İSTANBUL BİLGİ UNIVERSITY

INSTITUTE OF GRADUATE PROGRAMS

PHILOSOPHY AND SOCIAL THOUGHT MASTER'S DEGREE PROGRAM

AUTONOMY IN HUMAN AND ARTIFICIAL INTELLIGENCE AND ITS RELATIONSHIP WITH THE CONCEPTS OF DIGNITY AND HUMAN RIGHTS

Ayşe ACAR

118679001

Doç. Dr. Ferda KESKİN

İSTANBUL

2020

AUTONOMY IN HUMAN AND ARTIFICIAL INTELLIGENCE AND ITS RELATIONSHIP WITH THE CONCEPTS OF DIGNITY AND HUMAN RIGHTS

İNSAN VE YAPAY ZEKA'DA OTONOMİ VE OTONOMİNİN ONUR VE İNSAN HAKLARI KAVRAMLARIYLA İLİŞKİSİ

Ayşe ACAR

118679001

Tez Danışmanı: Doç. Dr. Ferda KESKİN …….

İstanbul Bilgi Üniversitesi

Jüri Üyeleri: **Dr. Öğr. Üyesi Zeynep Talay TURNER** …….

İstanbul Bilgi Üniversitesi

Prof. Dr. Ahmet Ayhan Çitil ……..

İstanbul 29 Mayıs Üniversitesi

Tezin Onaylandığı Tarih: 26.06.2020

Toplam Sayfa Sayısı: 70

Anahtar Kelimeler (İngilizce)	**Anahtar Kelimeler (Türkçe)**
1) Artificial Intelligence	1) Yapay Zeka
2) Autonomy	2) Otonomi
3) Kant	3) Kant
4) Human Rights	4) İnsan Hakları
5) Ethics	5) Etik

TABLE OF CONTENTS

SCHEMA LIST

ABSTRACT

Coming to the agenda as the fundamental subject in today's economy politics and technology, Artificial Intelligence separates itself from precursing technology elements with the concept of "autonomy''. The question by Alan Turing in 1950 asking "Can machines think?'' is viewed as the beginning of the efforts to make machines human-like. As autonomy, in Kantian philosophy, makes a reference to an intelligent ethical subject, the autonomy aimed at for machines is used for structures that will be reached at the stage of Human Level Artificial Intelligence, that have intelligence and are able to make independent decisions. The main question of this work is whether the problems arising in the realization of Kant's deontological ethics by a human subject can be overcome by another type of reason (artificial intelligence). Kantian approaches in Artificial Intelligence studies in cognitive sciences, and evaluation of Kant's deontological ethics and the subject of autonomy in terms of human rights and dignity in Machine Ethics will be examined. This work will deal with subjects in sections as follows, section 2.1 Autonomy from a Kantian Perspective, the Law of Freedom and Morality; section 2.2 Artificial Intelligence and Autonomy; section 3.1 "Life" in Kant's Critique of The Power of Judgment, Kantian Approach to Transcendental Mechanisms and Artificial Intelligence; section 3.2 The Problem of Applying Kantian Ethics to Structures with Artificial Intelligence; and section 4 Human Rights, Dignity and Autonomy.

Keywords: Artificial Intelligence, Autonomy, Kant, Human Rights, Ethics.

