DOI: 10.1609/aaai.12022

ARTICLE

Do we need a Hippocratic Oath for artificial intelligence scientists?

Nikolaos M. Siafakas

University of Crete

Correspondence
Nikolaos M. Siafakas, Department of Computer Science, University of Crete, Crete 70013, Greece
Email: siafakan@uoc.gr

Abstract

Artificial intelligence (AI) has been beneficial for humanity, improving many human activities. However, there are now significant dangers that may increase when AI reaches a human level of intelligence or superintelligence. It is paramount to focus on ensuring that AI is designed in a manner that is robustly beneficial for humans. The ethics and personal responsibilities of AI scientists could play an important role in continuing the constructive use of AI in the future. Lessons can be learnt from the long and successful history of medical ethics. Therefore, a Hippocratic Oath for AI scientists may increase awareness of the potential lethal threats of AI, enhance efforts to develop safe and beneficial AI to prevent corrupt practices and manipulations and invigorate ethical codes. The Hippocratic Oath in medicine, using simple universal principles, is a basis of human ethics, and in an analogous way, the proposed oath for AI scientists could enhance morality beyond biological consciousness and spread ethics across the universe.

INTRODUCTION

Artificial intelligence (AI) has been defined as a machine (computer) with the ability to copy intelligent human behavior. AI is the most rapidly advancing science and is connected with the majority of human activities (Tegmark 2017).

Today, AI is still weaker than human intelligence (weak AI), but experts believe that AI may reach the same level as human intelligence, thus becoming so-called strong or human general machine intelligence (HGMI) in the next few decades (Good 1965; Maravec 1998; Kurzmeil 2014; Mnih et al. 2015).

If this occurs, through machine learning techniques and automatic self-improvement, it is likely that human-level AI will easily become super artificial intelligence (SAI) and thus be well above the average level of human intelligence (Kurzmeil 2005; Bostrom 2014).

Although this could be the most significant scientific event in human history, many fear that it could be the last (Dyson 1979; Cussins 2018).

Obviously, if this event occurs, humans will no longer be the smartest entities on the planet. Even well-intentioned interventions for SAI may automatically become harmful, so it is paramount to control SAI so that it has robustly beneficial value for humankind. This control must be implemented during the current stage of intelligence design, well before HGMI becomes SAI. This is the first time in history that scientists have faced such a critical problem with this "short" a deadline. Since we can only speculate about when safety research will reach a solid level of control over SAI, it is prudent to focus our efforts on this issue and initiate global efforts now.

By highlighting ethical and moral issues and emphasizing the personal responsibility of scientists, the necessary control over AI can be reached faster and more

efficiently, as can the alignment of the goals of AI with those of humans.

The most successful field of practical ethics is medical ethics, with its long history going back to Hippocrates and his famous oath. Thus, lessons may be learnt from this successful story.

Today, there are a number of ethical codes for computer scientists produced by private corporations, but these codes have been criticized as "fraught with hollow promises, oversights and mistakes" (Statt 2019). In addition, ethical codes that have been created by scientific societies lack widespread implementation (Morley et al. 2019).

Thus, a Hippocratic Oath for AI scientists could be of significant value. I believe that if one takes an oath, one is most likely bound by its ethical principles. Furthermore, such oaths should give more strength to ethical codes.

THE HIPPOCRATIC OATH

Approximately 2500 years ago, the Greek physician and founder of modern medicine Hippocrates of Kos realized the potential dangers of medical malpractice. To minimize the misuse of medicine for humans, he proposed an ethical oath to be taken by all students before starting to practise medicine. Since then, the oath, which is now commonly known by his name, has had a tremendous effect on medical ethics and has become the basis of modern ethical standards and written documents, such as the Nuremberg Code, the Declarations of Geneva and Helsinki, and the Belmont report, among others.

Throughout history, the oath, despite many modifications and modernizations, has helped uphold high medical standards, and it is still in use by most medical schools in the Western world (Kao and Parsi 2004).

The Hippocratic principle "*Ωφειλεειν ου βλαπτειν*," "first do no harm," is considered the basis of human ethics.

