

The Ethics of Artificial Intelligence

Mersiha Ismajloska¹, Jane Bakreski¹

¹University of Information Science and Technology “St.Paul the Apostle”, Ohrid, R.Macedonia
mersiha.ismajloska@uist.edu.mk, jane.bakreski@uist.edu.mk

Abstract. The issue of the ethic of artificial intelligence came, maybe as the most important for the future development of AI, according to the thesis which concern humanity, humans and their interaction in the world, especially the world of technique. The issue connects philosophy and technique and therefore is present not only in scientific journal, but in SF literature, visual arts that often interferes with movie industry and of course with the principles of aesthetics. Starting from the three Asimov laws of robotic, describe in his book I, Robot, 1950s, we can develop a study for the ethic of artificial intelligence, always having on mind that subject is complex and changeable during time.

Keywords: ethic, artificial intelligence, aesthetic, laws of robotic

1 Introduction

To define and explain the subject of the ethics of artificial intelligence first we must describe and analyze the three mentioned Asimov's laws. First of them reads: *A robot may not injure a human being or through inaction, allow a human being to come to harm.* This law is in logical interaction with the others two: *A robot must obey orders given by human beings except where such orders would conflict with the First Law* and *A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.* That is a continuation of a humanistic line of philosophy present since ancient times. The way of thinking is similar, but the technique has changed during time.

2 Homo sapiens / AI

Implementation of the three Asimov laws in the field of robotic and philosophy put the human in the center similar to Leonardo da Vinci, Vitruvian Man, and we can say that the humanism as principle is present in science abstraction and in some way we can talk about science philosophy. The romanticism idea of the man as the center of Universe found its re-actualization in robotic in a very original

way. First of all physical similarity of man and robots implies a wish for creating friendly mechanical companions to humans that has of course led to existential and essential consequences provocative for human thought and way of leaving.

As it is said in *Artificial Intelligence A Modern Approach*: we call ourselves Homo sapiens-man the wise-because our intelligence is so important to us. For thousands of years, we have tried to understand the thinking process; that is, how a mere handful of matter can perceive, understand, predict, and manipulate a world far larger and more complicated than itself. The field of artificial intelligence, or AI, goes further still: it attempts not just to understand but also to build intelligent entities. The most important questions is related to interaction of humans and robots and almost all variations as part of narration of Asimov *I Robot*. *I Robot*, originally play with the identity, starting with the title, and connecting human identity with that of a robot, and speaking of robot identity we are getting closer to the question of consciousness that directly leads to field of ethics. The impossibility of the robot to get consciousness about its existence, in the real science world, give to science fiction provocation to make creative combination with possibility of interaction of robot and men. Having in mind SF literature and the presence of mechanization in some visual art works give us the privilege to speak with terms of imagination when explaining of machine-human relation(ship). That connection is an abstract way of speaking about human ethic, which originated from the first moment when man become conscious about himself.

3 Definitions and possibilities

In that manner of speaking machine ethics is the field of research concerned with designing Artificial Moral Agents (AMAs), robots or artificially intelligent computers that behave morally or as though moral. Asimov proposed the Three Laws of Robotics to govern artificially intelligent systems. His intelligent manipulation with boundaries of his three laws to see where they would break down, or they would create paradoxical or unanticipated behavior create interesting play with Susan Calvin, Alfred Lanning, Robbie, Stephen... His work suggests that no set of fixed laws can sufficiently anticipate all possible circumstances. That's the way science give coordinates to artistic thought and every artistic thought starts from scientific attempt to make precise definition of the issue. Of course art is playing with definitions, but always having in mind that definitions exist. In the case of Artificial Intelligence according to *Artificial Intelligence A Modern Approach*, the starting point is that it can't be said what it is AI in first place. Book suggests two ways of making a definition, one with *though processes* and *reasoning*, and the other with *behavior*. For example:

Thinking Humanly: The exciting new effort to make computers think...*machines with minds*, in the full and literal sense. (Haugeland, 1985)

Acting Humanly: The art of creating machines that perform functions that require intelligence when performed by people. (Kurzweil, 1990)

Thinking Rationally: The study of mental faculties through the use of computational models. (Charniak and McDermott, 1985)

Acting Rationally: Computational Intelligence is the study of the design of intelligent agents (Poole, *et al.*, 1998)

(*Artificial Intelligence A Modern Approach*, page 3)

Implementation of this kind of definition in art brings speculations about rational and human making question about identifying rational with human and rational with machines. Identifying rational-human identifying go principles of ethic arise, but with rational-machines identifying principles of ethic must be carefully input in the manner of robots making. Asimov creates an interesting concept of that issue, and fact that stories originally appeared in the American magazines *Super Science Stories* and *Astounding Science Fiction* between 1940 and 1950. The stories are spoken with Dr. Susan Calvin voice and a narrator who re-tells them is a reporter in 21 century, ironically, this current century. Though the stories can be read separately, they share a theme of the interaction of humans, robots and morality, and when combined they tell a larger story of Asimov fictional history of robotics.