ÖZET

Günümüz ekonomi politiğinin ve teknolojinin temel meselesi olarak gündeme gelen Yapay Zeka, kendinden önceki teknoloji unsurlarından "otonomi" kavramı ile ayrılmaktadır. Alan Turing'in 1950yılında sorduğu "Makineler düşünebilir mi?" sorusu makineleri insan gibi kılma çabalarınınbaşlatıcısı olarak değerlendirilmektedir. Otonomi, Kant felsefesi açısından akıl sahibi etik bir özneyegönderme yaparken makinelerde hedeflenen otonomi İnsan Düzeyinde Yapay Zeka aşamasındaulaşılacak akıl sahibi, bağımsız karar alabilen yapılar için kullanılmaktadır. Çalışmamızın ana sorusunu Kant'ın ödev ahlakının insan öznesi tarafından hayata geçirilmesinde yaşanan sorunların bir başka zihin türü (yapay zeka) tarafından aşılıp aşılamayacağı oluşturmaktadır. Yapay Zeka çalışmalarında bilişsel bilimlerde Kantçı yaklaşımlar, Makine Etiğinde Kant'ın "ödev ahlakı" veotonomi meselesinin insan hakları ve onur (haysiyet) açısından değerlendirilmesi irdelenecektir. Çalışmanın 2.1 bölümde; Kant Perspektifinden Otonomi, Özgürlük ve Ahlak Yasası, 2.2 bölümde; Yapay Zekâ ve Otonomi, 3.1 bölümde; Kant'ta Yargı Gücünde Yaşam, Transandantal Mekanizmalarve Yapay Zekaya Kantçı Yaklaşım, 3.2 bölümde; Kant Ahlakının Yapay Zekalı Yapılara UygulanmaMeselesi ve 4.bölümde; İnsan Hakları, Onur (Haysiyet) ve Otonomi, konuları ele alınacaktır.

Anahtar Kelimeler: Yapay Zeka, Otonomi, Kant, İnsan Hakları, Etik.

INTRODUCTION

As much as a knife made by chiseling a stone at the beginning of human history does not resemble today's tools of technology, as an artificially created tool it is a product of technology. When the stone age human took the knife in his/her hand and used it as a tool this made the humans of the era cyborgs or even if we can't say that smoke signal communication creates a kind of a cyberspace - which from some perspectives we actually can, since this situation references to a need, and we can say that is an attempt to create a second nature in the face of nature by humans. If we interpret the concept "technique" not only as a "tool" but as "a way to reveal" we will come across the Ancient Greek concept of "Tekhne".[1] Coming from the Greek language the word "technique" makes reference to the skill and undertaking of a craft and at the same time in its relation with "poiesis", to an artistic situation of bringing to being.[2] At the same time in the context of Plato Tekhne has the property of "setting the power to see in motion".[3] When this meaning in the concept Tekhne reaches a certain strength, it references the word "diamekhanesastha".[4] "Mekhane" at the root of this word means mechanism, tool and constructed structure and is the root of the word "machine" we use today.[5] Tekhne defined as a human activity in Ancient Greece or - although it does not exactly represent the same thing by today's terminology, technology "is not what we seek, but how we seek".[6] Today what we search for through technology is Artificial Intelligence and the answer to how we seek it is by modelling the human intelligence.

Frederick Betz who said *"The ruthlessness of technological change stems*

[1] Martin Heidegger, *Teknik ve Dönüş*, Çeviren: Necati Açа, Bilim ve Sanat Yayınları, 1998, p.17.
[2] Martin Heidegger, *Teknik ve Dönüş*, p.18.
[3] Oğuz Haşlakoğlu, *Platon Düşüncesinde Tekhne*, Sentez Yayınları, 2016, p.51.
[4] Oğuz Haşlakoğlu, *Platon Düşüncesinde Tekhne*, p.51.
[5] Oğuz Haşlakoğlu, *Platon Düşüncesinde Tekhne*, p.51.
[6] Gerd Leonhard, *Teknolojiye Karşı İnsanlık*, Translated: Cihan Akkartal, İlker Akkartal, Siyah X Yayınları, 2018, p.41.

from its own power that no society can resist and that is called compulsory technology," remarks that science and economy are directly related through technology and that this interaction is the *field of technological innovation*.[7] As a field in technological innovation, Artificial Intelligence has been in our lives for a very long time. This field of innovation concerns not only disciplines like computer science and cognitive science but many other disciplines such as philosophy, psychology, politics and law. Developments in this field are both a source of curiosity and anxiety for every human.

