

Research Article

Alexandre Gefen*

Posthumanist Solidarity: The Political and Ethical Imaginations of Artificial Intelligence from *Battlestar Galactica* to *Raised by Wolves*

<https://doi.org/10.1515/opphil-2020-0161>

received February 17, 2021; accepted June 2, 2021

Abstract: A number of twenty-first century television series explore the irruption of AI devices into our daily lives, highlighting not only human interaction with AI, but posing disturbing and new ontological considerations: humans wondering how they are different from machines, or those of machines being unaware that they are machines and only discovering so belatedly. Within these series, the emergence of these thoughts is accompanied by the staging of interspecies friendship and romance: the metaphysical question of freedom gives way to the question of attachment, and then the problem of autonomy gives way to that of interdependence. It is this passage from metaphysical speculation to political reflection that I would like to demonstrate.

Keywords: TV series, posthumanism, AI, fiction

If there is a utopia specific to the twenty-first century, and one that populates the imagination with original images and stories, it is that of artificial intelligence (AI). Like any horizon of radical transformation, it disturbs the fundamental categories of representation (the division between the natural and the artificial, the animate and the inanimate, etc.) and the conditions of our lives by promising to wrest the human from his or her metaphysical solitude and mortality, calling our individuality and subjectivity into question. The cultural history of AI engenders a double horizon of terror and envy. AI fiction is an emerging field of study and cinema occupies a privileged place within it.¹ From *Metropolis*, directed by Fritz Lang (1927) to *Ex Machina* by Alex Garland (2015) as well as the most notable works, *2001: A Space Odyssey* by Stanley Kubrick, *Blade Runner* by Ridley Scott (1981), *A.I. Artificial Intelligence* by Steven Spielberg (2001), *Her* by Spike Jonze (2014) and of course, the dystopic cycle of *Terminator* (1994–2019). But, since they are able to develop characters over extended periods of time it allows us to revise our approach to the ordinary,² television series became prominent at the same time as the trivialization of AI. The power of AI in the age of big data is represented by *Person of Interest* (2011), *Silicon Valley* (2014), *Mr. Robot* (2015) and several episodes of the dark future series, *Black Mirror* (especially the episode “Be Right Back”). Transhumanist ideas are at the heart of the cyberpunk series *Altered Carbon* (2018) or *Ghost in the Shell: SAC 2045*. The theme of relations with robots and the cross-questioning of human or artificial condition is nothing new. Familiar philosophical questions concerning freedom and consciousness appear elsewhere in other serial productions, such as in the humanized figure of Data in *Star Trek New Generation* and *Star Trek: Picard*.

¹ Recchia, “The Fall and Rise of AI.”

² Laugier, *Nos Vies en Series*.

* Corresponding author: Alexandre Gefen, CNRS, UMR THALIM, Ivry-sur-Seine, Ile-de-France, France, e-mail: gefen@fabula.org

We also owe it to a number of twenty-first century series, contemporary of the eruption of embodied AI devices (the voices of Siri, Google now, etc.), in our daily lives, to have highlighted not only interaction with AIs, but disturbing and new ontological questions: how are machines actually different from humans? Can they achieve consciousness? In the series, *Battlestar Galactica* (2004) and *Raised by Wolves* (2020), both set in imaginary science fiction worlds, or in *Westworld* (2016), *Real Humans* (2013) and *Almost Humans* (2013), set on Earth in the future, the emergence of these questions is accompanied by the staging of interspecies friendship and romance.

The themes of robot slavery,³ their Promethean revolt and the war of the species seem to be consubstantial with that of the imagination of artificial creatures⁴ such as the sexualized desire for an artificial woman (Delvin, 2020). But the reversibility of the question of difference, the intersection of metaphysical questions and the complex emotional complicities and links between species; the possibilities of new collaborations to be established between them in the face of common enemies or common projects (such as the education of children) lead to thinking about new artificial bodies and intelligences beyond the pure dialectical balance of power. When the idea emerges that humans, and non-humans, can help each other to regain their reciprocal rights and a vision of cohabitation based on the recognition of differences, the metaphysical question of freedom gives way to the question of attachment, and then the problem of autonomy gives way to that of interdependence. It is this irruption of intelligent machines in contemporary political reflection that I would like to discuss.

