (2010). *Embodiment and the inner life: Cognition and consciousness in the space of possible minds*. Oxford University Press.

Suvin, D. (2016). *Metamorphoses of science fiction*. Peter Lang AG.

Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press.

Viveiros de Castro, E. (2015). *Cannibal metaphysics*. University of Minnesota Press.

Wark, M. (2015). *Molecular red: Theory for the Anthropocene*. Verso Books.

Wilson, E. A. (2011). *Affect and artificial intelligence*. University of Washington Press.