Founding Director of Future of Humanity Institute, Oxford University Prof. Nick Bostrom, makes a warning that physical technology that today is supplementary to labor might become the replacement to labor in the chapter *Of Horses and Men*[8] of his work *Superintelligence* and gives horses as an example to make his point. In the beginning horse carriages and plows accompanied horses as a supplement and this was a big development that increased the productivity of horses. Later horses were replaced by automobiles and tractors. With time these innovations reduced the need for horse labor and the horse population plummeted. What was a 26 million horse population in 1915, became 2 million at the beginning of 1950s.[9] Bostrom tells that without doubt the most prominent difference between humans and horses is that humans have capital. Nobel Peace Prize laureate Al Gore reflects that in the past developments in economic productivity were reflected on wages, while today a much bigger part of the earnings will go to the investors.[10] The cumulative effect of the development speed of Artificial Intelligence causes workers to work for lower wages while causing more unemployment and even bigger inequality.[11] Artificial intelligence work that besides these negative developments caused many positive ones in many fields, especially the field of healthcare, began in 1950 with Alan Turing who is considered the founder of computer science and artificial intelligence. Although there have been stagnant periods called "artificial intelligence winter" since those years, today "human

[7] Frederick Betz, *Teknolojik Yenilik Yöntemi,* Tübitak Popüler Bilim Kitapları, 2010, p.1.
[8] Nick Bostrom, *Superintelligence*, Paths, Dangers, Strategies, Oxford University Press, 2014.
[9] Nick Bostrom, *Superintelligence*, p.161.
[10] Al Gore, *Gelecek*, Translated by: Çağlar Kök, Ebru Kızıldağ, MediaCat Kitapları, 2014, p.72.
[11] Al Gore, *Gelecek*, p.39.

level" and "beyond human level" artificial intelligence work have gained speed. While among these works approaches modelling the workings of the human intelligence by cognitive scientist creates the focal point, there are suggestions to apply Kant's transcendental mechanisms to artificial intelligence. While these suggestions are not the main subject of our work, they will be mentioned briefly to introduce Kantian approaches in Machine Ethics. The subject of whether the Kantian concept of morality can be a model for machines is already a current one in work on "Machine and Ethics". Kant's approach to the understanding of an object - in the second preface to the *Critique of Pure Reason* Kant tells that he has reversed the former definition of an object and that it is a *"Copernican Revolution"*[12]- also catches the attention of cognitive scientists in creating an artificial intelligence. Kant's understanding of morality that is thought to be based on computable judgement, caused the coinage of the concept "Kantian Machine". Kant's categorical imperative states; *"Act only according to that maxim whereby you can, at the same time, will that it should become a universal law."*[13] and it encompasses humans and all intelligent beings. On the one hand with the imperative encompassing not only humans but all intelligent beings and on the other with "autonomy", which is among the basic concepts of morality, being used today for machines besides humans leads us to the following thought; since the law of morality was designed by ignoring the emotional states and instinctual desires this creates problems for them to be implemented by "real human beings". Yet today's narrow scoped artificial intelligence structures do not possess instincts or emotional states. On the basis of this idea the question we will seek answer to will be; can Kantian moral law be applied to another kind of reason, that is structures with artificial intelligence?

[12] Immanuel Kant, *Critique of Pure Reason,* Introduction 1(Paul Guyer, Allen Wood), The Cambridge Edition Of The Works of Immanuel Kant, Translated and Edited by Paul Guyer and Allen W. Wood, Cambridge University Press, 1998, BXVİ, p.110.

[13] Immanuel Kant; *Critique of Practical Reason*, Introduction by Stephen Engstrom, Translated; Werner S. Pluhar, Hackett Publishing Company, 2002, 31, p.45.

