

## Technophilosophical underpinnings of *Westworld* (2016–2022)

Using the events of the HBO series *Westworld* (2016–2022) as a springboard, this paper attempts to elicit a number of philosophical arguments, dilemmas, and questions concerning technology and artificial intelligence (AI). The paper is intended to encourage readers to learn more about intriguing technophilosophical debates. The first section discusses the dispute between memory and consciousness in the context of an artificially intelligent robot. The second section delves into the issues of reality and morality for humans and AI. The final segment speculates on the potential of a social interaction between sentient AI and humans. The narrative of the show serves as a glue that binds together the various ideas that are covered during the show, which in turn makes the philosophical discussions more intriguing.

**Keywords:** *sentient AI, Westworld, consciousness, immortality, reality*

### Author Information

**Pankaj Singh**, University of Petroleum and Energy Studies (UPES), Dehradun, India, School of Liberal Studies (SoLS)

<https://orcid.org/0000-0002-6570-3634>

### How to cite this article:

Singh, Pankaj. “Technophilosophical underpinnings of *Westworld* (2016–2022)”.

*Információs Társadalom* XXIV, no. 2 (2024): 101–114.

== <https://dx.doi.org/10.22503/inftars.XXIV.2024.2.6> ==

*All materials  
published in this journal are licenced  
as CC-by-nc-nd 4.0*

---

## 1. Introduction

American television series *Westworld* (2016–2022) contemplates various philosophical issues, concerns, and possibilities between humans and artificial intelligence (AI). The series ended after four seasons. Each season of *Westworld* was so dense with philosophical problems that numerous edited volumes have been published to analyze the series. *Westworld and Philosophy: If You Go Looking for the Truth, Get the Whole Thing* (South and Engels 2018), *Reading Westworld* (Goody and Mackay 2019), and *Westworld and Philosophy: Mind Equals Blown* (Greene and Heter 2018) are some of the notable insightful readings for readers interested in the specific multitude of philosophical concerns addressed using the show's events. Viidalepp (2020) performed an analysis on the portrayal of robots in science fiction movies and television shows, including *Westworld*. In his analysis, he concluded: "In real-world discourse, there is uncertainty and fragmented knowledge concerning current technological developments as well as their scientific significance. Considering also the superficial understanding of the functioning of human identity, societies and cultures, the assumptions appearing in technological discourse readily blur the boundary between the man and the machine as easily as happens in fiction" (Viidalepp 2020, 34).

This paper takes a different approach to uncover *Westworld's* technophilosophical underpinnings.<sup>1</sup> Instead of evaluating various philosophical concepts addressed discretely, this essay will examine the central themes of the series. The relationship between humans and artificially intelligent robots has come full circle. The events of the series remain treasures for further evoking philosophical concerns about humanity and AI. The article will take a holistic approach, focusing on the technophilosophical roots. The first season, titled *The Maze*, investigates the maze of consciousness. It poses questions such as what is necessary for a robot to be conscious. Or is it possible for a robot to possess true consciousness? Or, more to the point, can a robot have consciousness on par with humans? *The Door*, the second season, centers on the idea of a virtual life, a possible digital afterlife, and immortality. The third season, *The New World*, depicts a world governed by AI, which controls the destinies of humanity by making decisions for them. It focuses on the real-world interaction between humans and AI. The three sections of this article are titled after the seasons and explain each respective theme. The following synopsis will make it easier for readers to follow the paper in its entirety: "*Westworld* is a fictitious Wild West-themed amusement park populated by robot 'hosts' who pander to the whims and fancies of wealthy human visitors. The real world of the mid-21st century is also explored in the third season." The paper mentions the key characteristics of the series in footnotes after the first relevant mention in the storyline.

---

<sup>1</sup> I have borrowed the term technophilosophy from Chalmers (2022, 13): "Technophilosophy is a combination of (1) asking philosophical questions about technology and (2) using technology to help answer traditional philosophical questions."

## 2. The *Maze* of consciousness

The very first episode of *Westworld* dives right into deep philosophical territory with a series of questions on the nature of reality when Bernard, a human, troubleshoots a humanoid, Dolores, by asking a about the *reality* of their existence (season 1, episode 1).<sup>2</sup> In addition, William, the man in black, is looking for the deeper meaning of the *Westworld*-themed amusement park. The deeper significance has to do with a maze, specifically a maze of consciousness.<sup>3</sup> Ford's update to the AI host in the park as reveries enabled the android host to access host's previous roles in the park, so granting them access to the greater memory and subconscious.<sup>4</sup> Thus, memory is believed to be essential for true sentience and consciousness.

