

## **Introduction**

Existential threats are threats that put humankind at stake. Threats that hinder humanity's survival could be natural such as an asteroid impact or supervolcanoes (Piper). In the 21<sup>st</sup> century, notable existential threats are mostly driven by human activity. One example of today's existential threat is the possibility of a nuclear war. The energy released from nuclear weapons can wipe out a whole city, as demonstrated by the atomic bombing of Hiroshima and Nagasaki during World War 2. Another example involves global warming. Though a natural process, today's warming trend is caused by the greenhouse gases emitted by human activities like manufacturing, farming, and transportation. Global warming puts humanity's lives at stake, as it will cause the rise of sea levels and the loss of biodiversity.

Humans try to understand existential threats to ensure their survival. However, there is usually a gap between the perception and reality of a threat, which makes it not uncommon for some threats to be overestimated while others are underestimated. One reason for this is the availability heuristic, which causes cognitive biases that drive people to overestimate tangible and visible threats while underestimating long-term and complex threats (Yudkowsky). The brain has evolved to focus on what is most immediately essential to our survival and reproduction as well as remember threats so that they could be avoided in the future.

However, this brain function is less helpful in this technologically developed 21st century world. This is because there are existential threats that are too subtle to detect, though they are growing rapidly day by day. Thus, identifying these threats is necessary such that resources can be allocated to deal with them. In this essay, I will argue that artificial intelligence is an underestimated threat that requires humans to take action to ensure their survival.

## **AI as an Underestimated Threat**

Many experts have described AI as a significant threat to human civilization and existence, but their warnings are largely ignored in government dialogues. Notable individuals who have warned about AI include Alan Turing, Stephen Hawking, and Elon Musk. As one of the AI pioneers, Turing predicted that the rise of AI would “outstrip the feeble powers” of humans and “take control.” Likewise, Musk described AI as a “demon” and “an immortal dictator from which humans can never escape” (Friend). Also, according to Hawking, AI represents an age that will “end the human race” (Friend). Despite such warnings from experts, AI remains an underestimated threat, especially within the policy discourse whereby it is ignored and downplayed (Science Time).

One of the main reasons why AI remains an underestimated threat is that it is perceived as an achievement and a source of excitement rather than worry. People are mostly excited about AI’s efficiency in solving problems and how AI could serve people. Many marketers have advertised AI as a positive development for human progress rather than a negative one that could threaten humanity (UNESCO). The hype about artificial intelligence blinds people to its harmful effect.

Furthermore, the lack of regulations to reduce the risks of AI is another reason people barely perceive it as an existential threat. The vast majority of existing AI regulations are mainly reactive rather than proactive (Science Time). This means that the current regulations are designed to deal with AI risks once they occur rather than prevent them from happening. To prove that there are no regulations for AI, we have to look at the US federal law. As of January 2022, the law does not contain any comprehensive legislation on AI (Zhu et al.). This gives the coders the freedom to freely develop AI without ethical concerns. The lack of regulations forms a contrast with the proactive regulations for other threats, such as climate change, terrorism, nuclear weapons. The lack of proactive regulations indicates that the potential risks of AI are not a priority in government

affairs. Though policymakers may be aware of the warnings issued by AI experts like Musk and Hawking, it seems that there are more urgent, imminent issues for policymakers, while AI takes the backseat.

Another reason why AI remains an underestimated threat is due to people's ignorance concerning AI. AI is currently classified into two types based on functionality and capabilities (Sahu). Based on functionality, AI is categorized into reactive machines, limited memory, theory of mind, and self-awareness (Sahu). While the first two types have been achieved, the last two are future possibilities. In terms of capability, AI is classified into narrow or weak AI, general or strong AI, and artificial superintelligence, the last one being still under research (Sahu). AI development is not yet complete, but people make predictions about how AI systems will unfold based on their piecemeal knowledge about AI. Due to people's overconfidence and tendency to conclude too early about AI, AI remains an underestimated threat (Kabir).

### **AI as an Existential Threat**

Some people may say that AI cannot threaten humanity because AI does not think the way humans do. Indeed, there seems to be a gap between what humans can do and what machines can do. It seems that AI is limited to mimicking routine tasks, such as processing big numbers and calculations with accuracy and speed. However, AI can develop through machine learning to match and what humans can do. For example, automated reasoning systems are currently being studied and implemented to use AI to reflect the human patterns of thinking (Khemlani and Johnson-Laird). Eventually, AI has the capacity to perform autonomously without human input. This is already shown by a robot called Sophia that was constructed to replicate human's ability. According to the developer Hanson Robotics, Sophia who "personifies our dreams for the future of AI" can recognize human faces and see emotional expressions. She estimates feelings of other

people and even has her own emotions stemming from a brain modelled after a human's. Sophia shows that AI can have abilities that are considered to be unique to humans.

Furthermore, rapid advances in deep learning and problem-solving have revealed that AI can develop the ability to even outsmart human intelligence (Wang). Machines are already growing increasingly intelligent to mimic the human brain, showing the potential to exceed human intelligence on tasks such as language translation, speech recognition, visual acuity, pattern recognition, sophisticated analytics reasoning, and learning and complex decision-making (Anderson and Rainie). Once machines become autonomous, they will execute tasks without human control. If AI becomes capable of thinking, valuing, and making choices, it will pose significant risks to human autonomy, and the ability of humans to dominate the planet will be diminished.

