HOSPITAL CHRONICLES 2019, 14(1): 3–6

EDITORIAL

Artificial Intelligence: Are we Creating a New Frankenstein?

Athanasios G. Yalouris, MD

Coordinating Director, Department of Internal Medicine, "Elpis" Hospital, Athens, Greece

KEY WORDS: *Frankenstein, artificial intelligence, materialism, dualism, learning machines, robot revolution*

Mary Shelley (1797-1851) is an English novelist best known for her Gothic novel* *Frankenstein or The Modern Prometheus*, written in 1818. In this novel, Victor Frankenstein, an excellent young scientist specialized in chemistry but also connoisseur of other sciences, develops a genius technique to impart life in a huge humanoid that he constructed using parts of dead human bodies. However, when he sees his creature come into life he abandons it terrified. As the creature wanders without an aim or help, it faces human enmity and that transforms it to a maniac for vengeance, extremely directed against its creator. It does not hesitate to murder the persons who are most precious to Victor, including his younger brother and even his bride at the night of their wedding. Victor starts a desperate chase of his creature that leads him to the North Pole, where he dies of exhaustion. The Creature, seeing him dead, mourns for him and, having decided to die too, drifts away on an ice raft and is soon "lost in darkness and distance", never to be seen again.[1] Although the "Creature" remains nameless in the novel, it is usually referred in every-day practice with the name of its creator. That's why the name "Frankenstein" is often used metaphorically to describe an evil existence that causes death and destruction (Figure 1)**.

FIGURE 1. Mary Shelley and her famous creature, Frankenstein.

Correspondence to:
Athanasios G. Yalouris MD
3 Tsaldari street,
153 43 Agia Paraskevi,
Athens, Greece
Tel.: +30 210 6013511
E-mail: yalourisa@gmail.com

* Gothic is a genre of romantic novel that combines fiction and romance with horror and death.

** https://www.google.com/search?q=mary+shelley&tbm=isch&source=iu&ictx=1&fir=ZCFYJRmvLQ4J pM%253A%252CYjcdiEbCdeJNnM%252C%252Fm%252F04_by&vet=1&usg=AI4-kRIqJpdXYyeHeRZTq G4dW2uvRZMiA&sa=X&ved=2ahUKEwjPpvzRu_XgAhXC16QKHYoxAQgQ_h0w FHoECAUQCA#imgdii=DoKQMV7djAVNaM:&imgrc=rc_O6_NT-51q6M:&vet=1

Frankenstein, although being a novel very interesting for its plot and originality, can also be read as a genius parable on the dangerous and actually catastrophic consequences of scientific progress if it is not self-limited by ethical principles or awareness of its possible risks. I believe that this second level of reading is responsible for the longevity of the popularity of this novel and the place it has taken in world literature. Furthermore it can be seen as a very useful warning to any future operation that aspires to make a breakthrough in science.

The term "Artificial Intelligence" (AI) has been introduced to define intelligence demonstrated by machines through a long series of algorithms. AI is not only a technological science targeting to the development of machines that would help the humans –and possibly compete with them- in several activities, mainly intellectual. It is also a psychological science which, by reproducing the main characteristics of human cognitive functions aims to increase our knowledge on the way human mind works.[2]

The prodromal form of AI was the scientific field of Cybernetics. This term was introduced in 1947 by Norbert Wiener and included the study of control and communication between living creatures and machines.[3] In machines the automatic control is achieved through a negative feed-back that tends to keep any present situation within certain limits non-significantly deviating from some standard conditions.[2] This is actually an imitation of the physiological phenomenon of homeostasis, well known to any doctor or biologist. Technological development has permitted the scientists to make a considerable progress on the subject.