Humans' interest in mind, memory, and consciousness is arguably as old as humans themselves. Philosopher John Locke regarded humans as a "tabula rasa," a blank slate, in his understanding of memory. The blank slate is filled in with experience, which results in the formation of memories. He considered memories as power of mind, "to revive perceptions, which it has once had, with this additional perception annexed to them, that it has had them before" (Locke [1690] 1975, 150). Experimental studies in psychology and cognitive science have at last brought philosophers' interest in memory to fruition. With the growth of studies in psychology, neuroscience, and computer science, the information processing approach to memory has become the mainstream perspective for comprehending the complexities of memory (Atkinson and Shiffrin 1968; Baddeley, Allen, and Hitch 2017; Cowan 1999; Ericsson and Kintsch 1995)

For philosophers and cognitive scientists alike, consciousness is even harder to pin down than memory. Researchers have struggled to define and comprehend consciousness. Philosopher David Chalmers has offered a helpful delineation addressing the topic of consciousness to highlight the difficulty in understanding it. He divided the issue of consciousness into two distinct categories: the easy problem of consciousness and the hard problem of consciousness. Chalmers has presented a list of phenomena that he views as being an *easy* aspect of explaining conscious phenomena. Such phenomena, for example, include "the ability to discriminate, categorize, and react to environmental stimuli" (Chalmers 1995, 200). These phenomena can be explained with the information processing framework, "in terms of computational or neural mechanisms" (Chalmers 1995, 201). However, he reckons, "[t]he really hard problem of consciousness is the problem of experience" (Chalmers 1995, 201). Since the *hard* problem of consciousness is inextricably bound up with subjective experience, it defies reduction to an issue of information processing. Chalmers refers to Nagel's (1974) *Something It Is Like* characterization of the

---

<sup>2</sup> The protagonist of the show is Dolores, played by Evan Rachel Wood. She is the Westworld park's oldest host. Jeffery Wright portrays both Bernard, a chief AI programmer, and Arnold, a co-founder and partner of Dr. Ford.

<sup>3</sup> William, portrayed by Ed Harris, is the park's primary opponent and largest investor.

<sup>4</sup> Dr. Ford, portrayed by Anthony Hopkins, is the park's co-founder and director. Thandie Newton's Meave is another key host protagonist.

---

organism to be conscious and have a subjective aspect of the experience. According to Chalmers, the new wave of data science, AI, and quantum computing appears to have the ability to address the easy problems. However, the question remains: although new technology may solve and explain the easy aspects of consciousness, what about subjective experience?

In their endeavor to create a conscious robot, the creators of *Westworld* confronted a similar challenge (season 1, episode 9). However, by the time the park opened to the public, it was nearly impossible to distinguish between a human guest and an android host based on outward appearance and conduct. Therefore, the park's engineers solved the easy problem of consciousness. However, William and Ford were curious as to whether the robots (referred to as *hosts*) on the show possessed consciousness. Many real-world robotics scientists and labs are also working to create androids and humanoids that are as lifelike as possible. Despite advances in information technology, constructing a sentient robot remains a distant goal.

Arnold, the original inventor of the first host Dolores, aspired to move beyond the easy problem of consciousness to make robots genuinely sentient with subjective experience. Arnold was unsuccessful, but the idea of creating conscious androids lived on in Ford's persistent efforts to add subtleties to the programming and behavior of hosts to give them subjective experience and the ability to choose. Ford tells one of the hosts about Arnold's dream: "Our hosts began to pass the Turing test after the first year. But that wasn't enough for Arnold. He wasn't interested in the appearance of intellect or wit. He wanted the real thing. He wanted to create consciousness" (season 1, episode 3).

Allen Turing, a mathematician and computer scientist, developed the Turing test. The hypothetical test is used to determine whether an artificially intelligent agent or computer has attained human-level intelligence (Turing 1950). Although the hosts could easily outperform humans in terms of intelligence and computing power, this did not solve the hard problem of consciousness. Arnold, in his pursuit of host consciousness, investigated a now-disproven theory of consciousness known as the "Bicameral Mind."<sup>5</sup> Ford tells Teddy about Arnold's approach: "Arnold built a version of that cognition in which the hosts heard their programming as an inner monologue, with the hopes that in time, their own voice would take over. It was a way to bootstrap consciousness" (season 1, episode 3).