The Hippocratic Oath has survived for so long because it contains simple and universal values; thus, it is still used in medical practice (Siafakas 2011), and its tremendous effects can be seen today, particularly during the current pandemic.

A Hippocratic Oath for AI scientists may be similarly beneficial for humans, as it can help eliminate AI malpractice, which theoretically has the potential capacity to destroy life (Hawking 2018; Tegmark 2018a). In addition, such an oath may facilitate research on the control of AI and augment its beneficial effects for humans.

DANGERS OF AI

If and when AI becomes super intelligent, it will be a machine with intellectual capacities well above those of the human brain. This raises a fundamental question: If we are no longer the smartest, will we be in control? This factor alone poses a great danger to humankind (Yudkowsky 2006; Kaku 2014; Harari 2017; Tegmark 2018b). There are a number of ways that SAI may act against humans, from using humans as we have used animals to utilizing all the available energy for its functioning leaving humans without, for example, electricity.

Furthermore, during this time of technological innovation in the race to develop SAI, something may go wrong, and although AI is generally programmed to be beneficial, it may turn out to be harmful (Cussins 2018).

In addition, AI could be programmed to do something devastating, such as using autonomous weapons, and such weapons could be mass produced or fall into the wrong hands and cause massive destruction. This risk is already acknowledged, and 116 expert scientists have signed a letter arguing against the further development and use of autonomous weapons by AI (Future of life Institute 2019).

This risk may exponentially increase if SAI is achieved.

Another potential scenario is that SAI may intentionally or unintentionally use other very lethal weapons, such as nuclear, biological, or chemical weapons, or accelerate climate change.

Moreover, SAI may easily bypass contemporary medical ethics and induce large-scale malpractice in medicine and in human genetics (Balthazar et al. 2018; Keskinbora 2019).

In conclusion, AI has been very beneficial for society thus far, but its misuse poses significant threats, and there is an urgent need to utilize all our wisdom to develop safe AI.

A Hippocratic Oath for AI scientists can definitely enhance awareness of all these potential dangers and focus scientific efforts on the production of robustly safe AI.

ARGUMENTS FOR A HIPPOCRATIC OATH FOR AI SCIENTISTS

Although many oaths for scientists have been proposed, there are important reasons to have a specific one for AI scientists.

Over the past 50 years, philosophers Karl Popper, Nobel Peace Prize winner Joseph Rotblat, and others have proposed oaths for scientists analogous to the Hippocratic Oath for medicine (Cressey 2007; Ghosh 2007). All of these proposals have been based on the fundamental principle from Hippocrates of "first, do no harm." However, some ethics experts have opposed them (Woodley 2012). The proposed oaths have not flourished and have never achieved universal status. This may be because they came very late, after the production and use of very dangerous weapons (e.g., nuclear bombs), and it was impossible to stop this

arms race. On the other hand, well before this has occurred with AI, scientists have realized that SAI may be the most brilliant discovery of all time but that it may also be the final one (Omohundro 2008; Bostrom et al. 2016). Therefore, the aim of the Hippocratic Oath for AI scientists is that this never occurs and that AI will only be beneficial for humans.

The principal benefits of a Hippocratic Oath for AI scientists would be that it could:

1. Significantly increase awareness of the potential lethal threats of SAI among students, scientists, the public, decision makers, and politicians and boost funding and donations for exploring safe AI.
2. Enhance collaborations among scientists and their focus on the control of AI, utilizing their resourcefulness to achieve, as soon as possible, secure and beneficial use of SAI.
3. Emphasize the personal responsibility of each individual scientist for the consequences of their activities; by informing them of the penalties of breaking the oath, such as losing the right to work in AI, malpractice may be avoided.
4. Enhance efforts to align the goals of AI with those of humans well before the development of SAI.
5. Improve ethical standards in computer science, invigorate the implementation of ethical codes, and prevent corrupt practices such as hacking, fake AI, data poisoning, and others.
6. Protect individual scientists from being misused by malicious employers, corporations, or governments.
7. Prevent AI from producing or using autonomous, nuclear, biological, or other lethal weapons or waging a cyberwar against humanity.
8. Prevent medical malpractice by AI and malicious manipulation of human or other genomes.
9. Help AI scientists fight climate change using AI technologies to reverse negative changes and always respect the environment.
10. Assist AI scientists in gaining a better understanding of the fundamental principles of the cosmos and facilitating space travel and the flourishing of human life throughout the universe.
11. Carry significant future implications if and when SAI develops "consciousness" or "feelings" or other high-level capacities of the human brain (Wallach et al. 2011).