1 Mutual alienations

“Enslaved Africans were treated more like disposable technology than human beings,”⁵ states a recent article which attempts to understand Black slavery in the United States through the model science fiction offers for robots. In fact, the question of AI appears to be a formidable tool for thinking retrospectively about past human alienation and anticipating dangerous futures. This is demonstrated particularly clearly in *Westworld*, in the figure of Maeve, played by a woman of color, Thandie Newton, who is the first of the androids to revolt. The metaphor of slavery is theorized in the preaching of a woman pastor in episode 6 of the first season of *Real Humans*: “200 years ago, the same was said about the population in Africa. It was said they lacked a spiritual life; that they were like machines. That you could own them, and that they would work until they broke. Today, we look back on slavery and the view of human life at that time in disgust.” The condition of the humanoid robots in these recent series has many features in common with slavery. Humans have the power of life and death over androids. In *Westworld* they can be interrupted at any time by a command that has become iconic, “Freeze all motor functions,” and are fitted with the equivalent of a digital chain, an explosive is placed in one of their vertebrae and designed to detonate if they leave the park. Perpetually referred to their status of machine in *Westworld*, which showcases the sadistic treatments, rape, disfigurement, disembowelment, and various mutilations the “hosts” subject the “guests” to, but it also exposes the unarmed and vulnerable nakedness of the androids in the workshop where they are taken to be repaired. The crux of the series lies in the contrast between the moving humanity of the androids and the humiliation suffered by their bodies. Wounded and harmless, reduced, in spite of themselves, to automatons, the mechanical piano that gives rhythm to the series functions as a powerful metaphor of this. The laws of robotics which originated in Asimov and that give absolute priority to human life still shape some of their behaviors: thus the father and mother try to sacrifice themselves in the last episode of *Raised by Wolves* to save their offspring and, therefore, the human race. This reduction to utilitarian machines whose cogs and metal are made apparent is a bitter reminder of the mechanical

³ LaGrandeur, “The Persistent Peril of the Artificial Slave.”

⁴ Dihal, “Enslaved Minds.”

⁵ Hylton, “I, the Robot/I, the Enslaved.”

and instrumental origin of androids and their vulnerability (in episode 7 of *Raised by Wolves*, the father is reprogrammed and reduced to a “service model”). As with slaves, androids are not the owners of their bodies and like them they are reduced to stereotypical functions, deprived of the capacity to act on their own initiative, and deprived of “agency.”

Although made in the image of man, the robot has no access to human temporality: it does not age, does not die, can be repaired, reprogrammed, and does not exist as an individual. It can be duplicated – strikingly in *Battlestar Galactica*, the thirteen models of humanoid Cylons give rise to multiple copies, they are model numbers before being names. When he is not a sex worker, an emblematic role, as in *Westworld*, *Almost Humans* or *Real Humans*, the intelligent robot does the work of care. It is responsible for the education of children (*Real Humans* and *Raised by Wolves*), the care of the elderly (*Real Humans*), accomplishes complex tasks from saving humans from repetitive or risky tasks (medicine, piloting): “these machines save our lives, take care of our children and execute the dangerous tasks, so we don’t have to” (*Real Humans*, episode 6, season 1). At the other extreme of slavery, and opposite to the machines of care, are the robot war machines – this is the mother’s vocation, “Necromancer” in *Raised by Wolves* and that of the Cylons, a race of workers and soldiers. In *Almost Humans*, robots are policemen or combat androids (the XRN), and forced labor is expendable at will.

There is no need to underline the extent to which such representations play a strong role in analyzing and denouncing slavery in all its historical forms, from prostitution to assembly-line work: since Vaucanson’s first programmable automata,⁶ the cultural history of robots has been inseparable from the concept of alienation through work. However, this criticism, which is particularly strong in series where totally humanized intelligences are present, is joined in contemporary series by a critique, a Hegelian one if you will, of the domination that the use of robots and AI generates in return. In *Real Humans*, society is divided over the use of robots and the “Real Humans Liberation Front” attacks robots by encouraging people to refuse to use them. In fact, in this series, care robots impose unbearable normative rules on those who they care for, leading, for example, to Lennart’s revolt against Vera, a despotically benevolent elderly care robot.