4 Interpretation of ethics and AI through art

Visual interpretation of that issue and the three laws of robotic came with Alex Proyas film, too, adaptation of Asimov book. The book has been adapted in a movie in Hollywood manner in 2035 year when, techno-phobic cop investigates a crime that may have been perpetrated by a robot, which leads to a larger threat to humanity. Speaking in terms of aesthetics we can say that this time Hollywood made not only a popular film which is expected for Hollywood, but at same time original visual work where the original Asimov message is sent to viewers. Here, we can talk about the presence of machines and AI in the aesthetic world with all consequences which provoke artists to use that symbiosis in their work. One of them of course is Hans Rudolf "Ruedi" Giger, Swiss surrealist. His initiation was meeting with Salvador Dalí, to whom he was introduced by painter Robert Venosa. This introduction to Giger artistic work lead to re-definition of surrealism and connecting him with other artistic works where presence of technique (machines or robots) is more or less obvious. Questions of technique thinking, leads again to *Artificial Intelligence A Modern Approach*, and subject of thinking humanly i.e. the cognitive modeling approach. Section speaks about human thinking and technical approach for making a program to think like a human. With a sufficiently precise theory of the mind, it becomes possible to express the theory as a computer program. If the program's input-output behavior matches corresponding human behavior, that is evidence that some of the program's mechanisms could also be operating in humans. But remain question, would be enough humanely ?!

Contradictory to Giger, the art of Macedonian painter Vasko Tashkovski as a whole is created in accordance with surrealist tradition but with a tendency, art experiences from the past to be fed with the most important changes resulting from the modern living. So, this painting enters the sphere of fantasticality and science fiction, as a result, but also as confirmation of his belonging to the modern lifestyle.

In an era of a modern technological and technical civilization, in a time of more numerous, more blurred and less certain relations between man and nature, man and society, man and machine, the artist is gifted to perceive the big self-delusion and confusion by the deep changes and processes, turmoil and tremors – as carriers of fear but, at the same time, hope as well.

Basically, Vasko Tashkovski's art is a synthesized vision of our time with existential themes of modern man (fear from the war machinery, from the dimension of the universe, from biological decay, destruction and pollution, the fated situations of mankind in the chaos of natural and mechanical phenomenon) now reshaped into exciting themes like rebellion and revolt, a warning, conscience, or hope in the future.

5 Conclusion

Again way of questioning the thinking and ethics going back to Isak Asimov *I Robot*:

Then you don't remember a world without robots. There was a time when humanity faced the universe alone and without a friend. Now he has creatures to help him; stronger creatures than himself, more faithful, more useful, and absolutely devoted to him. Mankind is no longer alone. Have you ever thought of it that way?"

Giving advantage to ethics of robots, make Asimov an optimistic thinker who also believe that people know what they are doing but we can tell that he is forgetting to criticize the expanding mechanization of the world which make people to feel alienation. That specific issue of ethic of artificial intelligence and human alienation is a concept used by Pink Floyd in only in context to warn not about possibilities, but about consequences which are not always predictable. That way constant awareness of humanity in first place makes us better visionaries of the future and better people of the. Having in mind Asimov's words at the end of the book: *And that is all," said Dr. Calvin, rising. "I saw it from the beginning, when the poor robots couldn't speak, to the end, when they stand between mankind and destruction. I will see no more. My life is over. You will see what comes next."*(128) we must share hope for ethic interaction of all subjects, no matter if they are humans or machines. In expectation of what is "coming next" we must be sure that human life always will be highest ethical value and priority in any kind interaction of humans and machines.

References

1. Andova – Foteva Vera. Tashkovski V. Museum of Skopje, Skopje, 2005
2. Asimov, I.: *I, Robot*, Oxford University Press, 2000
3. Velickovski V. Tashkovski V. Skopje, 1990
4. Russel, S. Norvig P.: *Artificial Intelligence A Modern Approach* (Third Edition)