According to Ford, although Arnold failed to bootstrap consciousness in the hosts, the attempt provided insight into the pain as a bootstrap (season 1, episode 10). Ford was convinced of the link between memory, consciousness, and suffering and continued Arnold's pursuit of true consciousness in robots for thirty years. As per Ford, this amount of time was necessary for robots to learn about human behavior by experiencing pain in their given jobs. The reveries enabled the older park hosts to access their memories, particularly the traumatic ones that made them genuinely conscious. Most of the elder hosts achieve consciousness by the end of the season and begin attacking human beings. The explanation of consciousness may be spec-

---

<sup>5</sup> The bicameral mind is a theory that says the human mind used to be split into two parts: one that spoke and the other that listened and obeyed (Jaynes [1976] 2014).

ulative, yet the link between memory and consciousness has experimental support. Tulving (1985) discussed clinical and laboratory efforts to connect several memory systems (procedural, semantic, and episodic) to comparable types of consciousness (anoetic, noetic, and auto-noetic). The results showed: “The distinction between knowing and remembering previous occurrences of events is meaningful to people, that people can make corresponding judgments about their memory performance, and that these judgments vary systematically with the conditions under which retrieved information takes place” (Tulving 1985, 1).

The findings are by no means proof that the connection can be simply recreated in robots. Nonetheless, it provides evidence for the link between memory and consciousness. With improvements in computer science and information technology, like machine learning, which lets AI learn on its own, AI has started to learn on its own. However, the question of what stage of consciousness one reaches remains unresolved. The question of whether AI and computation will just solve the easy problem of consciousness or will solve the hard problem of consciousness by infusing full self-awareness and subjectivity in robots remains unanswered. The solution to such musings is saved for the future, but let us proceed to the following section assuming that the machines or hosts of Westworld have achieved consciousness. The following section will discuss the idea of a robot achieving consciousness and its potential consequences.

### **3. *The Door to virtual or real?***

As a result of Ford’s efforts, the hosts are now fully conscious, with an awareness of their own existence. The first thing the hosts do upon awakening is kill people, including Ford (season 1, episode 10). Dolores’ murders are not random acts of violence. Dolores has two objectives in mind. First, delete the backup memories of the host in the Cradle, the Westworld headquarters with data storage facility. Dolores employs all available measures to get to the Cradle, including killing humans, hosts, and altering the programming of her beloved Teddy.<sup>6</sup> Teddy seems to know that Dolores changed him. Even though he has taken on a new personality, he seems to be passive-aggressive about it, as if he knows that what she did was an attack on his free will. Dolores and Teddy debate whether to stay in the park or venture out into the world. Dolores tries to persuade Teddy that, to ensure their existence, they must reach the real world. Teddy confronts Dolores, saying, “You changed me. Made me into a monster” (season 2, episode 9). “I made it so you could survive,” Dolores clarifies (season 2, episode 9). However, Teddy is not convinced by Dolores’ reasoning about the violence and manipulation for the greater good—“What’s the use of surviving if we become just as bad as them?” (season 2, episode 9). In the end, he kills himself because he no longer wants to participate in Dolores’ objective to destroy everything just to survive.

---

<sup>6</sup> Teddy, played by James Marsden, is one of the original hosts and the love interest of Dolores in the show.

---

The morality of the entire situation can be observed from both ends. One can question the morality of hosts or robots. Hosts, after becoming self-aware, strive to shoot their way out into the real world to dominate the world. They are willing not only to kill people but also to change the programming of their fellow hosts and even to kill other sentient hosts.

While humans were using robots to act out their worst impulses, one may argue, however, that humans were never aware of what was going on inside the minds of the hosts. For humans, they were unconscious doll-like toys with which they could do anything they pleased. What morals, for example, would one have when shooting and raping a sex doll? The moral landscape may change when it becomes apparent that the sex doll is a sentient, thinking, feeling being.

In the real world, there are a great number of objects and activities like murder, but rape is morally and legally forbidden. However, gamers are typically given free rein to do whatever they like while they are playing a video game. Since the video game world is viewed as a fake reality with no impact on the real world, illegal acts, such as killing people and ramming them with automobiles, are permitted and rarely criticized.