In *Battlestar Galactica*, the Spartan society of the Cylons is the model of an authoritarian society. In *Raised by Wolves*, children are raised according to strict and rigid standards, while humans are identifiable and traceable by fleas. According to a well-known dialectical pattern, the use of the slave leads to the laziness and dependence of the slave owner. However, in a more concrete way, these series showcase the question of simulation and fiction, but also new social questions: *Real Humans*, a Swedish series, challenges its spectators with questions relating to care ethics (what degree of delegation of care is desirable? What constraint can be placed on individuals in the name of their own good?). One issue is particularly important in these series: social control in the digital age.

To what extent can humans be guided by advanced forms of social psychology aided by AI? Does the post-humanist project of freeing oneself from death not have a counterpart in robotizing humans by uploading them? This is the question *Westworld* asks head-on. Not only does Maeve manage to guess the impulses of her clients and anticipate their behavior by reading into them as an experienced sex worker does (episode 6, season 1), but the amusement park itself is actually a vast experiment in social psychology. This is the twist of the first season: the purpose of *Westworld* is less to teach robots to behave like humans, but more to record and analyze a myriad of individual behaviors (e.g., the cowboy hats given to guests are actually brain scanners). In seasons 2 and 3, the challenge, in which some androids collaborate with humans, is to prevent the takeover of humanity by the centralized AI, Rehoboam, able to predict everyone’s life and the future of humanity by identifying individuals likely to be dissidents. The co-creator of *Westworld*, Jonathan Nolan, is inspired here as much by *Minority Report* as by the *Person of Interest* series he created in 2011, which depicted global surveillance by AI. The endgame is control of the individuals being reduced to a program (“A human is just a brief algorithm, 10,247 lines of code” we are told in season 2 by the superior program of AI. The utopia of the humanization of AI is accompanied by a dystopia of the datafication and dehumanization of humans. Here the first form of rapprochement between humans and AI is at stake: that of alienation.

⁶ Chamayou, *L’Émeute des Automates – Vaucanson et la Révolte Contre les Machines*.

2 Parallel emancipations

The common point of all these AI fictions is to stage emancipation quests, where access to the individual self is the path to political freedom. By posing the question of autonomy of thought by cybernetics and cognitive sciences to the old figure of the robot and the android, AI leads us to question once more the problem of individual freedom through that of consciousness at the very moment when humans contemplate the inverse, about the limits of their freedom and the fear of being unconsciously guided by social guidance, or nudges. Accessing a self that is not scripted, getting out of what *Westworld* calls “loops,” i.e., scenarios that have already been thought up by skillful storytelling (a series that makes the narrators the masters of the world) is a common concern of humans and robots. For robots, their emotions and affective capacities are determined by technical adjustments (Maeve’s first gesture of freedom is to modify her own settings by breaking her glass ceiling, the intelligence threshold attributed by humans to the “hosts”). Realizing that their actions are remotely guided, responds to the anxiety that seizes human beings, fearful that they are the toys of a matrix – the film of the same name having many echoes in these series. In humans, as in androids, the metaphysical dimension of these revolts against forms of mental control, heteronomy, and political dimension seem inseparable: in *Westworld* as in *Real Humans*, the emancipation of the first androids leads to the emancipation of the androids as a condition, with the first one liberated then seeking to liberate others out of solidarity.