AI machines are currently used in several aspects of everyday practice. As regards to Medicine they are promising to offer several advantages over diagnosis and treatment of human disease. It is important to mention that AI projects involved in healthcare had the highest financial support among all other sectors in 2016.[4]

AI supporters claim that the machines may modify medical practice by relieving clinicians from dull data collection and analysis and leaving to them enough time to focus on more essential medical work. Computers can work out patient data with considerably higher speed and reliability and so help in organizing medical files of patients with chronic diseases (such as diabetes, hyperlipidemia or hypertension) demanding frequent follow-up. An AI machine may rapidly review the whole file and highlight the points that should be carefully studied in each visit of the patient, define the optimal time for a follow-up visit or even suggest necessary actions to be performed. There will be also an economical benefit by eliminating –through relevant and complicated algorithms- care that patients don't need.

Today, neural networks tend to imitate human brain by using numerous interconnected neurons and the whole process is rapidly improving by the time. They are already capable to approach rather complicated clinical problem solving. More importantly, they are so programmed as to progressively increase their abilities by incorporating new data obtained by "experience", thus reaching a level of acquisition of new knowledge. In other terms we are now referring to "learning machines". Such a machine can be exposed in a very short time –possibly minutes- to a number of cases that a clinician will need a whole lifetime to obtain.[4]

Such AI machines can support diagnosis based on imaging techniques (X-rays, hypersonograms, CT scans, MRIs), e.g. by revealing slight changes in successive tests invisible to the human eye. Furthermore they can help in evaluating chest radiograms sent to a center from remote areas of countries with high prevalence of tuberculosis and lack of specialized doctors.[5] AI may also be useful in clinical diagnosis. Esteva et al, working in California U.S.A., trained a convolutional neural network in the differential diagnosis of dermatological malignancy. They used a dataset of 129,450 clinical images consisting of 2,032 different diseases. They tested its performance on clinical images with two critical binary classification use cases: a. keratinocyte carcinomas (the commonest skin cancer) versus benign seborrheic keratoses. b. malignant melanomas (a usually lethal skin cancer) versus benign nevi. The final diagnosis was defined by skin biopsy. The AI machine succeeded in diagnosing skin cancer equally well to a group of 21 board-certified dermatologists.[6] Other, more specific, benefits cannot be ignored. For example, AI may prove more reliable than a specialist in defining the exact target areas for head and neck radiotherapy to avoid needless expose to irradiation.[7]

Some scientifically emerging countries, such as China, tend to be pioneers in the use of AI in Medicine, possibly because they lack a sufficient number of trained doctors. In China, "Doctor vs. machine" competitions are frequently organized and presented on television. Sometimes the results are in favor of the machines, as in a recent (2018) competition in Beijing concerning the diagnosis of brain tumors or the prediction of the expansion of brain hematomas or bruises. In both fields the Biomind AI system won a team of 15 expert doctors from top hospitals across China.[8]

It is generally expected that AI will have a positive effect in every human activity with the machines doing for us the difficult, dangerous or boring work, so that we will be free to act in more intellectual fields. If this proves to be true in medical practice, some of the roles of physicians will have to change and they must undergo a specific training in order to be able to coexist and collaborate with the AI machines. Some medical specialties may be more seriously affected and possibly their specialists will have to deviate to a different mode of working. Ethical problems may also arise concerning the use and elaboration of patients' data by non-medical stuff that is not bound by an ethical code, such as the Hippocratic oath.

Of course, AI is not -with the present status- expected to replace clinical doctors. Diagnosis may be assisted, treatment

recommendations may be suggested but the final decision must be kept for a human mind. However, the question that emerges is unavoidable: will the present status remain unchanged or we shall face in the future radically different conditions?

Up to date the activities of AI machines are limited by their programming. Although day by day they become even more complicated, they remain under human control. However AI intelligence is gradually increasing. Can it happen that one day it will surpass human intelligence? There are several reasonable arguments against this possibility. A human being will not be able to create a machine cleverer that himself. The increase of AI is provoked by humans through proper challenges that are offered to the machines. If there is not an additional challenge, there will be no additional intelligence. Furthermore there is not a unique solution to the several problems that have to be solved in practice. Algorithms can be effective for a certain target but are completely useless for any other. For example, a chess-playing AI machine can play chess equaly well to a Chess World Champion and possibly win him. To succeed in this field, there is no need to "learn abstract concepts, think cleverly about strategy, compose flexible plans, make ingenious logical deductions" etc.[9] A special-purpose algorithm can sufficiently cover this field. But, of course, a machine highly intelligent in chess playing will be absolutely idiotic in any other intellectual field. On the contrary, a human being can face numerous everyday problems through a brain that concurrently uses millions of neural structures in different combinations. A machine will never be able to imitate these mechanisms.