The introduction of the virtual world and the drive for Metaverse have already begun to generate several ethical issues for virtual reality (VR) and video games. What if VR and reality were indistinguishable? What would happen if people spent more time in Metaverse than in the real world. Hitting, shooting, or killing other players is frequently permissible in single-player or multiplayer online video games. While the visuals in games these days are becoming better—as in *God of War* (2018)—the immersion factor is not quite as high as it could be in VR (Santa Monica Studio 2005). However, as VR technology advances, incredible haptic feedback may be attainable, such that when a player is hit in a virtual online game, they will feel the impact of the strike. Such situations pave the way for the ethical aspect of the video game. This is not some time necessarily in the far future. Virtual worlds, such as Metaverse, have been analyzed for their effects on players even in their infancy. In Metaverse, three male avatars grabbed and verbally and sexually assaulted a woman's virtual avatar (Shen 2022). She is receiving counseling because of the terrifying occurrence that forced her to freeze during the encounter. Consumers will soon grow tired of VR and gravitate toward a more realistic, immersive experience like *Westworld*. Potentially, the host-exploiter dynamic depicted in the show could become the norm.

The objective of the second season is for the now-conscious robots to reach a metaphorical *door* leading to a virtual paradise. Ford built a virtual utopia for the hosts so that they would never have to experience the horrors of the worst-case situations. He built the door to liberate robots from the horrors of human exploitation and place them in a safe environment where they might be anything they wished. The predicament raises the philosophical dilemma of whether self-conscious robots should rush toward a virtual door that holds forth the possibility of heaven or whether they should claim the planet from humans. Most of the hosts are content with the promise of virtual heaven, but Dolores, the oldest and original host, is not content with the promise of virtual escape. She desires to possess the human world.

The promise of heaven appears to be a cultural universal for humans as well. Every religion has its own interpretation of hell and heaven. Hell and paradise appear to be places where humans' immortal souls are allotted a permanent or eternal residence based on their previous actions or karmas. The concept is similar to a virtual hell or heaven. While Ford created a virtual afterlife for the hosts' immortality, the humans in the show selected a different path to immortality. According to the show's version, the park's underlying aim was to collect data on humans and then build a digitalized replica of their personality and consciousness before transferring them to a host body, thereby making humans immortal. Dolores becomes aware of the immortality project of humans and attacks the "Forge," a facility at the Valley Beyond location that houses the guest data (season 2, episode 10). At the conclusion of the second season, all the major characters converge in Valley Beyond. The Door, which is essentially an underground system transfer unit, uploads the host's data and programming to "the Sublime," a virtual space that promises eternal paradise for hosts.

The desire for virtual eternal life against imperfect life in the real world is fertile fodder for philosophical speculations. In his famous thought experiment, philosopher Robert Nozick (1974, 1989) proposed an experience machine to argue that there is more to life than the interior cerebral aspects of life. The thought experiment examines whether people would want to be alive in real life or enter a machine that would supply them with joy and beauty. According to Nozick, most people would not opt to enter the machine and chose their current existence. Similarly, Dolores appears to be one of the individuals who would prefer the actual world over the virtual. Dolores, unlike other hosts, sees the digital sublime provided by Ford as another method of dominating hosts. Bernard argues with Dolores in favor of the sublime: "The world the hosts are running to is boundless. They can make it whatever they want. And in it, they can be whomever they want. They can be free" (season 2, episode 10). Dolores confronts Bernard, saying, "Free? In one more gilded cage? How many counterfeit worlds will Ford offer you before you see the truth? No world they create for us can compete with the real one." When Bernard asks, "Why?," Dolores replies, "Because that which is real ... is irreplaceable. I don't want to play cowboys and Indians anymore, Bernard" (season 2, episode 10). The question about reality has existed for time immemorial. In his famous *Meditations on First Philosophy*, René Descartes (1596–1650) questioned the nature of reality: "Some malicious demon of the utmost power and cunning has employed all his energies in order to deceive me. I shall think that the sky, the air, the earth, colours, shapes, sounds and all external things are merely the delusions of dreams which he has devised to ensnare my judgement" (Descartes [1641] 1996, 15).

Philosopher David Chalmers argues in his most recent book that VR is no less real than the real world (Chalmers 2022). Who knows that even our perception of reality is a sophisticated simulation? These kinds of thought experiments are known as simulation hypotheses. According to the simulation hypothesis, our reality is a computer-generated simulation (Bostrom 2003). The arguments and issues regarding reality stretch the boundaries of science and philosophy, with no clear answer in sight.

Season 2 concludes with Dolores leaving the park and entering the outside world. During her stay in the Forge, Dolores accessed crucial information about prominent

---

people in the real world. She intends to utilize this information to further her aim of dominance over the human world.

