Humanity Concerns of AI Nanny (by proving differences between AI Nanny and Government), the Protection of Mankind in Singularity

Yanfu Ding[1,*]

[1] *Crespi Carmelite High School*
[*] *Corresponding author. Email: dingda@celts.crespi.org*

ABSTRACT
AI Nanny was proposed by Ben Goertzel to delay the singularity of technology to protect humanity. Its most important purpose is to maintain the interests of mankind. The importance of hypothetical AI Nanny is well-presented by Goertzel because technical singularity will gradually lead humanity to decline. However, Goertzel drew a parallel between government and AI Nanny by claiming they are no qualitative differences between the two. After reconstructing and reorganizing Chalmers and Goertzel's work of technical singularity and AI Nanny, the reasoning clearly illustrated there are differences between government and AI Nanny. Furthermore, the differences should be noticed before humans considered using AI Nanny in the future to prevent singularity. From the perspective of privacy, algorithmic errors, and the common interest of humans, it will not stably bring us benefits. Its concept is not easy to realize in society, because it needs to monitor our society in all aspects. This will also bring greater danger to people's privacy.

Keywords: *"Singularity,""AI Nanny," "humanity issues,""future technology,"*

1. INTRODUCTION

In this volatile era, people are eager to pursue the endless convenience and enjoyment that technology brings to life. Will people one day lose control of the earth and the universe? At that time, people may realize that the artificial intelligence and many technologies we created are destructions to ourselves. More specifically, the future point at which artificial intelligence exceeds human intelligence and then continues to improve itself exponentially, soon reaching such a superhuman level of intelligence that humans cannot even imagine what it would be like.

Obviously, the second door is where we stepped in, but what is waiting for us is the technological singularity.

The general setting of my investigation is that technical singularity will subvert the stable development of our current society. The future point at which artificial intelligence exceeds human intelligence and then continues to improve itself exponentially, soon reaching such a superhuman level of intelligence that humans cannot even imagine what it would be like. For example, an ant couldn't possibly fantom human-level thinking with its ideas, thoughts, emotions, and feelings. If it encounters a human being, people might not necessarily want to hurt them by purpose. However, if the ants were in the way of human progress or improvement, we annihilate them without hesitation. The ants here are considered too insignificant to people. Therefore, humans should be concerned whether the future exponentially improved technologies or robots would treat us by the same logic.

This future we anticipated could be catastrophic, however, it may arrive at any specific time period in the future. Humans do not know when the singularity will come, and people cannot be prepared to adapt to it. According to Chalmers, the singularity scientist, the best choice people can make is to "integrate" into this society, but it will change our existing intelligence [1].

Therefore, what human beings should truly understand and clarify is if there is a human-friendly singularity that potentially that helps humanities develop or maintain the present situation, then imposing an encouragement on its development is needed. Not only the device helps and protect humanity, but human beings could also control it to achieve new goals.

However, Ben Goertzel, who brought up the concept of "AI Nanny", has mentioned in his work that AI Nanny

is not the ideal choice of human beings to rely on and is very problematic in post-singularity era. Therefore, scientists know that AI Nanny would be very problematic. Secondly, AI Nanny's works are not totally transparent to human beings which could be an issue in the future. In addition, while the technologies and societies are still under the control of humans, potential problems should be overviewed before realizing AI Nanny in the future.

This article hopes to clarify that humans need to know that there is a big difference between a government and a friendly singularity (AI Nanny). After understanding this difference, human beings will find that we do not favor AI Nanny very much because what it will bring us is not the perfect situation that humans imagined. While preventing the technological starting point from corroding human nature, it has made our society lose too much value and diversity, which the existence of a government composed of humans will never give up.

The significance of this research is that this decision will affect the direction of our future technology. When we already realized that the arrival of the singularity will cause a huge blow to mankind, we obviously have to calmly think about our countermeasures to the dangerous situation. If AI Nanny is going to become a necessity for society one day, then the detailed analysis of this paper will help people analyze the difference between it and the government, so as to help people see that this is quite different from our present life (negatively) and not something that humans can rely on.

2. RECONSTRUCTION

The term "singularity" is a concept proposed by Vernor Vinge in 1983. Chalmers defines it as the new superintelligence in the era that will continue to upgrade itself and make technological progress at an increasingly faster rate. Chalmers summarized a process of AI "evolution" based on the research of many scholars. In this format, AI will begin to expand and develop step by step.