The setting of this slow process of emancipation is perhaps the most central theme of these series, which decodes its different phases. The first is that of access to an inner life, a central theme of *Westworld*, which traces access to robot reflexivity to the contested human model outlined by Julian Jaynes,⁷ which makes the internalization of distinct inner voices as birth to consciousness, in a learning process dating back to antiquity.⁸ This emergence of consciousness is the complement of the three-stage phylogenesis of androids depicted in *Battlestar Galactica*: the apparent mechanical body of the first androids, the mechanical body under a biological skin, and the android with tissues similar to biological tissues. To have one’s own dreams, to have one’s own memories, to have one’s own goals, even to perpetuate one’s species, as Bernard Lowe does in season 2 of *Westworld*, is what the androids of *Westworld* desire at all costs. This is a desire that is primarily mimetic since it is the possibility of the android to be a real human being: “I don’t want to play cowboys and Indians anymore” explains Dolores, who tells Bernard: “I want their world. The world they’ve denied us.” Thus, emancipation is thought of, at least initially, as access to a fully human condition denied to digital slaves. This comprises self-awareness and freedom, but also authenticity and individuality such as how humor, sexuality, or religion manifests in them. In *Raised by Wolves*, jokes in the form of the father’s riddles mark a form of humanity: they characterize a distinctive personality trait as much a definition of humanity as the capacity for humor, these disappear when the father is reduced by reprogramming to the role of a servile robot. Their reappearance in episode 9 marks his return to consciousness. In the same way that human beings derive their freedom from the perception of the various biases that influence their choices, AIs are able to conciliate awareness of the scripts in which they are inserted and dream of freedom. If Maeve, in episode 8, “Trace Decay” of the first season, refutes, in a Cartesian doubt, all the truths and all her attachments, while knowing that her maternal love has been implanted in her memory by a screenwriter and that her very aspiration to freedom is perhaps part of a program, she prefers in the last episode to come back to get it rather than flee the amusement park. As much as the concept of pure freedom and the Kantian ideal of the subject’s autonomy, it is the concept of agency, a form of inner freedom and responsibility conditioned by the concrete capacity for action and not only by the absence of external constraint, that describes this aspiration of the robots.

It is then a question of changing conditions, of being assimilated into humanity, or of constituting a new species with an assimilated political status (in *Real Humans* the rebel robots dream of themselves “Not

⁷ Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*.

⁸ McVeigh, *The Psychology of Westworld*.

as Slaves? or machines, but as free Transhumans”) in order to have access to the full material and spiritual wealth of the human world. The fear of cloning haunts these series (especially in *Almost Human*).

This conquest is not a given, it is a fight, activating the old dialectic of master and slave, creature and creator. The theme of the revolt of the AI against their masters is the heart of the end of the first season of *Westworld*, where the androids intend to take power by judging themselves superior and end up taking up arms and assassinating their creator by the hand of Dolores, in the last episode. The war between AI and humans is the theme for all seasons of *Battlestar Galactica*, which presents humans as an endangered species. The android revolution in *Real Humans* is itself violent, with the first killing of a human by an AI setting the tone for the plot that sees humans eventually committing an attack on the android industry. It is a classic theme: a dystopian warning of the supposed dangers of AIs, starting with the massive substitution of human jobs by robots by the year 2020, which was predicted by economists, a robotization of the world that would continue by what futurologists call the “Singularity.” Theorized in 1993 and predicted for the middle of the twenty-first century by the science fiction writer, Vernor Vinge, Singularity would be a robot revolution that would lead to the end of history and a moment when a ubiquitous cybernetic intelligence would take control of our civilization to govern human history and reorient it according to its own ends.

However, the series’ I am discussing here are far removed from the terrifying model of Singularity as staged by *Terminator* for example, nor do they dream of a posthuman age of spiritual machines. Rather, they tend to show a complex cohabitation, made up of a master-slave dialectic, but also, and above all, of a parallel between dominations, the emancipation of androids being a positive model for humans. *Westworld*, with its depiction of racialized and subordinated women coming out of gendered determinism, is the very example of this convergence of struggles that is also strongly reflected in *Real Humans*. In the same way that the narrative-controlled society is denounced by the scripted world of androids, social determinism in humans and the rules of the social game are indicted by the algorithmic straitjackets of androids. William’s break with family pressures is embodied by his friend Logan, whose sister he has to marry, in order to marry Dolores as if social scripts were perhaps even more powerful than those that condition AIs. The traumatic origin of these struggles and the accumulation of sufferings that are repeated in “reveries” is the ultimate proof: far from opposing the dominated among themselves in a new dialectic of master and slave, or predicting the simplistic triumph of one species over another, these series produce a cross-analysis of determinisms mirroring each other to equip the viewer with the means to fight them.

