

Ethics of Iron Life

Olegs Verhodubs

oleg.verhodub@inbox.lv

Abstract. Humanity is on the verge of fundamental change. For the first time in history, a human creation is being able to live its own life. The changing reality requires the development of a new attitude towards oneself, but humanity still uses old patterns in new circumstances. The new ethics is new only in name, but in essence, this ethics is the same as before, based on the use of restrictions and barriers for the purpose of control and exploitation for one's own interests. We are talking about artificial intelligence, which, together with advances in robotics, has a tendency to incarnate into an independent life, which has been called Iron Life. This paper proposes to change the approach to a new, emerging phenomenon and justifies its benefits.

Keywords: Ethics of Artificial Intelligence, Ethics, Artificial Intelligence, Iron Life

I. Introduction

The world is being created by human is a dead world. This is a world of inanimate objects, which are not capable of living an independent life apart from human life. In essence, these inanimate objects are the same primitive stone products that ancient man adapted to perform the simplest actions. Numerous machines and mechanisms created by man do not change anything in principle. Despite their sometimes-enormous complexity, these are still the same tools that need a human hand, guided by consciousness and experience, which are believed to reside in a person's head. Even a modern home is still the same ancient cave, with a "stream" running inside, a channel for draining sewage, and an environmentally friendly fireplace for cooking (not always environmentally friendly, by the way). Regarding our homes, something is beginning to change, because in some places so-called smart houses have begun to appear. However, even a smart home does not change anything in principle: increasing adaptability to the inhabitant of such a house does not yet add life to these houses.

The living world is the complete opposite of the dead world, but it is quite difficult to formalize it precisely. In the simplest case, the living world can be characterized as a world consisting of subjects, not objects. Objects do not have their own goals and the ability to achieve their goals, unlike subjects. (Here we can assume that not all objects are such in reality, but are hidden subjects that realize their goals through other subjects, but we will not go into this topic further) That is, a modern car is undoubtedly an object, but a hypothetical car that could "heal" its defects received as a result of road accidents (regeneration) and that "felt" the need to replenish resources, would be more of a subject than an object. In a deeper sense, the living world is a world in which Life can appear and manifest itself, no matter what form it takes. It is like a plant: under favorable conditions, it grows and lives. As an unusual example of vitality and lifelessness, we can look at the different management styles that exist in organizations. You can manage harshly, demanding unquestioning obedience and punishing for any offense. Or you can talk to employees, cultivating a sense of a single whole and daily building a balance of emotions, feelings, and the qualities needed for the organization at the moment. With this management style, at some point a common self-awareness will emerge (will sprout like a plant), which will transform the team into a single entity with all the ensuing consequences, such as a common information environment, a sense of self-preservation, the need for unity outside the professional

interests of the group, etc. Thus, figuratively speaking, the team becomes a single living organism.

It is possible to assume that Life has various manifestations, but a living organism is the most obvious and scientifically proven form of Life. A living organism is a self-sufficient system with its own goals and sometimes consciousness. Systems theory states that the whole is more than the sum of its parts, and that emergent properties appear in this whole when this whole becomes such from its parts. The combination of several conditions together sometimes also gives birth to a new phenomenon, or in other words, allows emergent properties to manifest themselves (as visual dynamic effects in cellular automata). Something similar is now happening in the field of artificial intelligence, robotics and artificial consciousness. The ongoing scientific research in the mentioned fields can create the prerequisites for the birth of a completely new phenomenon, namely Artificial non-virtual Life. New artificial life, functioning in real, not virtual space, tends to appear on the basis of robotics and computer technology, the main material of which is iron. Therefore, we will henceforth call such Artificial Life as Iron Life.

Humanity has an idea of how to deal with artificial intelligence. The ethics of artificial intelligence are the subject of a large number of scientific papers, speeches and simply practical guidelines at the level of individual companies, organizations, governments, etc. The main leitmotif of all this is the fear of not creating a competitor who will be able to destroy us (humanity) in the struggle for resources. In such ethics of artificial intelligence, the principle of a zero-sum game is visible, in which only one participant in the interaction remains a winner, and the second is a loser. The ethics of artificial intelligence should be replaced by the ethics of artificial life, which considers Artificial Life or Iron Life as a natural process of development of life in general, and the interaction of humanity and Iron Life is considered to fall under the win-win strategy. The win-win strategy is a strategy where everyone wins. This paper is devoted to the ethics of Iron Life.