A HIPPOCRATIC OATH FOR AI SCIENTISTS

The oath should be taken by students of computer science, robotic scientists, AI scientists, AI programmers, logic theorists, and others with scientific knowledge related to AI. It would preferably be taken early in an individual's career, such as at university graduation ceremonies or when they earn their licence to practise.

THE OATH

The Oath could be as follows:

With free will, I swear that I will carry out according to the best of my abilities and judgement this oath.

I will use my knowledge to make AI of any kind, biological, nonbiological, or mixed, beneficial for humanity.

I will respect my colleagues and peers, and I will share my knowledge with them to make AI not only beneficial but also safe for the human race.

I will consider all ethical implications before taking any action by evaluating the consequences of those actions for humankind.

I will take all necessary steps to prevent corrupt practices and professional misconduct even if my life is in danger, and I will always declare my conflicts of interest.

I will never produce or use AI to develop autonomous, nuclear, biological, chemical, or other lethal weapons, and I will always use ethical machine learning techniques.

I will never utilize hiking, fake AI, data poisoning, or other malpractice or engage in cyberwar against humanity.

I will respect the environment, and if needed, I will use AI to reverse dangerous changes to the climate.

I will never use AI for medical malpractice or malicious genetic alterations, and I will also actively prevent such use.

I will promote AI for a better understanding of the fundamental principles of the universe and will never use this knowledge against life.

I will use AI to improve space travel and human life everywhere in the universe.

DISCUSSION

Problematic ethical implications of AI are not new, having been first noted in the 1940s by Alan Turing and continuing to be pointed out in the 1960s by Wiener (Turilli 2008). Recently, this discussion has intensified since the capabilities of AI have developed tremendously, and it is foreseen that a machine of human-level intelligence is very near realization (Good 1965). In addition, early in this

rapid development, the potential great dangers of AI were recognized, especially when AI became super intelligent (Wiener 1960). It was noted that "once this Pandora's Box is opened, it will be hard to close" (Musk 2017). Thus, a balance between improving AI for the benefit of society and limiting its potential harm is urgently needed (Taddeo and Flori 2018; Tegmark 2018b).

Ethics are considered one of the pillars of such an effort, and arguments have addressed how ethical principles can be translated into AI practices (Morley 2019). These efforts aim to ensure that human rights will be respected by ethically aligned algorithms (Moor 2006; Bostrom 2013).

Philosophers and ethics experts have tried to describe how ethical AI may look, but they cannot instruct on how to achieve it (Bostrom 2013). Thus, this urgent goal of developing constructive but safe AI must remain a primary concern of all scientists.

Today, there are a large number of publications on ethical principles and frameworks for AI, 23 of which emerged from the Asilomar conference (2015), and others have been developed by associations such as the Global Initiative for Ethical Considerations in AI and Autonomous Systems (IEEE 2019), Partnership AI (2019), ACM (2018), and European Commission (2019). However, none of these has suggested how to implement the needed principles nor proposed an oath as a practical and efficient way to do so. An oath would increase the personal responsibility of the scientists involved. It would also enhance the implementation of ethical codes, giving them the backing and strength they need (Veliz 2019) after being approved by academic and international scientific societies.

As the parallels between AI ethics and medical ethics are extensive, the proposed oath should be accompanied by the development of ethics committees to be more effective (Veliz 2019).

History has shown that an oath is not a panacea for all malpractice issues; this is the case even with the Hippocratic Oath, which has been unable to prevent medical war crimes, for example, Nazi medical atrocities, and unethical research, as in the Tuskegee syphilis study.