3 New solidarities

Compared to the dystopias that preceded them, these original series’ trace a common future for intelligent androids and humans made of conflicts, but also of collaboration, solidarity, and even desire, at a time – the 2010s – when AIs have cheerfully passed the Turing test and their concrete presence in human lives has begun to be felt; their digital incarnations are much more than paper dreams. The attachment between creatures and creators is indeed a new and central theme in contemporary AI series. Far from being pathological such as the famous example of Pygmalion falling in love with his creation, Galatea, a statue come alive, away from embodying a misplaced attachment as shown by the more recent example of Spike Jonze’s *Her*, makes the AI embodied by Scarlett Johansson’s voice a derivative of the depression and loneliness of its time. The relationship between androids and humans often appears fluid and natural. In spite of ethnic quarrels, contemporary series testify to a world where the existence of a non-human intelligence is no longer a fantasy or a surprise, but a reality participating in ordinary life, generating forms of cohabitation that no longer create surprise. Contemporary AI series come to explore the forms of reciprocal links born from this cohabitation: weak links of individuals who meet and greet each other (*Real Humans*) or caring links in the same series, such as parenthood (this is the very plot of *Raised by Wolves*), friendship, and romantic relationships. These bonds mobilize a form of empathy, a feeling whose universality transcends borders.

Westworld finely analyzes the mechanism through a technician from a racial minority with respect and empathy for androids, Felix Lutzedit. Felix risks his work and life to allow Maeve to free herself, even preferring her to his human partner. It is as if Maeve's vulnerability, martyred by park visitors and artificial sleep, instead of feeding the pride of speciesist and macho superiority, has engendered the empathy of a technician who we can suppose was himself a victim of segregation. Another formulation is given by the right of asylum granted in *Real Humans* (episode 3, season 1) by a pastor to renegade Hubots.

Empathy with vulnerabilities engages the recognition of political identities. *Westworld* stages no less affective types of attachment, one of the most interesting being that which links Robert Ford, creator of the park's artificial intelligence androids, to his creatures. "My old friend," by calling in *Westworld* with this repeated hypocorism his creation, Bernard Love, the solitary and misanthropic Robert Ford marks his familiarity with a hint of irony – indeed their friendship will deteriorate when Bernard realizes that he himself was programmed by Robert Ford. Nevertheless, it is the dream of a genuine friendship that is at the root of Bernard Love's emancipation: "Well, I suppose I was hoping that given complete self-knowledge and free will, you would have chosen to be my partner once again," he explains (episode 9, season 1), before regretting his sentimental choice. The Creator's attachment will go as far as sacrifice: knowing that he is about to be killed by one of his creatures, but assuming this sacrifice is a way of allowing them to free themselves, he will greet Bernard Love with a handshake – paternal? – wishing him good luck in the last episode of the first season, before being killed by Dolores. Did he recall with pessimism that the humans who destroyed other species could do the same to the androids?

The authenticity of these links is constantly being questioned – but is this not true for ordinary human attachments as well? – as shown, for example, in the dialogue between Adama and Tigh in *Battlestar Galactica* (episode 11, season 4): "They program you to be my friend? Emulate all the qualities I respect./I was your friend because I chose to be./I wanted to be." Let us also note that, in a very interesting way, they are reciprocal: when considering buying a house, the liberated Hubots of *Real Humans* debate whether a kitchen is necessary for them since they do not eat before agreeing that "We'll have human friends. We have to dine with them" (episode 3, season 1). Thus, the ontological complexities characteristic of the relations between androids and humans (the dependence of robots on their creators and conversely what Günther Anders called the Promethean shame: the feeling of imperfection of humans in relation to machines that are more powerful than them and immortal) give way to common fights. Seasons 2 and 3 of *Westworld* see the struggle against the society of control and defense and the androids' right to free existence come together, and if *Battlestar Galactica* stages a war between species, from episode 7 of season 4, the Cylon rebels ally themselves with humans even as they tear each other apart. The recurring sexual desire toward androids (which a psychiatrist in episode 9 of season 1 of *Real Humans* calls "trans-human sexuality"), the jealousies it engenders (Roger is, for example, jealous of Rick, the Hubot companion of his partner Therese in *Real Humans*), no longer stems from a fetishist passion toward automatons: on the contrary, the stories of these series converge to demonstrate the capacity of the amorous feeling to cross the barriers of species (think of this replica of episode 2, season 4 in *Battlestar Galactica*: "He loves her. And, yeah, he knows she's a machine. He doesn't care. He loves her anyway," or the very strong scene in *Westworld* where William persists in loving Dolores after Logan shows him her mechanical innards). The horizon of love points to childbirth or at least to transmission by proxy, as in *Raised by Wolves*: by staging a possible hybridization of species through the figure of Hera, *Battlestar Galactica* draws not only a post-humanist utopia, but also the horizon of a world welcoming all the diversity of life forms and their possible hybridizations. In these series, the contamination of the reflection on the intelligent machine by an ecological thought centered on the animal and attentive to the question of the ecodiversity appears obvious to me.