Can we be sure that this way of thinking is right? I 'm afraid, not. The idea of a robot revolution first appeared as a clever science-fiction topic. In the current status it may be more than that. Scientists working on this field –as Eliezer Yudkowsky, a renowned AI researcher, working in Berkeley, California- have discussed the possibility that intelligent machines will become conscious and pose a great threat to humanity. It may be only a nightmare scenario, but can we reject the possibility that "intelligent" machines will one day start acting independently from humans? What will happen then? It is possible that they will proceed to production of new machines with progressively increasing intelligence. If that is really achieved, will they finally create machines with much higher intelligence than that of the humans? One can reasonably conclude that in this case, we humans will appear to the machines just as animals appear to us and that these superintelligent machines will no more be our servants but revolt against us.[10] Another interesting point has been suggested: Let's suppose that we create an AI machine that is programmed to have a positive feed-back when an action it exerts causes our satisfaction, shown by our smiling or our sense of joy. There is a possibility that a superintelligent machine, when it fails to produce our satisfaction with its actions, will attempt not to change its action but to cause our smiling (e.g by paralyzing our face musculature) or change our brain function in order to produce our satisfaction (e.g. by implanting electrodes into the pleasure centers of our brains) that is needed for its positive feed-back.[9] Here we return to possibilities looking as science-fiction. But we must not forget that several times in the past a science-fiction scenario became true after years. Jules Vern, possibly caused his contemporaries' smile with his fantastic travels, but after years some of them came actually to reality.

Let's now turn to a philosophical question. Can a machine be really independent from humans? Such a possibility has as a precondition that machines will acquire some kind of sentiments and a sense of self-existence. Is that possible to happen through algorithms or other technical procedures? Here is the philosophical point: Materialists argue that only matter exists and even mental phenomena are the result of interactions of matter.[11] So, what we call "mind" is nothing more than one group of cerebral functions mediated through the same chemical or physical mechanisms as any other brain activity. Today, artificial neural networks try to imitate the biological mechanisms of brain function. A better understanding of human brain neurophysiology may furthermore improve these systems and this is currently one of the main goals of AI research.[12] Wiener suggests that if we ever construct a machine with a mechanical structure accurately imitating human brain physiology, we will have a machine with spiritual abilities equal to those of human beings.[3] In other words, if we can make a machine that would be nearly as "clever" or even superior to a human being why not reach a point where this machine will also "feel" joy or sorrow?

On the other side, mind–body dualists believe that some mental phenomena are non-physical, since mind and body are distinct and separable.[11] Dualism is based on the work of René Descartes who described mind as a nonphysical -and therefore, non-spatial- substance,[13] closely related with consciousness and self-awareness.[14] Dualists parallel machines to the animals. Humans know that they are conscious and have free will. We don't know whether animals are conscious or not. Some of them, e.g. chimpanzees, have a high level of intelligence but don't care about beauty nor would spend a lot of time in activities which they don't need in order to survive, such as the arts.[15] According to the Dualists this happens because beauty and ugliness are understood through non-physical or spiritual mechanisms that belong to the "mind". In such a case AI machines may be very useful as our servants but they will never be able to obtain "spirit", possess sentiments or come to such an intellectual level that would threaten our authority and power.[16]

Some scepticists about AI suggest another possibility. AI research includes intensive efforts to better understand human brain physiology. If one day we can reach a point where all secrets of cerebral functions will have been revealed, why not try – instead of transferring our knowledge to machines- to

interfere with human neurons and improve their function. In other words instead of creating cleverer machines why not create cleverer humans? This seems an intriguing possibility, although far distant. However, can we be sure that such an intervention -even if it proves effective in increasing human intelligence- will not have other unwanted side-effects? Will such a superintelligent human being remain mentally sane or we will produce a very clever but insane or highly malevolent personality?