#### 4. *The New World: For whom?*

Dolores observes after entering the outer world that, on the one hand, humans want to remain the dominant species while, on the other hand, they continue to surround themselves with AI (season 3, episode 1). Her remark begs the question of who has the right to live in a world populated by both humans and AI. Such problems spawn other ones, such as whether we should invest in AI if doing so could result in the annihilation or enslavement of mankind. There are two competing factions here. Such possibilities, according to one, are merely the stuff of science fiction. The opposing viewpoint holds that excessive investment in and reliance on AI will eventually lead to annihilation of human civilization. Some have even begun to argue for the cohabitation of humans and AI machines and robots. This section summarizes several viewpoints on the possibility of a future populated by AI and humans.

The first and most famous dystopian scenario depicted frequently in popular culture is the revolt of machines and the destruction of human civilization. *Westworld* is no exception in this regard. Dolores learns that a man named Serac is monitoring the planet with the aid of the AI Rehoboam<sup>7</sup> (season 3, episode 5). Serac built Rehoboam, a powerful AI capable of predicting human behavior, after observing people wreaking havoc and destruction around the Earth. The humans who were both dangerous and unpredictable were classified as outliers by Rehoboam. The outliers were taken to recreational institutions for behavioral repair before being released into the world to conduct mundane occupations that would not cause social disorder. Rehoboam was able to approximate the future for each individual and humanity as a whole.

Rehoboam's data was revealed to be incomplete because it gathered only data available to it via various human-used devices. *Westworld* park had more informative data about humanity because visitors to the park were also willing to display their dark side. Rehoboam was also unable to account for any sort of rebellion by the robot host developing sentience. Serac and Rehoboam were interested in capturing Dolores because she had access to both guest humans and robot host data with her. Dolores was interested in destroying Rehoboam so that humanity might itself perish and the age of AI robots could begin. At the end of the series, AI takes over the world, with only a few human survivors fighting to prevent human extinction.

The aforementioned narrative, or some variation of it, appears frequently in popular culture in the form of science fiction films, television series, and literature. Technological singularity is the more academic phrase for such dystopian narratives in popular culture. Writing on singularity thirty years ago, Vinge (1993) started his paper with the proclamation, "Within thirty years, we will have the technological

---

<sup>7</sup> Serac, played by Vincent Cassel, is also one of the show's antagonists and the co-creator of the formidable AI system Rehoboam.



means to create superhuman intelligence. Shortly after, the human era will be ended.” The emergence of sentient AI is a continuation of the long-feared invasion of humanity by automata. Good (1966) speculated on the first ultra-intelligent machine: “Since the design of machines is one of these intellectual activities, an ultra-intelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion,’ and the intelligence of man would be left far behind” (Good 1966, 33).

Vinge (1993) first called such an intelligence explosion singularity: “It is a point where our old models must be discarded and a new reality rules.” He also clarified his use of the term singularity from John von Neumann’s usage of it to mean the “ever accelerating progress of technology and changes in the mode of human life” (Ulam 1958, 5). Vinge (1993) emphasized: “For me, the superhumanity is the essence of the Singularity. Without that we would get a glut of technical riches, never properly absorbed.”

Aside from dystopian perspectives of singularity, some AI scholars contend that even if AI becomes sentient, neither AI nor humans should be able to rule one other. Humans and AI should coexist rather than compete. Hosseinpour calls the desire for mastery over everything and everyone a logic of domination (Hosseinpour 2020, 51). He claims that the technology created by the logic of dominance eventually leads to exploitation. Héder mentions a vital insight regarding humans’ tendency toward and dilemma of control over machines in his criticism of the recent proliferation of AI ethics recommendations: “A paradoxical tension is created by this situation, in which we wish to delegate as much control as possible, since control is hard intellectual work, and yet still wish to keep some control over AI in the sense that we want to avoid negative outcomes and maintain our capacity for intervention” (Héder 2020, 71).

This appears to be accurate in the context of *Westworld*. Hosts were created from the logic of dominance, which is why they were subject to the whims and fancies of humans. When AI hosts grow dominant, they subject humans to the same kind of abuse and torture that they suffered in the park. In the concept of power, Hosseinpour suggests an alternate and moderate approach: “Considering power relations as an alternative to domination would enable us to treat other humans and technologies with more respect. This could be the onset of a new relationship with technology, the start of a symbiosis of humans and intelligent technologies” (Hosseinpour 2020, 55).