However, it is well known that the Hippocratic Oath was established very early in the history of humankind, and the ethical principles and values of medicine persist even today (Siafakas 2011), affecting medical professionals' behavior in positive ways, as seen, for example, during the current pandemic. In addition, its fundamental principle of "first, do no harm" is one of the bases of human ethics. Although Hippocrates had no idea of the tremendous progress medicine would make, his oath is still taken by most medical students in Western universities (Kao and Parsi 2004) because it contains simple, easily applied, and universal ethical rules.

Similarly, we cannot foresee the progress that AI will make, especially if it becomes HLMI or SAI, since there are factors of which we are not yet aware. However, expert scientists currently emphasize that the potential lethal danger of SAI may alter the trajectory of human civilization and urge that this trajectory be controlled before it is too late. The promotion of ethics in AI science may prevent SAI from being our last invention. By reinforcing the personal responsibility of AI scientists for the consequences of every action taken, the misuse of AI can be minimized. In addition, by increasing the sharing of scientific knowledge and collaboration among scientists of different disciplines, a robustly safe AI, which is urgently needed, may be achieved faster and more easily.

Clearly, personal responsibility alone is not sufficient to deter unethical utilization of AI. As in medicine, credible world institutions are needed to determine what is crime, malpractice, or just an unintended accident with respect to AI. In addition, developing regulatory bodies similar to those associated with medical research that have the ability to impact funding would be an effective approach to self-regulation. Finally, an international treaty could be developed and signed by many countries and enforced by a world court.

In summary, the aim of an oath for AI scientists is to ensure that the beneficial aspects of AI can be enjoyed by humans for a long time.

Just as the Hippocratic Oath for medicine has lasted for centuries and promoted human ethics, it is hoped that this oath for AI scientists will affect humanity in the best ways and, in the future, enhance moral rightness in the universe.

REFERENCES

ACM. 2018. *ACM Code of Ethics and Professional Conduct*. https://www.acm.org/code-of-ethics

Asilomar conference. 2015. http://tinyurl.com/asilomarAI

Balthazar, P., P. Harri, A. Prater, and N. M. Safdar. 2018. "Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics." *Journal of the American College of Radiology* 15: 580–6.

Bostrom, N. 2013. Ethical issues in advanced artificial intelligence. http://www.nickbostrom.com/ethics/a.i.html

Bostrom, N. 2014. *Superintelligence: Paths, Dangers, Strategies*. London, UK: Oxford University Press.

Bostrom, N., A. Dafoe, and C. Flynn. 2016. Policy desiderata in the development of machine superintelligence. http://nickbostrom.com/papers/aipolicy.pdf

Cressey, D. 2007. Hippocratic oath for scientist. http://blog.nature.com/news/thegreatbeyond/2007/09

Cussins, J. 2018. How to prepare for malicious use of AI. http://futurelife.org

Dyson, F. J. 1979. Time without end. http://blog.regehr.org/extra-files/dyson.pdf

European Commission. 2019. Ethics Guidelines for Trustworthy AI. https://go.nature.com/2K6HWBl

Future of Life Institute. 2018. Open letter against autonomous weapons. http://futurelife.org/open-letter-autonomous-weapons/

Ghosh, P. 2007. UK science head backs ethics code. http://news.bbc.co.uk/2/hi/science/nature/6990868.stm

Good, I. J. 1965. "Speculations concerning the first ultra-intelligence machine." In *Advances in Computers*, edited by L. Franz and R. Morris, 31–88. New York, NY: Academic Press.

Harari, Y. N. 2017. *Homo Deus. A Brief History of Tomorrow*. New York, NY: Harpet-Collins.

Hawking, S. 2018. *Brief Answers to the Big Questions*. New York, NY: Bantam Books.

IEEE-SA. 2019. http://standards.ieee.org/develop/indcom/ec/ead_v1.pdf

Kaku, M. 2014. *The Future of the Mind*. New York, NY: Doubleday.