I would like to finish by suggesting a personal approach. Several technical progresses –and we have lived too many of them in the last decades- have as an indirect result a decrease in some human capabilities that tend to be less exerted. We nowadays have cars, busses and agricultural machines but we don't have so strong musculature as our ancestors who had to walk long distances or dig their fields with their hands to obtain their crop. To use an example from Medicine, we modern doctors are not so keen in chest or heart auscultation as were the doctors in the first half of the 20th century or even earlier who could not count on a support by ultrasonograms or computerized tomographies. It is a common rule in Nature that any organ or ability that is not in a constant working, tends to atrophy. If we pass too much intellectual work to AI machines may we lead our brain to a state of atrophy? In other words is there a danger that by increasing artificial intelligence we tend to decrease human intelligence?

There are several questions about AI and most of them are discussed worldwide. Of course, Science progresses and no possible skepticism can arrest it. However, I cannot avoid thinking of the possibility that we are actually trying to create a new Frankenstein, only to face the unavoidable consequences of our thoughtlessness if we finally succeed in it. Mary Shelley's allegory has several times been confirmed in current times. Let's hope that she will not prove an everlasting prophet and the doom for humanity will not come -at least through the AI procedure.

REFERENCES

1. Shelley M. *Frankenstein, or, The Modern Prometheus*, Penguin, London, 1994.
2. Cordeski, R. Artificial intelligence and cognitive sciences. In Eco U, Fedriga R (eds). *The History of Philosophy. Vol. 8. 20th Century: Philosophies and Sciences*, Pedio, Athens, 2018, pp. 188-200.
3. Wiener N. *Cybernetics and Society*. Papazisis editions, Athens, 1970.
4. Varun HB, Irfan A, Mahiben M. Artificial intelligence in medicine: current trends and future possibilities. *Brit J Gen Pract* 2018; 68:143-144. DOI: https://doi.org/10. 3399/bjgp18X695213
5. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* 2017; 284:574–582.
6. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542:115–118.
7. Chu C, De Fauw J, Tomasev N, et al. Applying machine learning to automated segmentation of head and neck tumour volumes and organs at risk on radiotherapy planning CT and MRI scans. *F1000 Research* 2016; 5*:*2104.
8. Xiaodong W. AI defeats top doctors in competition. China Daily. Available at: http://africa. chinadaily.com.cn/a/201807/02/WS5b397076a3103349141e006b.html. Accessed May 31, 2019.
9. Bostrom N. *Superintelligence. Paths, Dangers, Strategies*. Oxford University Press, Oxford, 2014.
10. Yudkowsky E. Artificial Intelligence as a Positive and Negative Factor in Global Risk. In Bostrom N, Ćirković M (eds). *Global Catastrophic Risks*, Oxford University Press, New York, 2008, pp. 308-345.
11. Angeles, P. *The Harper Collins Dictionary of Philosophy*, Collins Reference, New York, 1992.
12. van Gerven M, Bohte S. Artificial Neural Networks as Models of Neural Information Processing. *Front Comput Neurosci* 2017; 11:114. doi:10.3389/fncom. 2017.00114.
13. Cottingham J. *Philosophy of Science. A. The Rationalists*, Polytropon, Athens, 2003.
14. Pellegrinis Th. *The five epochs of Philosophy*, Pedio, Athens, 2015.
15. Savain, L. Why the Superintelligent Machine Is a Materialist Pipe Dream. Available at https://medium.com/@ Rebel Science/why-the-superintelligent-machine-is-a-materialist-pipe-dream-9caae341548. Accessed December 10, 2017.
16. Yalouri A. Artificial intelligence as a point of conflict between materialists and dualists. Proceedings of the Philosophy Conference for Students "Summa Studiorum Philosophiae, Novi Sad 2018. (to be published)