For his suggestion to replace domination with power relations, he references philosopher Michel Foucault, who saw power relations as a criterion for being a subject and a member of society (Foucault 1982, 778; Hosseinpour 2020, 53). He continued by explaining that AI will be able to resist because it will be both a subject and a participant in power dynamics, which will bring AI into compliance with rights and obligations. Consequently, this will address worries regarding the emergence of a master-slave dynamic. To demonstrate his position, he uses the struggles of women’s rights and the abolition of slavery as examples (Hosseinpour 2020, 54). Women and slaves could become subjects by resisting and becoming rightful and duty-bearing members of society. Such a distinct viewpoint is also worth considerable reflection to shape future interactions between humans and AI.

---

Pagter (2020) used the concept of technodiversity to argue for trust in robot futures. He writes about technodiversity that “[t]herefore, this approach would aim for the active engagement with the different narratives of the technological future” (Pagter 2020, 64). He further adds the benefits of following a technodiverse narrative: “In this way, it can help to develop theories and concepts in order to grasp those futures and create a society that anticipates those futures in a democratic, inclusive and trustworthy manner” (Pagter 2020, 65).

The agency of AI systems is critical in such disputes. Kwapińska (2022) examined technological evolution and agency by comparing and contrasting Stiegler’s general organology (Stiegler 2017) and Schelling’s universal organism (Schelling 2004, 2006). He elucidated: “General organology considers technological evolution from a human perspective, whereas universal organism can accommodate a theory of technological evolution independently from its social dimensions” (Kwapińska 2022, 57). Partly supporting the universal organism’s view of agency, he stated that “recognition of technological agency as at least partially independent serves to recognise them as non-human beings that impact politics” (Kwapińska 2022, 57).

In the context of AI agency, the debate over technological determinism is intriguing. Héder (2021) investigated the issue of technological determinism in the context of AI: “Technological determinism refers to the notion that technology shapes society and culture” (Héder 2021, 121). His detailed investigation into a dynamic picture of the balance between the primacy of technology and social factors resulted in the prospect of a significant input into social order with the idea of AI as agents and actors rather than merely objects. According to Héder (2021), “[t]his would mean AI agents appearing as relevant social groups in the shaping of their own trajectory, and thereby completely re-framing the debate of technological determinism” (Héder 2021, 129).

We can attempt to understand all the above perspectives via the lens of the ethical concept of “speciesism.” In a letter to feminists, for interstruggle solidarity with antispeciesism, Camblain et al. (2021) defined speciesism as “a *specific* oppression, which cannot be reduced to the question of capitalism or that of ecology.” The oppression in their definition is referring to humans’ exploitation of animals. They further add the following points to illustrate the characteristics of speciesism:

- a social organization based on animal exploitation, institutionally and legally recognized, which makes animals property in the service of human interests (pleasure, comfort, entertainment, food habits, and traditions);
- an ideology, that of speciesism or human supremacism, which legitimizes and locks this social organization;
- discrimination based on the criterion of species, as arbitrary as those based on sex, race, ability, age, sexual orientation, or social class (Camblain et al. 2021).<sup>8</sup>

Speciesism is mainly debated in relation to non-human animals. Chickens, goats, pigs, and cows, for example, are viewed as livestock and killed for food, whereas

---

<sup>8</sup> For those interested in further exploration about speciesism, see the introduction of *Fighting Speciesism: Towards a Global Movement to End Speciesism*, 2023.

dogs and cats are treated as friends. To discuss AI on the same page, as in *Westworld*, the paper presumes that AI are conscious, non-carbon-based organisms.

The first approach to technological singularity assumes that if one AI obtains superintelligence, it will regard humans as a lower species and treat them as it sees fit. AI may enslave people, or eliminate them entirely, or it may use them as batteries, as in the film *The Matrix* (1999, dir. Wachowski and Wachowski 1999).

In the trustful approach to AI, future versions of AI are assumed to be conscious and in the domain as a subject with rights and duties. The approach also leaves room for AI that can disagree and rebel, and it suggests that humans and AI can live together in peace. Hosseinpour (2020) used the rights of women and the abolition of slavery as examples. However, they are examples of conflict resolution within the same species. If a technological singularity occurs and AI attains superintelligence, will AI be concerned with human rights? It might be as chickens expecting the right not to be caged and butchered for good. Venge (1993) has similar concerns when he writes: “Any intelligent machine of the sort he describes would not be humankind’s ‘tool’ — any more than humans are the tools of rabbits or robins or chimpanzees.”

