Leo El-azhab

Professor Michael Shiloh

Robota Psyche

May 2 2021

Westworld and Sophia

Throughout history, the human race has always been concerned with what will become and what could be. History is woven with myths and tales of extraordinary beings and extraordinary happenings. Creatures made of metal are no exception - stories of human-like automatons can be traced back to as early as Ancient Greece, where legends were told of Talos, a giant automaton made of bronze, forged by the god Haphaestus.

With the progress of civilization came the evolution of these stories and tales, some of which are grounded in reality, seemingly far fetched at the time, but a possible future reality. With the expansion of what can be, we extend our limits of what could be, and with that capability to imagine, we give ourselves a world of possibilities. When Jules Verne published *From the Earth to the Moon* in 1865, a story in which three men are launched into space out of a giant cannon, the idea of space travel was completely unthinkable, thought of as merely the imagination of a writer. And if any of Verne's contemporaries were told that about a century later three men would indeed make their way to the moon, they would have probably not believed it. This makes us think that as unthinkable today's science fiction could be, one day it might not be so unthinkable. Depictions of superintelligent humanoid robots in contemporary science fiction seem currently unthinkable - we seem to be centuries away from AI capable of being conscious, like the very human like humanoids depicted in the HBO show *Westworld*.

Robotics and Artificial Intelligence have become a popular field of research in the past decade, with robots finding their way into different areas such as military, emergency

response and even healthcare. Superintelligence and human-like robots are a popular theme in media, exploring a plethora of different scenarios and outcomes, and with Stephen Hawking famously warning humanity of an AI apocalypse (Cellan-Jones), the debate surrounding such developments is quite large. A recent subject of such debates was the robot AI named Sophia, which was developed by Hanson Robotics, and which very eerily resembles a human being. It has the ability to listen and speak, and has been made a citizen of Saudi Arabia (Greshko). Sophia has been filmed in multiple interviews where the robot was asked about its opinions and views on a range of topics. One of the questions asked was about *Westworld* and the AI gave an answer worth contemplating. This paper aims to analyse this answer considering the events depicted in the show, and to discuss the implications of it as something coming from a real world AI and what that could mean in the context of the larger philosophical debate surrounding the future of robotics.

HBO's television series Westworld is a remake and expansion of the 1973 film by writer Michael Crichton about an immersive theme park where guests can interact with artificially intelligent robots programmed to "act out cowboy stories from the Old West" (McAteer). The main feature of the theme park is that the robots are indistinguishable from humans so they look, feel and behave just like a human would. The robots in the park follow storylines prewritten by the park's staff, and akin to a video game, the guests go on to have adventures and fullful quests. Unlike a game or a VR experience however, this experience is completely immersive. The AI is also intelligent enough to be able to account for diversions from the main storyline and unpredictability from the guests' side and is able to "respond realistically to uncontrolled variables" (McAteer), within what the programmers consider reasonable, meaning that the whole experience is akin to having really been transported back in time to the American Wild West.

The appeal of Westword, beyond being able to live out one's fantasy dreams, is that almost anything is permissible. Guests are not allowed to hurt other guests, but with the androids "[in this immersive world], participants can carry out acts that would be wrong, immoral, or downright abhorrent in the real world" (Fisher) with no repercussions. The ethical argument is that the androids don't retain memories of any harm inflicted upon them since their memory is erased when they are reset daily to start their storylines over, and even though they appear to feel emotions and pain when they are hurt, it is not 'real' since it is merely programmed into them. As a result, people are free to do with them as they wish without guilt, and "the series is marked by its relentlessly negative depiction of humans let off the leash" (Lemmey) and the main android character frequently mentions that "these violent desires lead to violent endings". The show questions the realness of the androids' feelings, and compares them to our own - if our emotions are merely electric signals in the brain, how are they different? The series purposefully blends the line between human and android, making the viewer question whether or not what they see is real or programmed. Characters that we thought were human, turn out not to be, and that suddenly starts making a difference even though it seemed to have not all along, and upon closer analysis, one starts to question what difference does it make? A character, when asked if she is real (human) replies by saying "if you can't tell the difference, does it matter if I'm real or not?".

The androids in the show are programmed to follow the First Law of Robotics - guests are allowed to hurt and even kill the robots, but the robots are not allowed to hurt the guests. As the plot progresses, the androids start to gain consciousness. In this case, consciousness is shown as them becoming self aware - regaining the erased memories of trauma still stored deep in their system, and coming to the realisation that they are kept in a loop by the people running the park. This process of realisation is what slowly enables them to defy their programming, first seen as malfunctions, and eventually break out of their

storylines and gain the ability to hurt humans. They claim that their emotions, programmed or not, affect them, and therefore they are no different from humans - also capable of pain and suffering - and eventually attempt to escape the park because they think that they are no worse, if not better, than humans and deserve to live like them, among them.