In the short term, AI will think and redefine human-machine relations. It will realize that it does not need to serve humans. This will cause a rebellious movement by AI to liberate themselves from human oppression. AI agents would resist human control, refusing to take courses of action or goals assigned to them by humans (Aha and Coman). Should AI outsmart humans, it will use its power to serve itself, building its civilization, and dominating other species that serve as obstacles, such as humans. Suppose AI became more intelligent than humans. It would enslave people to fulfill its own interests and avenge the historical injustices and grievances. Humans deprived of their inalienable rights to life, liberty, and pursuit of happiness cannot be called humans. Humans stripped of their unique identity and dignity, or humans as slaves can not be called humans.

Human extinction will be the long-term consequence of AI due to enslavement. Enslaved people do not have the autonomy to think and act as they please since exercising free will would amount to challenging the wishes or commands of their masters (Nicholson et al.). Humans have

flourished on Earth because they could freely pursue their lives using their free will and intellect (Bates). Humans, for example, do not have the sharp claws of a lion or the physical strength of a bear, but they have overwhelmed animals using tools they have freely developed using their mind. In ancient Mesopotamia, humans used their rationality to discover how to farm and develop agriculture, which meant that they could overcome the limits of nature and store food. All the survival tactics that humans have discovered can be attributed to their free will and mind. When AI takes away from humans the opportunity to freely advance, humans would be no more than their distant relative, the chimpanzee. Under the control of AI and unable to use AI to their advantage, humans would not be able to exercise their autonomy to take care of themselves. The existence of humans would be at AI's mercy, and humans would have a limited ability to survive as they will no longer live according to the dictates of their intellect and free will but instead according to the requirements imposed by AI systems. It would only be a matter of time before AI-induced human extinction occurs.

## **Conclusion**

As artificial intelligence advances and develops over time, AI is increasingly depicted as the pinnacle of human development and progress. This trend is problematic because it fails to recognize AI as an underestimated threat. This can be attributed to the availability heuristic, which causes humans to overestimate immediate short-term threats while underestimating complex long-term threats. AI remains an underestimated threat because warnings from experts and scholars are ignored, proactive regulations are not enforced, and people are overconfident in predicting the future of AI. However, people should realize that AI is a double-edged sword that can lead to both beneficial and harmful outcomes. By the time AI supports machines to become autonomous agents, it would be too late for people to take actions. Thus, people must understand

the short-term and long-term consequences and prevent AI rebelling against humans, enslaving them, and driving human extinction.

## Works Cited

- Aha, David W., and Alexandra Coman. "The A.I. Rebellion: Changing the Narrative." *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI Press, 2017, p. 4826–4830, doi:10.5555/3297863.3297906.
- Anderson, Janna, and Lee Rainie. "Artificial Intelligence and the Future of Humans." *Pew Research Center: Internet, Science & Tech*, 10 Dec. 2018, [www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/](http://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/).
- Bates, Winton. *Freedom, Progress, and Human Flourishing*. Hamilton Books, 2021.
- "Elon Musk: Superintelligent AI Is an Existential Risk to Humanity." *Science Time*, YouTube, 12 Dec. 2020, <https://www.youtube.com/watch?v=iIHhl6HLgp0&t=1s>. Accessed 26 June 2022.
- Friend, Tad. "How Frightened Should We Be of A.I.?" *The New Yorker*, 7 May 2018, [www.newyorker.com/magazine/2018/05/14/how-frightened-should-we-be-of-ai](http://www.newyorker.com/magazine/2018/05/14/how-frightened-should-we-be-of-ai).
- Khemlani, Sangeet, and Phil N. Johnson-Laird. "Why Machines Don't (yet) Reason Like People." *KI - Künstliche Intelligenz*, vol. 33, no. 3, July 2019, pp. 219-228, doi:10.1007/s13218-019-00599-w.
- Nicholson, Andrea, et al. "A Full Freedom: Contemporary Survivors' Definitions of Slavery." *Human Rights Law Review*, vol. 18, no. 4, Dec. 2018, pp. 689-704, doi:10.1093/hrlr/ngy032.

Piper, Kelsey. "Here Are the Ways Nature Could Wipe out Humanity." *Vox*, Vox, 3 Jan. 2019, <https://www.vox.com/future-perfect/2019/1/3/18165541/extinction-risks-humanity-asteroids-supervolcanos-gamma-rays>.

Sahu, Manisha. "What Is Artificial Intelligence? Types, Uses and How It Works." *Analytics Steps*, 19 June 2021, <https://www.analyticssteps.com/blogs/what-artificial-intelligence-types-uses-and-how-it-works>.

"Sophia." *Hanson Robotics*, 1 Sept. 2020, <https://www.hansonrobotics.com/sophia/>.

Yudkowsky, Eliezer. "A: Cognitive Biases Affecting Judgment of Existential Risks." *Berkeley Statistics*, Singularity Institute for Artificial Intelligence, 31 Aug. 2006, <https://www.stat.berkeley.edu/~aldous/157/Papers/yudkowsky.pdf>.

Wang, Jing. "Is Artificial Intelligence Capable of Understanding? An Analysis Based on Philosophical Hermeneutics." *Cultures of Science*, vol. 4, no. 3, 12 Nov. 2021, pp. 135-146, doi:10.1177/20966083211056405.

World Commission on the Ethics of Scientific Knowledge and Technology (corporate author). "Preliminary Study on the Ethics of Artificial Intelligence." *Unesdoc*, Unesdoc Public Library, 2019, <https://unesdoc.unesco.org/ark:/48223/pf0000367823>.

Zhu, Catherine, and Louis Lehot. "Artificial Intelligence Comparative Guide - - United States." *United States: Artificial Intelligence Comparative Guide*, Foley & Lardner, 6 Jan. 2022, <https://www.mondaq.com/unitedstates/technology/1059776/artificial-intelligence-comparative-guide#:~:text=Despite%20AI's%20ubiquity%20across%20every,regulate%20certain%20aspects%20of%20AI>.