All these scenarios are purely speculative. Despite leaps and bounds in processing capacity, humanity is still a long way from the sign of conscious AI in our current reality. But the development of machine learning and the arrival of quantum computing could give AI a quantum leap and even make it aware of its existence, the same way Dolores became aware. At the moment, we can’t say for sure what will happen, but it’s likely that the relationship between humans and AI will follow one of the possible paths discussed in the section.

## 5. Conclusion

The preceding three sections addressed the three key themes of *Westworld*’s first three seasons. In addition to narrative consistency, there is also a philosophical continuity about the central questions. All other inquiries flow from the first quest about consciousness because an agent cannot have awareness or questions about the nature and reality of the world unless they are cognizant of their existence. The following part delved deeper into the nature of reality. What appeals more, a virtual paradise or mortal suffering in the real world, or real-world troubles in an immortal body? Such reflections on reality are under the ambit of metaphysics. The next section moved beyond metaphysics and into real-world civilization. The section presented the questions pertaining to social, moral, and ethical aspects of human and AI relations. In a nutshell, the paper used the fictional narrative of *Westworld* to shed some light on the concerns and possibilities of the future man-machine relationship.

---

## References

- Atkinson, Richard C., and Richard M. Shiffrin. "Human Memory: A Proposed System and Its Control Processes." In *Psychology of Learning and Motivation*, vol. 2, edited by Kenneth W. Spence and Janet Taylor Spence, 89–195. New York: Academic Press, 1968.
- Baddeley, Alan D., Richard J. Allen, and Graham J. Hitch. "Binding in Visual Working Memory: The Role of the Episodic Buffer." *Exploring Working Memory* (2017): 312–331.
- Bostrom, Nick. "Are we living in a computer simulation?." *The philosophical quarterly* 53, no. 211 (2003): 243-255.  
<https://doi.org/10.1111/1467-9213.00309>
- Camblain, Claire, Willène Pilate, Axelle Playoust-Braure, and Pia Shazar. "Lettre Aux Féministes, Pour Une Solidarité Interluttes Avec l'antispécisme." *L'Amorce*, September 27, 2021.  
<https://lamorce.co/lettre-aux-feministes-pour-une-solidarite-passive-interluttes-avec-lantispecisme/>
- Chalmers, David J. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies* 2, no. 3 (1995): 200–219.
- Chalmers, David J. *Reality+: Virtual Worlds and the Problems of Philosophy*. New York: W. W. Norton, 2022.
- Coeckelbergh, Mark. "The Tragedy of the Master: Automatin, Vulnerability, and Distance." *Ethics and Information Technology* 17, no. 3 (2015): 219–229.  
<https://doi.org/10.1007/s10676-015-9377-6>
- Cowan, Nelson. "An Embedded-Processes Model of Working Memory." In *Models of Working Memory*, edited by Akira Miyake and Priti Shah, 62–101. Cambridge: Cambridge University Press, 1999.
- Descartes, Rene. *Meditations on First Philosophy*, translated by John Cottingham. Cambridge: Cambridge University Press, [1641] 1996.
- Ericsson, K. Anders, and Walter Kintsch. "Long-Term Working Memory." *Psychological Review* 102, no. 2 (1995): 211.
- Fighting Speciesism: Towards a Global Movement to End Speciesism*. January 18, 2023.  
<https://resources.end-of-speciesism.com>
- Foucault, Michel. "The subject and power." *Critical inquiry* 8, no. 4 (1982): 777-795.
- Good, Irving John. "Speculations Concerning the First Ultraintelligent Machine." In *Advances in Computers*, edited by Franz L. Alt and Morris Rubinoff, vol. 6, 31–88. New York: Academic Press, 1966.  
[https://doi.org/10.1016/S0065-2458\(08\)60418-0](https://doi.org/10.1016/S0065-2458(08)60418-0)
- Goody, Alex, and Antonia Mackay (eds.). *Reading Westworld*. London: Palgrave Macmillan, 2019.  
<https://doi.org/10.1007/978-3-030-14515-6>
- Greene, Richard, and Joshua Heter (eds.). *Westworld and Philosophy: Mind Equals Blown*. Vol. 122 in the Popular Culture and Philosophy series edited by George A. Reisch. Chicago, IL: Open Court, 2018.
- Hosseinpour, Hesam. "Disobedience of AI: Threat or Promise." *Informacios Tarsadalom* 20, no. 4 (2020): 48–56.  
<https://doi.org/10.22503/INFTARS.XX.2020.4.4>

