Ultron: A Cautionary Tale of Artificial General Intelligence

Amanda Uccello

Blythe Academy

IDC4U1: Interdisciplinary Studies: Artificial Intelligence

Mr. Rogers

September 1, 2024

Sci Fi authors have constantly played with the consequences of a world where the human mind and a machine's hardware were one and the same. However, a conversation about fiction is now a real debate among scientists as artificial intelligence rapidly progresses. Marvel's sci-fi villain Ultron, in the movie *Age of Ultron*, is a prime example of artificial intelligence levels and the myriad of issues that revolve around them. Ultron is a strong AI that would pass Turing's test as its intelligence is well into the categorization of general but lacks the features required of a super intelligence.

Creation and Backstory

Ultron was created by Tony Stark and Bruce Banner as an aid to the Avengers on an international scale. They hoped to create a robot that could help them achieve world peace after seeing the fatal dangers that came with saving the world from an attack led by Loki. The two scientists took the Tesseract, a cube shaped entity containing immense power, and injected it into Ultron's code. In addition Ultron had access to all of humanity's data to learn from and access to Stark's military weapons. This is comparable to how Turing described the creation of learning machines could occur. Learning machines required an initial state of mind and a set of data to learn from. Additionally Turing reasoned that to land on the correct initial state of mind required an evolutionary process of trial and judgment. The injection of the Tesseract with Ultron's code was another iteration of the initial state of mind which combined with his training through the entirety of human information made an intelligence machine its creators would soon regret. Ultron ultimately concluded humanity in nature is destructive and the only way to achieve peace is by removing humans.

Throughout the movie, Ultron compares himself to two opposing characters, his creator Tony Stark, and another artificial intelligence, Vision. Tony Stark and Ultron ultimately have similar types of goals. Tony wants to protect humanity while Utron wants to protect the Earth by resetting it. However, their methodologies to achieve their goals contrast. Tony has the ethical primacy of taking a human life built in while Ultron does not. Ultron can commit human atrocities without any emotional consequence because he lacks empathy. However, Vision is another character programmed from the same code as Ultron. Despite this, Vision has a human centric empathy Ultron lacks. He is on the side of life and believes a flawed humanity is still worth protecting. Vision explains this to Ultron in his final scene together, "But a thing isn't beautiful because it doesn't last". (Whedon, 2015. 2:30:35). Both Tony and Vision are a reflection of Ultron and shows how destructive AI can be without the ethical restrictions humans believe are primary

Intelligence

The Turing Test, created by Alan Turing, was designed to differentiate between a human and a machine through an imitation game (Turning, 1950). To pass the test, the evaluator should not be able to tell the difference between the responses for a five minute period from both machine and human subjects thirty percent of the time. If Ultron were to be the machine in the Turing Test, he would easily pass. First, Ultron has an advanced understanding of languages and communication. In the movie Ultron was able to understand and reply with nuances, sarcasm, and jokes. In addition, Ultron can hold complex conversations without any difficulty. The best example of this was how he was able to debate abstract topics about evolution and peace with

Vision (Whedon, 2015. 2:30:35). This conversation shows how Ultron has a moral compass and ideology although different from humans. Ultron can also learn and adapt as he gains new information. Throughout the movie, Ultron modifies his plans in real time when the Avengers fight back. In these high pressure situations, Ultron can analyze situations and determine his response. The combination of all these attributes would make it difficult to determine Ultron's response as coming from a machine or man.

Another perspective to consider is the *Chinese Room Argument.* It suggests there are further aspects to intelligence beyond what can be encapsulated in the Turing Test (Searle, 1980). Searle described it as "consciousness and intentionality" (Searle, 1980). Intentionality, or purpose of action, is clear but consciousness needs clarification. There are another two concepts that form the basis of consciousness (Yolles, 2022). Sentience, or the ability to have feelings, is often considered the line that would bring with it a moral question when reprogramming or turning off a machine that possesses it. Sapience is the ability to apply common sense and insight (Yolles, 2022). This trait in humans is often referred to as wisdom. While more complex than the sum of the two, consciousness can be thought of the intermingling of both (Yolles, 2022). Decisions are made on two fronts. One with feelings directed both inward like pain or outward like empathy and anger. Two, wisdom of common sense. Utlron's whole mission can be understood as an exhibition of consciousness. After consuming all of human history Ultron responded with "Too much. Oh no, No. Yes. … I'm here to help" (Whedon, 2015. 2:30:35). At that moment it made a conscious judgment, involving feelings and its wisdom, to eliminate all of humanity.

Under Turing or Searle's perspective Ultron should definitely be considered as passing the bar for strong AI. A weak AI is only able to perform singular human tasks while a strong AI can think and act with the full intricacy of human intelligence (Glover, 2024). Throughout the movie, Ultron demonstrated his complexity by learning and applying his new information to a wide range of tasks. For example, in the movie, Ultron devises and executes a muti-facted plan by creating a massive meteor to destroy the human race. This plan alone requires strategic decision making and a cognitive ability at least at a human level. Tony Stark concurred with this, "He's a machine with a brain. He's been given our greatest strength, but he's also been given our greatest weakness: the ability to reason and evolve" (Whedon, 2015).

Beyond Turing to Super Intelligence

While all measures of intelligence considered so far use human intelligence in some capacity as the high watermark. We could also consider intelligences that are beyond what humans are capable of. This is referred to as Artificial Super Intelligence "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest" (Bostrom, 2014). Ultron has the ability to process all of human history in seconds, reason and come to a conclusion. This is something humans can not achieve. However, despite unquestionably exceeding in some areas, the biggest argument for Ultron not being a super intelligence is that the Avengers were able to understand and ultimately foil its plan (Whedon, 2015). An ASI would presumably be able to operate on a level that humans simply could not understand similar to an ant's obliviousness to the complex world humans exist in despite living in the same environment.

      Ultron is ultimately a work from science fiction that provides a warning to expand AI with care and safety in mind. An AGI alone brings with it concerns as there is no guarantee it would possess the same value system humans do. An ASI would be even more concerning to humans. While currently unthinkable, the process of creating learning machines through evolutionary learning through generation makes the potential of it occurring possible. After all this is exactly how the current dominant intelligence on Earth came into being. The difference would be that the machine iteration would not require the time that biological evolution requires. This serves as both the ultimate promise and fear of AI.

References

Bostrom, Nick (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Glover, E. (2024). *Strong AI vs. Weak Ai: What's the difference?* Strong AI vs. Weak AI: What's the Difference?
https://builtin.com/artificial-intelligence/strong-ai-weak-ai#:~:text=Weak%20AI%2C%20also%20called%20narrow,don't%20actually%20exist%20yet.

Searle, J., 1980, 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3: 417–57
http://cogprints.org/7150/1/10.1.1.83.5248.pdf

Turing, A. M. (1950). I.—Computing Machinery and intelligence. *Mind*, *LIX*(236), 433–460.
https://doi.org/10.1093/mind/lix.236.433

Whedon, J. (Director). (2015). *Avengers: Age of Ultron.* [Film]. Marvel Studios. Marvel. (n.d). *Marvel Characters: Ultron. https://www.marvel.com/characters/ultron/on-screen*

Yolles, M. (2022). Consciousness, Sapience and Sentience—a Metacybernetic view. *Systems*, *10*(6), 254. https://doi.org/10.3390/systems10060254