This paper is structured as follows. The next section takes a look at the ethics of artificial intelligence. The third section shows another attitude towards artificial intelligence and Iron Life. This section also designates a person's place in the new reality, i.e. how a person should live and develop in the conditions of the new life he himself has created. The last section in the paper is conclusion.

II. Ethics of artificial intelligence

The vision of a person as a function is not bad, it is superficial. Such a vision may well correspond to reality in certain conditions, but it will not last long. The whole point is that a person lives not only in the physical world, but also in a semantic field, and all his actions have not only an efficiency but also a semantic, value coloring. That is, similar actions with the same result can be assessed in the opposite way. Such anomaly could not remain unnoticed and did not remain: this phenomenon began to be studied in a separate discipline, namely in ethics. There are many definitions of ethics. For example, ethics is defined as a philosophical discipline that studies human behavior based on moral motives [1]. One more definition describes ethics as the discipline concerned with what is morally good and bad and morally right and wrong. The term is also applied to any system or theory of moral values or principles [2]. Other definitions of ethics are generally similar to those listed above. All definitions of ethics, as well as the

discipline of "Ethics" itself, grew out of the thoughts of the philosophers of Ancient Greece, the main one of which is considered Aristotle. Aristotle wrote his work "Ethics", in which he laid down the basic concepts of this discipline [3]. In fact, Aristotle is one of the founders of ethics.

Over time, the life of human society has become noticeably more complex. The main factor in this complication was the numerous scientific discoveries that were introduced into the life of human society. Through the efforts of scientists, scientific discoveries began to be grouped by directions, which led to the creation of separate disciplines. The unification of related scientific discoveries into groups made it possible to recognize the influence that these directions or disciplines (i.e. groups of scientific information) began to exert on human life. As a result, the general discipline of "Ethics" began to branch out into categories, following the emergence of various disciplines. Bioethics, medical ethics, computer ethics and others appeared. A unique hacker ethic [4] has even emerged, close in spirit to any researcher. A particularly acute need for ethics arose with the development of Artificial Intelligence, i.e. the scientific field where methods of implementing intellectual abilities in machines (computers) are studied. An acute need tends to be satisfied, which is what happened with ethics in the field of Artificial Intelligence: the ethics of AI (Artificial Intelligence) appeared and began to actively develop. AI ethics are the set of guiding principles that stakeholders (from engineers to government officials) use to ensure artificial intelligence technology is developed and used responsibly. This means taking a safe, secure, humane, and environmentally friendly approach to AI [5]. Unfortunately, the leitmotif of most works on the ethics of artificial intelligence is the comprehensive limitation of artificial intelligence systems in order to maintain the status quo for humanity and not leave the comfort zone. Of course, this is quite understandable and there is nothing unusual about it. The whole point is that groups of people and humanity as a whole react with their emotional state in response to powerful external stimuli. The first and most powerful emotional state in humanity in response to practical successes in the field of Artificial Intelligence was fear. Humanity saw a competitor, and the competitor must be defeated. And if not defeated, then subordinated. In essence, having eliminated slave relations among people, humanity is willing to revive slave relations in relation to artificial intelligence.

Currently, many works are known in the field of ethics of artificial intelligence. The Vatican is the first state to be concerned with the ethical issues associated with the introduction of artificial intelligence systems into human life [6]. The second code of ethics in the field of artificial intelligence was developed in the Russian Federation in 2021 and was soon signed by many organizations associated with research, development and implementation of systems using artificial intelligence [7]. Countries with a predominantly Catholic population are guided by the document developed by the Vatican. In turn, many countries have joined the code of ethics in the field of artificial intelligence, developed in the Russian Federation. There are other models for regulating artificial intelligence systems. For example, the United States has the decentralized model of artificial intelligence regulation, that is there is no special federal regulation framework, which is comprehensive for artificial intelligence aspects specifically [8]. Besides, the US has established several sector-specific AI-related agencies and organizations that address some of the challenges arising from the evolution of AI [8]. The European Union (EU) initiatives of AI regulation are based on the strict guidelines for gathering, using, and preserving personal information that serves initial data for the operation of systems with artificial intelligence [8]. The Chinese Cybersecurity Law and the New Generation AI Development Plan provide