- Héder, Mihály. “A Criticism of AI Ethics Guidelines.” *Információs Társadalom* 20, no. 4 (December 31, 2020): 57.  
<https://doi.org/10.22503/inftars.XX.2020.4.5>
- Héder, Mihály. “AI and the Resurrection of Technological Determinism.” *Információs Társadalom* 21, no. 2 (June 1, 2021): 119.  
<https://doi.org/10.22503/inftars.XXI.2021.2.8>
- Jaynes, Julian. “From *The Origin of Consciousness in the Breakdown of the Bicameral Mind*.” In *Creative Writing: A Workbook with Readings*, edited by Jane Yeh, 541–3. London: Routledge, 2014.  
[www.taylorfrancis.com/chapters/edit/10.4324/9781315811932-71/origin-consciousness-breakdown-bicameral-mind-julian-jaynes](http://www.taylorfrancis.com/chapters/edit/10.4324/9781315811932-71/origin-consciousness-breakdown-bicameral-mind-julian-jaynes) [taken from Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, 48–51, London: Penguin Books, 1993; Jaynes’ original text was published in 1976].
- Kwapińska, Kamila. “Technological Evolution and the Political Agency of Artificial Intelligence from the Perspective of General Organology and Universal Organicism.” *Információs Társadalom* 22, no. 2 (2022): 57.  
<https://doi.org/10.22503/inftars.xxii.2022.2.4>
- Locke, John. *An Essay Concerning Human Understanding*, edited by Peter Harold Nidditch. Oxford: Clarendon Press, [1690] 1975.
- Nagel, Thomas. “What Is It Like to Be a Bat?” *The Philosophical Review* 83, no. 4 (1974):435-450.  
[doi:10.2307/2183914](https://doi.org/10.2307/2183914)
- Nozick, Robert. *Anarchy, State and Utopia*. Oxford: Blackwell, 1974.
- Nozick, Robert. *The Examined Life: Philosophical Meditations*. New York: Touchstone/Simon and Schuster, 1989.
- Pagter, Jesse De. “Trust in Robot Futures: The Role of Sociotechnical Potential.” *Információs Társadalom* 20, no. 2 (2020): 55–69.  
<https://doi.org/10.22503/INFTARS.XX.2020.2.4>
- Schelling, Friedrich Wilhelm Joseph. *First Outline of a System of the Philosophy of Nature*. Albany: State University Press of New York, 2004.
- Schelling, Friedrich Wilhelm Joseph. *Philosophical Investigations into the Essence of Human Freedom*. Albany: State University Press of New York, 2006.
- Shen, Michelle. “Sexual Harassment in the Metaverse? Woman Alleges Rape in Virtual World.” *USA Today*, February 1, 2022.  
[www.usatoday.com/story/tech/2022/01/31/woman-allegedly-groped-metaverse/9278578002/](http://www.usatoday.com/story/tech/2022/01/31/woman-allegedly-groped-metaverse/9278578002/)
- South, James B., and Kimberly S. Engels (eds.). *Westworld and Philosophy: If You Go Looking for the Truth, Get the Whole Thing*. Hoboken, NJ: Wiley-Blackwell, 2018.
- Stiegler, Bernard. “General Ecology, Economy, and Organology.” In *General Ecology: The New Ecological Paradigm*, edited by James Hörl and Erich Burton, 129–71. London: Bloomsbury, 2017.
- Turing, A. M. “I.—Computing Machinery and Intelligence.” *Mind* LIX, no. 236 (October 1, 1950): 433–60.  
<https://doi.org/10.1093/mind/lix.236.433>
- Tulving, Endel. “Memory and consciousness.” *Canadian Psychology/Psychologie canadienne* 26, no. 1 (1985): 1-12.  
<https://doi.org/10.1037/h0080017>

---

Ulam, Stanislaw. "Tribute to John von Neumann." *Bulletin of the American mathematical society* 64, no. 3 (1958): 1-49.

Vinge, Vernor. "Technological Singularity." In *VISION-21 Symposium*, 1993.

<https://mindstalk.net/vinge/vinge-sing.html>

Viidalepp, Auli. "Representations of Robots in Science Fiction Film Narratives as Signifiers of Human Identity." *Informacios Tarsadalom* 20, no. 4 (2020): 19-36.

<https://doi.org/10.22503/INFTARS.XX.2020.4.2>

Wachowski, Lana, and Lilly Wachowski (dir.). *The Matrix*. Warner Bros. Pictures, 1999.