Premise 1: There will be AI.

Premise 2: If there is AI, there will be AI+.

Premise 3: If there is an AI+, there will be A++.

Conclusion: there will be A++ (which means singularity will occur) [1].

AI is the current intelligence of human beings. AI+ is higher than human intelligence. AI++ is a higher level of intelligence produced by a higher intelligence than humans, which directly brings scientific singularities. Here, we need to introduce intelligence explosion and absent defeaters. The intelligence explosion assumes that technology can improve significantly upon human intelligence by creating an AI that's considerably smarter than the smartest humans. Chalmers suggested "defeaters" which refer to factors that prevent the technical improvement to singularity [1]. After facing many doubts about the singularities of science, Chalmers had to explain his premise. Some people think that the algorithm of the brain and thought is not reproducible, because it has no rules. However, Chalmers argued that their behavior can be used to measure intelligence. If our intelligence has evolved, then intelligent systems have the ability to learn to reach human intelligence. Through the advancement of the human level and the discovery of technology, AI+ was born. AI+ is more powerful than humans in designing and improving its own intelligence, which brings some expected results.

2.1. Accelerating Change (Speed Explosion)

This situation refers to the rapid acceleration of technological progress. Although human beings do not define intelligence with real measurements or dimensions, singularity scientists still believe that it is easier for AI to progress independently to construct smarter AI than to build AI that is smarter and can be integrated into the human brain and nervous system. We can no longer use the current speed of technological development (AI era) to predict the speed of technological development and technological revolution brought about by AI+ because there will always be more and more intelligent intelligence.

2.2. Intelligence Explosion

This result is an era of smarter and smarter intelligence because technology has far surpassed human intelligence.

2.3. The Event Horizon

When the singularity gradually comes, not all social and technological progress will be created by the human brain in the future. If we only think from the perspective of humans, we simply cannot anticipate or even construct many superintelligences.

At this time, humans gradually discovered that singularity has brought many threats. Unless humans integrate ourselves into a higher level of intelligence (become super intelligent machines), we will only be left with delayed technology or extinction. Obviously, not many people could agree these are good choices [2].

When people realized the crisis brought about by the singularity of science, Ben Goertzel proposed an idea based on scientists such as Chalmers, AI Nanny. AI Nanny, as an Artificial General Intelligence (or a Friendly AI), can be used to delay the arrival of singularities, thereby protecting humanity from danger. Ben Goertzel put forward the three most important conditions created by AI Nanny [4]. First, we need to

realize superhuman general artificial intelligence. Secondly, we need Nanny AI to obtain the relevant global surveillance network. Finally, AI Nanny needs the ultimate control of all robots Although these theories seem very convincing, Goertzel himself considers objection against AI Nanny.

Human-level Intelligence represents that machines or technology can replicate human behavior, where they are considered by Ray Kurzweil to arrive via human-brain emulation. However, he believes that AI Nanny belongs to a kind of Super Intelligence, that is to say, a higher level of intelligence (ability and thought) than all aspects of the human brain [8]. Although AI Nanny can achieve superhuman general artificial intelligence in a seemingly reasonable way because it is designed in accordance with the interests of human beings, people still cannot be sure whether the creation of AI Nanny is reasonable before the AI++ era. Although people start to change the direction of technology, we cannot ensure that everyone will stop exploring smarter areas. If these smarter technologies (not to be friendly singularities) are a blow to human nature, then the existence of AI Nanny is necessary.

Although it may circumvent some of the dangers brought about by the singularity, its rationality is still a problem. After human beings accept its help, it will leave like Nanny, instead of leading us for a long time.

“I suspect government could be done a lot better than any country currently does it — but I don't doubt the need for some kind of government, given the realities of human nature. It may be that the need for an AI Nanny falls into the same broad category. " [4] The author argues that the government has many problems, but it exists for human nature. AI Nanny exists just like the government. However, the author later explained that we did not need AI Nanny before we reached the AI++ era, just as we did not need the government in the Stone Age.

When we need it, it will help us as the government. The existence of the government has helped to stabilize the turmoil in the society, but many problems have not been resolved. Goertzel believes that the role of the government is irreplaceable. The author believes that AI Nanny will be a more offensive existence.