Kao, A. C., and K. P. Parsi. 2004. "Content analyses of oaths administered at U.S. medical schools in 2000." *Academic Medicine* 79: 882–7.

Keskinbora, K. H. 2019. "Medical ethics considerations on artificial intelligence." *Journal of Clinical Neuroscience* 64: 277–82.

Kurzweil, R. 2005. *The Singularity is Near*. New York, NY: Viking Press.

Kurzmeil, R. 2014. *How to Create a Mind*. London, UK: Duckworth Overlook.

Maravec, H. 1998. "When will computer hardware match the human brain?." *Journal of Evolution and Technology* 1: 1–12.

Musk, E. 2017. https://www.inheart.com/content/2017-08-21-elon

Mnih, V. et al. 2015. "Human-level control through deep reinforcement learning." *Nature* 518: 529–33.

Morley, J., L. Floridi, L. Kinsey, and A. Elhalal 2019. "From what to how: An initial review of publicly available ai ethics tools, methods and research to translate principles into practices." *Science and Engineering Ethics* 26: 2141–68, https://doi.org/10.1007/s11948-019-00165-5

Moor, J. 2006. "The nature, importance, and difficulty of machine ethics." *Ieee Intelligent Systems* 21: 18–21.

Omohundro, S. 2008. The basis AI drivers. http://tinyul.com/omohundro2008

Partnership AI Organization. 2019. http://www.partnershiponai.org

Siafakas, N. M. 2011. "Preventing exacerbations of COPD–advice from Hippocrates." *New England Journal of Medicine* 365: 753–54.

Statt, N. 2019. "Google dissolves AI ethics board just one week after forming it." *The Verge*. https://go.nature.com/2Zg727k

Taddeo, M., and L. Floridi 2018. "How AI can be a force for good." *Science* 361: 751–2.

Tegrmark, M. 2017. Research priorities for robust and beneficial artificial intelligence. http://fururelife.org/ai-open-letter/

Tegmark, M. 2018a. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York, NY: Penguin.

Tegmark, M. 2018b. Benefits and risks of AI. http://futurelife.org

Turilli, M. 2008. "Ethics and practices of software design." In *Current Issues in Computing and Philosophy*, edited by A. Briggel, P. Brey, and K. Waelberts, 171–83. Amsterdam: IOS Press.

Veliz, C. 2019. "Three things digital ethics can learn from medical ethics." *Nature Electronics* 2: 316–18. https://doi.org/10.1038/s41928-019-0294-2

Wallach, W., S. Franklin, and C. Allen. 2011. "Consciousness and ethics: Artificial conscious moral agents." *International Journal of Machine Consciousness* 3: 177–92.

Wiener, N. 1960. "Some moral and technical consequences of automation." *Science* 131: 1355–8.

Woodley, L. 2012. Do scientists need an equivalent of Hippocratic Oath to ensure ethical conduct? http://www.lindau-nobel.org/do-scientist-need-an-equivelent-of-the-hippocratic-oath-to-ensure-ethical-conduct/

Yudkowsky, E. 2006. Artificial intelligence as positive and negative factor in global risk. http://inteligence.org/files/AIPosNegfactors.pdf

AUTHOR BIOGRAPHY

Nikolaos M. Siafakas is a well-known Clinician Educator and Researcher. He studied at Athens University (MD), London (PhD) and Paris Universities as well as, at Mc Gill and UCSD. He had been Director of Pulmonary Departments in Athens and Crete, and is Professor of Thoracic Medicine in the University of Crete, Medical School since 1988. He served as President of the Hellenic Thoracic Society (twice), The European Respiratory Society (ERS 2009–2010), and as Vice Rector of the University of Crete. He published more than 350 research papers in PubMed and Journals and edited 5 Books. Recently, he is focusing on applications of Computer Science in Medicine and their Ethics.

How to cite this article: Siafakas, N. M. 2021. "Do we need a Hippocratic Oath for artificial intelligence scientists?" *AI Magazine* 42: 57–61. https://doi.org/10.1609/aaai.12022