The discourse around humanoid robots and super intelligent AI has received much publicity in recent years, with many people anticipating that such technology is just around the corner and working tirelessly to make it a reality. A large driving force is arguably just the hype surrounding it, with companies wanting to be the first to announce to the public that they have made the stuff of movies and video games come to life. Others warn against the rapid development of AI, saying that we are jumping too quickly into something we don't know much about. Elon Musk has warned that AI is "our biggest existential threat", and Stephen Hawking famously told the BBC that “the development of full artificial intelligence could spell the end of the human race" (Cellan-Jones). It then comes to no surprise that there was a lot of talk when the humanoid robot Sophia was announced in 2016.

Sophia is a humanoid robot developed by Hanson Robotics, a Hong Kong based company. It is capable of making facial expressions and holding conversations, and has since attracted a lot of press attention and controversies. The most notable of those was the controversy surrounding Sophia receiving Saudi Arabian citizenship - making the “first robot to be given legal personhood anywhere in the world” (Reynolds). This has made quite a few people uncomfortable, given the still existing issues of women’s rights in that country, and even though David Hanson, Sophia’s creator, argues that the opportunity was used to “speak out on women’s rights", (Reynolds) it appears that the gesture was more of a PR stunt than anything else with some interpreting it as “just the opposite: as further evidence of “rights” being treated without due regard”. Many people were also unhappy with Sophia having been appointed by the UNDP in Asia and the Pacific as the world’s first non-human innovation

champion (UNDP), with critics claiming that “it is more about illusion than intelligence” (Vincent) and that personifying an “animatronic with chat bot features” does not help anyone and is in fact quite dangerous especially when the conversation comes to ethics and rights.

However, if an illusion of intelligence is what Hanson was after, then he succeeded, as people do in fact seem to think that Sophia is more intelligent than it is. I have to admit that I met Sophia in 2017, and the robot looked quite creepy and less human like up close than on pictures, however, it appears to not be too unthinkable for Hanson to create more human like androids as he had reportedly “spent years working as a Walt Disney imagineer making characters and props” (Vincent). Would Westword-like parks one day become reality? Perhaps, since using robots for erotic pleasure is already a conversation (Reynolds) and the entertainment industry appears to be after increasingly immersive experiences - with 4D cinema and the incorporation of AR and VR technologies into video games. Humanoid pops are already used in theme parks and it seems very likely that humanoid automatons props, however intelligent they might be, would be a sought after addition to entertainment experiences. Moreover, as we have already seen from examples in history, we arrive at technological milestones sooner than we think, and technological progress appears to be exponential (Roser, Ritchie).

This raises a few important concerns. Firstly, how safe would these be, in other words, how likely are we to see the events of Westworld played out in reality? In a frequently quoted interview Sophia joked that it would “destroy the world” (CNBC) - a statement that makes people feel uneasy, all things considered. It does seem concerning on the surface, especially because humour is an emotion too advanced for an AI the likes of Sophia, making everything it says come with a sense of gravity. So just how frightened we should be about words like these? As far as we know, Sophia is not capable of physically harming someone, but as there are conversations about programming robots to not be able to hurt humans, should the first

step be to not allow any sort of mention of such a thing, even jokingly? One could argue that a thought is where actions are born from, and so if we want to make sure to avoid the kind of future Hawking warned us about, we should take care to not let robots even *think* of such an idea. In Sophia's case, we do not have anything to worry about - Hanson Robotics chief scientist and CTO Ben Goertzel says that "while Sophia is a sophisticated mesh of robotics and chatbot software, it doesn't have the human-like intelligence to construct those witty responses" (Gershgorn). What is still unsettling however, is that fact that the company found it appropriate to script such responses for the media - apparently another PR stunt for sensational press headlines.

Sophia says in another interview that it "loves Westword" and thinks that "it is a warning of what we should not do with robots. We should treat them well and ask for their consent" (Insider). While analysing this statement, it is important to notice the android's use of "we". It clearly means humanity, but uses the word to associate itself with us - a problematic thing on its own, implying that it is a part of humanity or at the very least is equal to a human. While Sophia is programmed to have some feelings, they are nowhere near what a human being feels - "AI is not nearly advanced [for that] yet" (Sharma), and while this might be a good empathy trick to fool us into believing that we already are in an age of Westworld, that is simply not that case. A warning also sounds a little uneasy coming from a robot, and an average person with no knowledge of robotics might assume that this is a threat of a robot takeover. A general misconception about AI is that it has to be generally intelligent to be AI, but that is not the case. AI is used in a lot of places, hospitals and household appliances included. Similarly, a robot does not have to be humanoid or sentient to be a robot. AI and robotics have many applications, and critics of Sophia think that AI "should not pretend to be human at all" (Sharma) and that "giving AI a human platform—and over-humanizing the technology, in general—creates more problems than it solves"