3. CONTRIBUTION

In the singularity, human beings will face ubiquitous dangers and challenges. AI Nanny proposed by Goertzel will help humans and nurture humanity. Through thinking and summarizing, we believe that there are several qualitative differences between AI Nanny and the government.

3.1. The government can review problematic decisions and people, but we cannot find errors in AI Nanny's algorithm.

The government can find its own mistakes. In a democratic society, the government will pass bills (direct democracy or representative democracy) that allow people to assemble, march, or elect reasonably. Nevertheless, this still does not mean that all bills and decisions are just. Therefore, society sets up agencies to check the correctness of the government and its decisions. Taking slavery as an example, the government once believed that slaves were the property of his owner, so it promulgated Fugitive Slave Laws. This bill requires all slaves to return to the owner, even if he is in a slave-free state. Adult white Americans who could vote at the time were very satisfied with this policy, believing that this was the government's commitment to fulfilling the Fifth Amendment (the state protects personal property). In 1865, fifteen years later, the government passed the 13th and 14th constitutional amendments with a majority advantage in the House of Representatives and the Senate (both with a pass rate of more than 2/3) [7]. These two amendments abolished slavery, and at the same time, they gave African Americans citizenship rights. Therefore, we can see from this example that the government is not perfect because it allows African Americans to be discriminated against and abused in society. When people discover the situation of African Americans, human reflections and social progress make us review the previous problematic bills. Therefore, many of the government's mistakes can be corrected.

In Ben Goertzel's article, he wrote: "General intelligence somewhat above the human level, but not too dramatically so — maybe, qualitatively speaking, as far above humans as humans are above apes." [4] AI Nanny. The author's metaphor illustrates a problem. Apes do not understand how humans work, because humans are of higher intelligence. When AI Nanny has higher intelligence, we will not understand its operation. Facts have proved that AI Nanny needs to have the cognition of the target and reasonable calculation of the action, and the complex algorithm is not reliable [3]. A large amount of algorithmic bias reinforced the idea, which created lots of ethical problems and unsatisfaction.

3.2. The government is composed of humans, however, the AI Nanny couldn’t take human interest into its awareness.

Although different governments have different political pursuits and strategies, mankind still has one of the most basic common interests, which is survival. Therefore, the existence of the United Nations and various international alliances has consolidated this. The environmental protection actions and garbage classification proposed by the United Nations are carried

out on the basis that human beings realize that environmental problems threaten the stability of the earth and biological safety. Everyone believes that a human government is capable and necessary to maintain security. When life safety is protected, people begin to pursue that the government can bring enough benefits to people. Therefore, government is an organization composed of human beings dedicated to bringing humanical services and regulations in accordance with people's basic interests.

AI Nanny does not have the same attitude when dealing with these issues. When it discovers that humans encounter a potential threat, it will only stop the threat from happening as a real Nanny. However, we are not a baby, but human beings with independent thinking. Human behavior can be replicated by AI, but the human brain is biologically unique. There are people with multiple ideas in society, and they will form various groups. No matter how good a government is, it can't get everyone or a group to agree to one unique decision, but AI Nanny will send out many robots and its own technology to counter this threat, which brings a singularity that is unfriendly to humans. But if that threat is in line with people's current interests and development, AI Nanny will also destroy it without hesitation. Moreover, some experiments (such as cloning organisms) are just repeated experiments that violate humanism and bring danger, but it does not mean that science does not have the need for development and exploration. AI Nanny could guarantee the safety of human beings, but humanity and human society have been going into the direction that could cause distruction. Human society will lose some diversities, for example, everyone may eat the same healthy food every morning (just because AI Nanny believes that this guarantees people's safety and health, and is the best choice for humans who want to live a long life). The author Ben Goertzel himself admitted that AI Nanny is not the best choice for mankind or even a poor choice. Although it can ensure that mankind stays away from disasters brought about by technology, this did not bring benefits to mankind.

3.3. AI Nanny will compromise more privacy than a human government.

The government usually opposes infringement of personal privacy, but there are still times when agency-authorized inspections are inevitable. These are justice in the judicial process. The protection of privacy is a symbol of human freedom and independence, and it is also part of the government's satisfaction with the common interests of the people. The government does not monitor every actions and decisions in people's daily lives and remains some distance in order to respect and trust people. People's concern for privacy is not simply because we are a civilized society, but it builds the people's sense of security in life emotionally and physically.

