Maximillian V. Phillips

Dr. Jennifer Keating

ENGCMP 0425: Digital Humanity

2 December 2021

All's Fair in Robot War

Throughout the course of Digital Humanity, it has become abundantly clear that a great deal of care must be taken in deciding and distributing responsibility in society's technological advancements. The parties that create and use these innovations both hold a great deal of power, so the actions of technological agents must follow some sort of defined guideline to point fingers in the case of accidents, atrocities, and all other use case scenarios with severe implications on humanity. One piece of particular note on this topic was Lucy Suchman's *Wishful Mnemonics and Autonomous Killing Machines*, especially the section "Anthropomorphism and wishful mnemonics in AI." In the piece, Suchman details the naming schemes given to technology by its creators: "wishful mnemonics" that present a false quality of agency. This is most apparent in daily life in virtual assistants such as Siri and Alexa, whose human names and voices divorce these artificial intelligence systems from their algorithmic bases and ascribe a sense of independent decision-making separate from any individual companies or creators; when Siri "mishears" us, we get mad at "her." This terminology anthropomorphizes these systems, casting the creators' responsibility solely on the systems themselves, rather than the responsible parties of origin. This draws parallels with the science fiction canon's archetype of the "rogue AI," as seen in films such as in *2001: A Space Odyssey.* In the film, the ship's onboard computer HAL attempts to kill the entire crew to carry out its mission as programmed, only to be shut down while singing. Despite taking place in a fictional future, the idea of HAL declining to open the

pod bay doors out of its own volition reflects our society's greater casting of responsibility for the actions of the technology we create upon the technology itself.

While the casting of responsibility occurs across technologies, Suchman's main concern is the "legal and political discussion about enabling autonomous lethal machines" (17). If we are able to so easily shift the blame onto trivial systems of autonomy and artificial intelligence, the possibility of shifting the full blame onto autonomous killing machines specifically created to replace humanity in war is deeply concerning. Suchman urges for the recognition of the underlying mechanisms of lethal machines, avoiding misnomers that placate the true nature of the technology. All autonomous technology should be addressed as it is and ascribed to who it should be, but if autonomous war machines are an inevitability, recognition of the "true nature" of military systems should be top priority. We cannot expect these systems to act "humanely" (as paradoxical as that is in war) if we do not know the humans that gave them the ability to "think" like they do. To even consider these machines "humane," however, creates its own plethora of responsibility issues: because these systems inherently cannot act "humane" (as in *like humanity*, as defined in *Keywords for Today*), the use of the term *requires* a human layer of accountability to place blame in military applications. As much as the language used by pro-autonomous weapon warmongers creates an illusion of human-like agency in their war machines, these wishful mnemonics cannot avoid the fact that these machines cannot and will not "think" for themselves; responsibility lies solely with humanity, despite mechanical action.

The idea of technological anthropomorphism and Suchman's call to avoid it in order to disambiguate responsibility present a challenge to humanity coded into its very nature. In other words, if ascribing human characteristics to technology is to be avoided, our own human characteristics make it very difficult to do so. Neuroscience and psychology tell us of

"pareidolia," the phenomenon of seeing patterns in noise, such as human faces in inanimate objects. David Danks in "Thinking Through Autonomy" states that our language lacks specific words for what computers do: computer vision systems are not "seeing" objects and artificial intelligence does not "learn." In this way, the language we co-opt from biological systems to describe mechanical systems is ultimately limiting in how we understand the "chain of command" in artificial intelligence decision-making. In the future, I hope that such linguistic misattributions can be avoided. Furthermore, a new set of terms specific to technological systems (and, by extension, inapplicable to biological systems) would be helpful in avoiding any confusion that these systems are separate from humanity but not independent.

As technology moves forward and our society becomes more deeply entrenched in it, our language still tends to anthropomorphize devices; while writing this essay, I had to make careful considerations to avoid assigning gender pronouns to systems like Siri or HAL to avoid affording them constructs from the human world. Despite the touchy language, the actual disambiguation of machine from human ironically becomes *more easy* as they become "more human," approaching the inflection point of the uncanny valley in which there is something subtly wrong with each simulacra. This notion of uncanniness paired with a "virtual assistant fatigue" from constant inundation gives hope that society most likely will not suspend their disbelief in truly autonomous machines: there is *always* a responsible human party. If we can extend this rejection beyond our everyday lives, we can say no to the autonomous killing machines Suchman is so wary of.

Works Cited

Dunlap, Ken. *Why Ethical A.I. Needs a Focus on the Fundamentals w/ David Danks*. https://soundcloud.com/thinkingthroughautonomy/why-ethical-ai-needs-a-focus-on-the-fundamentals-david-danks. Accessed 5 Dec. 2021.

MacCabe, Colin, and Holly Yanacek, editors. *Keywords for Today: A 21st Century Vocabulary: The Keywords Project*. Oxford University Press, 2018.

Sharkey, Noel, and Lucy Suchman. "Wishful Mnemonics and Autonomous Killing Machines." *Proceedings of the AISB*, vol. 136, May 2013, pp. 14–22.