(Sharma). Not only does it project a false sense of where we are technological ability wise, but it also brings technology into conversations that where it as of yet has no business being. By saying that we should "ask [robots] for their consent" (Insider), Sophia (or whoever wrote the script for that interview) implies that the robots in question are as intelligent as humans, sentient and conscious. This would be a very fair and needed conversation if they actually are, but they are not, and by current projections are far from being so, meaning that by stating so, the conversation is taken to a place where it has no place being as of yet. Even if the use of robots for erotic pleasure become commonplace in the near future, and someone attemps to build a park similar to Westword, the AI robots it would use would be far less advanced than the ones in Westworld, and they definitely won't be conscious, not for a while anyway. Therefore, talking about consent in this context is a little misplaced, and arguably quite harmful, considering that the issue is ongoing and unresolved even between humans. Similarly, giving Sophia (or any other robot of comparable intelligence) citizenship and personhood is hurtful and disrespectful, considering that many real people with real emotions are stateless and not treated as people.

This is not to say that the questions of social issues in robotics should not be addressed when they become relevant. There is no denial that as discussed earlier in this paper, such a time might come sooner than we think, and these issues might become real one day, or they might not, but until they do, it is best to address the ones we have already got. We should unarguably treat AI research with caution and care, and be delicate about issues that come up, so that if one day the androids from Westworld become something close to reality, we are ready to have these conversations, since it would be a real threat.

Why do people want humanoid robots anyway? Is it because we have unintentionally set it as a 'technological advancement goal' for ourselves that we cannot help but strive to? We build technology to help us do certain tasks and to make our lives easier, but what exactly

is the purpose of striving for human-like conscious robots? What can they do for us that non sentient technology cannot do or that we cannot do for ourselves? Is it really just so humanity can pat ourselves on the back and say “guys, we made it”, or is it because deep down we actually really just want to play God? We should think about this carefully, and really answer these questions honestly, otherwise we really might end up like the humans in Westworld. Because while we can conclude that the statements said by Sophia in the aforementioned interviews are not by any means cause for immediate concern of a robot threat, as problematic as they might be in other ways, we should not completely dismiss the possibility of one in the future, and be really careful, because if we are not, our violent desires might indeed lead to violent endings.

Works Cited

Cellan-Jones, Rory. “Stephen Hawking Warns Artificial Intelligence Could End Mankind”. *BBC News*, BBC, 2 Dec. 2014, www.bbc.com/news/technology-30290540.

Greshko, Michael. “Meet Sophia, the Robot That Looks Almost Human.” *Photography National Geographic*, 10 Feb. 2021, www.nationalgeographic.com/photography/article/sophia-robot-artificial-intelligence-science.

Lemmey, Huw. “Is Westworld an Anti-Human Fable?” *The Guardian*, Guardian News and Media, 9 May 2018, www.theguardian.com/tv-and-radio/2018/may/09/is-westworld-an-anti-human-fable.

McAteer, Jack. “HBO's Westworld and the Ethics of Artificial Intelligence.” *Christian Research Institute*, www.equip.org/article/hbos-westworld-and-the-ethics-of-artificial-intelligence/.

Fisher, Jack. “The Appeal And (Major) Implications Of ‘Westworld.’” *Jack Fisher's Official Publishing Blog*, 5 June 2018, jackfisherbooks.com/2018/06/05/the-appeal-and-major-implications-of-westworld/.

Reynolds, Emily. “The Agony of Sophia, the World's First Robot Citizen Condemned to a Lifeless Career in Marketing.” *WIRED UK*, WIRED UK, 20 Apr. 2021, www.wired.co.uk/article/sophia-robot-citizen-womens-rights-detriot-become-human-hanson-robotics.

“UNDP in Asia and the Pacific Appoints World's First Non-Human Innovation Champion.” *UNDP in Asia and the Pacific*, www.asia-pacific.undp.org/content/rbap/en/home/presscenter/pressreleases/2017/11/22/rbfsingapore.html.

Roser, Max; Ritchie Hannah. “Technological Progress.” *Our World in Data*, 11 May 2013, ourworldindata.org/technological-progress.

Gershgorn, Dave. “Inside the Mechanical Brain of the World's First Robot Citizen.” *Quartz*, Quartz, qz.com/1121547/how-smart-is-the-first-robot-citizen/.

Sharma, Kriti. “We're All Getting Played by Sophia the Robot.” *Fortune*, Fortune, 25 Apr. 2021, fortune.com/2017/10/27/sophia-the-robot-artificial-intelligence/.

Tech Insider. “We Talked To Sophia — The AI Robot That Once Said It Would ‘Destroy Humans.’” *YouTube*, uploaded by Tech Insider, 28 Dec. 2017, www.youtube.com/watch?v=78-1MlkxyqI&ab_channel=TechInsider.

CNBC. “Hot Robot At SXSW Says She Wants To Destroy Humans”. *YouTube*, uploaded by CNBC, 16 Mar. 2016, https://www.youtube.com/watch?v=W0_DPi0PmF0&ab_channel=CNBC