The AI Nanny proposed by Ben Goertzel will "Interconnection to powerful worldwide surveillance systems, online and in the physical world. Control of a massive contingent of robots (eg service robots, teacher robots, etc.) and connectivity to the world's home and building automation systems, robot factories, self-driving cars, and so on and so forth. "[4] This sentence shows that AI Nanny will be rooted in every corner of our lives in order to protect us, so humans will compare information explosions The singularity of technology. However, these behaviors undoubtedly add an unprecedented monitor to our lives. It will not only monitor us but also affect our lives. Therefore, we will have no privacy at all. Although we currently do not have the existence of AI Nanny, the transaction of big data also pushes down human privacy again [5]. Internet users generate about 2.5 quintillion bytes of data each day, while 97.2% of technological organizations and specialists feels necessary to invest in big data and AI [6]. Then, we can't imagine that in the future, there will be an AI Nanny who monitors us all day, and our privacy and information will no longer kept a secret.

4. CONCLUSION

The era of AI++ has not yet come, so our guesses will not be accurate. Although the renewal of technology will lead us to an unknown, the basic interest of mankind is to survive. Human Beings cannot know whether the general intelligence or the government will lead us in the future. The government and general intelligence will have many problems, but they are by no means the same. The government can check its own mistakes to come up with better solutions to contribute to the people, but when humans go to review the mistakes of the machine, they will be prevented by complex algorithms (couldn't find out the problem). The complex algorithm leads to not only the beginning of distrust to AI Nanny but also the manifestation of lack of transmission. At the same time, even if AI Nanny protects humans and prevents technological singularities from harming us, it cannot guarantee the interests of the people like the government do. The difference between groups of people makes the government or the country more diversified. Therefore, the decisions made by the government are more representing the people. AI Nanny may lead us to a society under a dictatorship in this way. Society will lose its diversity. Finally, AI Nanny will no longer respect the protection of personal privacy, because its characteristic is to catch any technology or AI++ that may threaten humans (maybe some people will maliciously create it, a terrorist organization or a country). After that, it will eliminate this threat in its own way. Again, we can't predict whether AI Nanny is going to occur in future or not, but we only know that this is an ideal way to counter

singularity disaster, since singularity will be taking place in our future. After all, the long-term development of human society is still a topic that needs to be discussed.

ACKNOWLEDGMENTS

My Gratitude goes to Dr. Cavan at Laurel Springs School. She helped me build some foundations of philosophy and anthropology. Without her help, I would not have enough interest and confidence in this subject. After her guidance, I was able to develop my own propositions and logical thinking. I also want to thank the physics teacher for his help, because he told me how to look at the world more rationally. He cultivated my ability to analyze things objectively. My appriciation belongs to them for their inspiration so that I can progress and develop.

REFERENCES

[1] Chalmers, D. (2009). The singularity: A philosophical analysis. Science fiction and philosophy: From time travel to superintelligence, 171-224.

[2] Chalmers, D. (2012). The Singularity: a Reply. Journal of Consciousness Studies, 19(7-8), 141-167.

[3] Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: review, synthesis, and future research directions. European Journal of Information Systems, 1-22.

[4] Goertzel, B. (2012). Should humanity build a global AI nanny to delay the singularity until it’s better understood?. Journal of consciousness studies, 19(1-2), 96-111.

[5] Stahl, B. C., & Wright, D. (2018). Ethics and privacy in AI and big data: Implementing responsible research and innovation. IEEE Security & Privacy, 16(3), 26-33.

[6] Petrov, Christo. “27+ Big Data Statistics - How Big It Actually Is in 2021?” TechJury, 6 Dec. 2021, https://techjury.net/blog/big-data-statistics/#gref.

[7] Landmark legislation: Thirteenth, Fourteenth, & fifteenth amendments. U.S. Senate: Landmark Legislation: Thirteenth, Fourteenth, & Fifteenth Amendments. (2021, January 11). Retrieved December 30, 2021, from https://www.senate.gov/artandhistory/history/common/generic/CivilWarAmendments.htm

[8] Goertzel, B. (2007). Human-level artificial general intelligence and the possibility of a technological singularity: A reaction to Ray Kurzweil's The Singularity Is Near, and McDermott's critique of Kurzweil. Artificial Intelligence, 171(18), 1161-1173.