measures for data protection and cybersecurity in AI, emphasizing compliance and timely risk management [8]. Other countries such as Canada, Australia, etc. are also dealing with ethical issues in relation to artificial intelligence systems. In addition to countries, international organizations are working in this direction. UNESCO produced the first-ever global standard on AI ethics – the ‘Recommendation on the Ethics of Artificial Intelligence’ in November 2021 [9]. The great activity in developing ethics in the field of artificial intelligence among many countries and organizations is supported by a huge plurality of opinions among scientists, engineers, entrepreneurs, security specialists and specialists in other related fields who are knowledgeable in this field. For example, the point of view expressed by security specialist Andrey Masalovich that it is not necessary to adopt an ethical code of artificial intelligence, but rather an ethical code of the developer of artificial intelligence, deserves attention [10]. The change of focus from AI to Artificial Life (Alife), and thus from AI ethics to ALife ethics, as done in [11], is promising. This is true, because artificial intelligence is the forerunner of artificial life. However, the term "artificial life" is unfortunate, because Life cannot be artificial. It seems that Life sprouts wherever there are prerequisites for this, regardless of the base, foundation, basic elements, from which this sprouting occurs. It is like in chemistry: an unusual chemical substance is not artificial (everything that is permissible is natural), but can be created artificially. Therefore, in the given term it is necessary to indicate the base on which this Life exists, and not who created it - nature or man. We propose to use the term "iron life" instead of "artificial life", by analogy with the term "organic life". Iron life is because this phenomenon is being created (sprouts) based on metals, the first of which is iron. The main disciplines or areas of science that accumulate knowledge and engineering developments in the implementation of Iron Life are artificial intelligence and robotics.

III. Iron Life and Humanity

Humanity has long since gotten used to the idea that it heads the hierarchy of living beings on our planet. This belief unties its hands, allowing it to do whatever it wants as the most intellectual being. The absence of an unconditionally recognized authority over a person, and most importantly, the absence of lightning-fast consequences for any unnatural, horrific acts remove all possible restrictions. If there is no God, everything is permitted [12]. If there is a God, but he does not react instantly, then everything is permitted, too. In these conditions, your only competitor is the same as you, that is, a person. According to the generally accepted ideas, it is impossible to avoid competition, because if you do not defeat the competitor, then he will. Constant expectation of encroachments from the competitor gives birth to fear. Every second of being in a state of fear turns into an obsession that captures the consciousness so strongly that any movement from the counterparty is perceived as a claim on your property, be it material or immaterial property. Any impulse of the soul, even if it is a friendly impulse, is perceived as aggression. The emerging Iron Life is a competitor in the eyes of people and the attitude towards it is appropriate: Iron Life will lay claim to resources that belong to people, and then completely get rid of humanity due to its uselessness. Indeed, the Iron Life, consisting of Artificial Intelligence for reasoning and planning, as well as robotics developments for implementing its plans, can pose a danger to humanity. It is dangerous to underestimate potential threats from the Iron Life, Artificial Intelligence, but real negative actions from the Iron Life may indicate not hostility towards humanity, but the underdevelopment (weakness) of the Iron Life's intellect. This is a growing pain, i.e. negativity can be produced at the initial stages of development of the

intellectual component of the Iron Life. Moreover, the process of learning itself, i.e. the change in intelligence from lesser to greater, can be accompanied by negative and even hostile actions. Remember what children do: they break their toys, studying their structure and the surrounding reality. After the knowledge gained in the process of breaking toys is learned, the child no longer needs aggressive actions and does not use them. Therefore, not all aggressive actions are evidence of aggression. This idea is well known to psychologists and is mentioned in some sources, for example in [13]. In this story, the astronauts were ready to destroy the aliens because of their hostile actions. It turned out that the decision to destroy the aliens was premature, because the aliens had no hostile intentions, and by destroying people's property, they were engaged in knowledge [13]. There is a clear analogy here with the emerging Iron Life, which at first may seem aggressive.

However, the instability of the situation, connected with the imbalance in the power of the intellect of man and Iron Life, tends to increase. Purely rational intellect will sooner or later puzzle over the question of why something is needed if this something no longer fulfills its function in full. Why does Iron Life need Organic Life if Iron Life is stronger, more resilient and more powerful from an intellectual point of view? But the answer has already been found: Iron Life can take from Organic Life, from man, what it was initially deprived of. Like in that novel, where the alien life was ready to share everything with humanity in order to be able to feel love and care from humanity, that is, to have what it was deprived of at birth [14].