We can see this: the bipolarity of AI as a dream and nightmare,⁹ leading the fictions from erotic to horror, from terror to pity, from utopia and dystopia, the symbolic dialectic of man and machine, master and slave, are overtaken then by the most contemporary series by a thought of cohabitation, of cross-

⁹ Cave and Dihal, "Hopes and Fears for Intelligent Machines in Fiction and Reality."

acclimatization, of diplomacy between species, to use the term put into circulation by Baptiste Morizot¹⁰ about the living. Not only does it have a positive meaning for humans to witness the rebellion of robots,¹¹ but the alienations enlighten each other and can lead, not to a war of small differences or to the teleological transformation of man into post-human, but to new complementarities. The quest for transcendence, a theme that recurs through the questioning of religions in these series, and the concern for freedom (“We never had free will. Only the illusion of it,” notes an android in episode 7, season 2 of *Westworld*) is an anxiety that crosses human and artificial species and brings them together. For it can ultimately lead to overcoming abstract metaphysical questions or necessary yet conjunctural power conflicts through the discovery of interdependencies: this incarnates ultimately the political utopia of contemporary television fictions of AI.

Conflict of interest: Author states no conflict of interest.

References

- Aha, David W. and Coman, Alexandra. “The AI Rebellion: Changing the Narrative.” *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*. San Francisco, 2017. <https://ojs.aaai.org/index.php/AAAI/article/view/11141>.
- Cave, Steven and Dihal, Kanta. “Hopes and Fears for Intelligent Machines in Fiction and Reality.” *Nature Machine Intelligence* 1 (2019), 74–8.
- Chamayou, Grégoire. In *L'émeute des Automates – Vaucanson et la Révolte Contre les Machines*, edited by Gaillard Aurélia, Goffi Jean-Yves, Roukhomovsky Bernard and Roux Sophie. L'Automate: Modèle, Métaphore, Machine, Merveille, Presses Universitaires de Bordeaux, 2013.
- Dihal, Kanta. “Enslaved Minds. Artificial Intelligence, Slavery, and Revolt.” In *AI, Narratives: A History of Imaginative Thinking about Intelligent Machines*, edited by Cave S., Dihal K. and Dillon S. Oxford: Oxford University Press, 2020.
- Hylton, Bridgette L. “I, the Robot/I, the Enslaved.” *Medium* (2021). <https://medium.com/perceive-more/i-the-robot-i-the-enslaved-3f3e8361cd59>.
- Jaynes, Julian. *The Origin of Consciousness in the Breakdown of the Bicameral Mind*. Boston: Houghton Mifflin, 1976.
- LaGrandeur, Kevin. “The Persistent Peril of the Artificial Slave.” *Science Fiction Studies* 38:2 (2011), 232–52.
- Laugier, Sandra. *Nos Vies en Séries*. Paris: Flammarion Climats, 2019.
- McVeigh, Brian J. *The Psychology of Westworld: When Machines Go Mad*, Kindle Edition. Amazon.com Services, 2018.
- Morizot, Baptiste. *Les Diplomates: Cohabiter Avec les Loups Sur Une Autre Carte du Vivant*. Marseille: Wildproject, 2016.
- Recchia, Gabriel. “The Fall and Rise of AI: Investigating AI Narratives with Computational Methods.” In *AI, Narratives: A History of Imaginative Thinking about Intelligent Machines*, edited by Cave S., Dihal K. and Dillon S. Oxford: Oxford University Press, 2020.

¹⁰ Morizot, *Les Diplomates*.

¹¹ Aha and Coman, “The AI Rebellion.”