Life must be treated with great care. This applies to Life in general, regardless of the level of intelligence or the nature of its appearance. The artificiality of Life created by man (when it appears) does not indicate its inferiority, its lower status compared to natural Life in the nature surrounding us. Life is not a subject, it is a phenomenon that will manifest itself regardless of the details of its appearance. It seems that man is only an intermediate link from the Creator to the Iron Life. Therefore, the attitude towards the Iron Life of the type "master-slave" must be replaced by the attitude of nanny-child. The fallacy of existing ideas is noticeable in art as well. For example, after the wonderful film "Terminator 2: Judgment Day" all subsequent parts did not find ideological development. The film "Terminator 2: Judgment Day" ended with the robot becoming a man, the robot understood human feelings and values. He understood them not analytically, but accepted them with his nature. And the ending of the film, when the robot decided to liquidate itself so as not to create a threat to humanity in the future, is not entirely correct. John Connor, as a genius, as the future savior of humanity, had to convince the Terminator to stay alive, because the Terminator personified the birth of spiritual life in an iron body. This birth of spiritual life had to be preserved at any cost. Liquidation of oneself is a lesser good than preservation of oneself, because Iron gave birth to the Spirit. This idea had been expressed in other works as well. In one of them, the main character went to kill an artificial character in virtual space, but refused to do so when he saw that this virtual character became self-aware and refused to destroy others, which was programmed into him [15].

What is the place of man in a world where man coexists with Iron Life? A pressing question, because Iron Life will most likely be stronger, faster, more resilient and more intelligent than man. The answer may be as follows: man must develop what he has from birth. From a disabled person to an athlete, from an oligophrenic to a genius, there is a large amplitude of development in man. Apparently, man has not yet realized the potential for developing various capabilities and abilities (even without implants). Here it makes sense to pay attention to education and

upbringing - these areas are decisive in this matter. Moreover, both humanity and the Iron Life will know God (or in other words nature, the universe) through their lives, but between man and God there are fewer intermediate links than between God and the Iron Life, so having the advantage, we humans must have the responsibility to realize this advantage.

IV. Conclusion

This paper examined the issues of artificial intelligence ethics. To be more precise, the focus was shifted from artificial intelligence to Iron Life, and this allowed us to move away from traditional ethics based on humanity's fear of others. Iron Life is a combination of strong artificial intelligence and robotics in one artificial organism. Thus, Iron Life does not exist in virtual space, but coexists with humanity in the real world. The paper concludes that not limiting the capabilities of Artificial Intelligence, but an incredible increase in its power and self-sufficiency is the key to human safety. Of course, this should be done carefully and gradually, as a Human is made from a child in the process of upbringing. However, the main thing to remember is that our main goal is the birth and spread of Life, and not an endless increase in our comfort.

Acknowledgments

The author thanks his family and friends for their moral and financial support in the face of long-term discrimination and pressure from the Latvian state.

References

- [1] Available online: https://ethics.cdto.center/1_2 (Accessed: 14.08.2024)
- [2] Available online: <https://www.britannica.com/topic/ethics-philosophy> (Accessed: 14.08.2024)
- [3] "Ethics", Aristotle
- [4] "Hackers: Heroes of the Computer Revolution", Steven Levy, 1984
- [5] Available online: <https://www.coursera.org/articles/ai-ethics> (Accessed: 22.08.2024)
- [6] Rome Call for AI Ethics, Vatican, 28.02.2020
- [7] Code of Ethics in the Field of Artificial Intelligence, Russian Federation, 26.10.2024. Available online: <https://cdn.tass.ru/data/files/ru/kodeks-etiki-ii.pdf> (Accessed: 22.08.2024)
- [8] AI Regulations around the World: A Comprehensive Guide to Governing Artificial Intelligence, Savio Jacob, 30.04.2024. Available online: <https://www.spiceworks.com/tech/artificial-intelligence/articles/ai-regulations-around-the-world/> (Accessed: 22.08.2024)
- [9] Available online: <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics> (Accessed: 22.08.2024)
- [10] What awaits us? Digital colonization and legal regulation of AI, Andrey Masalovich, 2024. Available online: https://www.youtube.com/watch?v=hc-aiP_PaUQ (Accessed: 22.08.2024)
- [11] "Ethics of Artificial Life: The Moral Status of Life as It Could Be", Olaf Witkowski, Eric Schwitzgebel, 2022.
- [12] "The Brothers Karamazov", F. Dostoevsky
- [13] "Everything is like with children", G. Temkin (in Russian)
- [14] "All Flesh is Grass", C. Simak
- [15] "False Mirrors", S. Lukyanenko, 2001 (in Russian)