1.1AUTONOMY IN HUMANS AND ARTIFICIAL INTELLIGENCE

1.1.1 Autonomy, Freedom and Moral Law from Kant's Perspective

When we look at the history of thought in terms of concepts such as "reason, metaphysics, morality, autonomy, and freedom…" we can clearly say that it can be examined in two stages; that is before Kant and after Kant. This dual separation makes itself apparent in the general philosophy of Kant as well. Kant is a systemic philosopher that makes distinctions, and then tries to create a relationship among them. In the conclusion to the *Critique of Practical Reason* he states; "*The starry heavens above me and the moral law within me. I see them before me and connect them immediately with the consciousness of my existence."*[14] This motto can be considered a summary of Kant's philosophy in general.[15] The starry heavens begins outside, in the world of the senses, where the human being is found. The law of morality however lets a human who is an intelligent being get to know himself/herself in a general and compulsory connection.[16] This opportunity is at the same time an opportunity for a relationship between the theoretical and practical uses of reason. In the third antinomy of the *Critique of Pure Reason* it is stated that according to the law of nature causality is not the only causality, and that to explain appearances there is a need to assume a causality through freedom.[17] Kant, in the preface to the second print of the *Critique of Pure Reason* (B) tells that in order to make space for the human freedom there is a need for epistemological distinction. By making space for freedom in this manner, a transition from theoretical intelligence to practical intelligence can be facilitated. According to Kant reason proves its own reality when it transitions to the practical stage as pure reason.[18] The practical use of reason is what takes us to morality. In Kantian philosophy of ethics, freedom and autonomy can be understood through their relationship with each

[14] Immanuel Kant; *Critique of Practical Reason*, 162, p.203.
[15] Immanuel Kant, *Critique of Pure Reason,* Introduction, XXI.
[16] Immanuel Kant; *Critique of Practical Reason*, 162. p.203.
[17] Immanuel Kant, *Critique of Pure Reason,* B 474, p.484.
[18] Immanuel Kant; *Critique of Practical Reason*, Preface, p.4.

other and the moral law. Putting autonomy at the basis of the value of a human and any rational being, Kant's autonomy serves the function of the categorical imperative of the moral law.[19] Establishing the basis of the moral law not on the natural disposition of human but as to encompass all rational beings, according to Kant[20] freedom coincides with autonomy in terms of will putting itself into law.[21] To understand what autonomy and consequently what as an autonomous being a human being means for Kant you need to take a look his "Moral Law".

The law examined in detail in *Critique of Practical Reason* and in *Groundwork of the Metaphysics of Morals* separate Kant from philosophers before him who did ethical analysis. In the process from the Ancient Greek to Kant the most important problem of ethics has been "the highest good" (summum bonum). The solution to the problem generally follows a "eudaimonistic" line.[22] Aristotle in the introduction to *Nicomachean Ethics*[23] defines good as *"something that is desired by everything"*. Every art, every research, every action and every choice desires the good. But what is the nature of this Good? Aristotle emphasizes that in an attempt to define the nature of good one must begin with what is known. However what is known, multiple goals and tendency created by singularities makes it hard to understand what this nature exactly is. Nevertheless he gives us the highest Good, the science of sciences "Politics".[24] When we look at what the telos, the good it is directed at is, we come across the concept of *eudaimonia*. In translations to other languages from Ancient Greek this concept passes as "happiness". Here what needs cautions is that neither Aristotle nor Plato use the *eudaimonia* concept to mean "happiness". For Aristotle *eudaimonia*, makes a reference to a form of activity instead of a psychological state.[25] The concept formed from the words *Eu* and *Daimon* means "in harmony with Daimon". Daimon is the supernatural power that mediates the relationship between Psukhe and what

[19] Immanuel Kant, *Groundwork for the Metaphysics of Morals,* Edited and translated by Allen W. Wood, Yale University Press, New Haven and London, 2002, p.157.
[20] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.171.
[21] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.189,190.
[22] Doğan Özlem, *Kant Üzerine Yazılar*, Notos Kitap, 2014, p.23.
[23] Aristotle's *Nicomachean Ethics*, Book 1, Translated, With An Interpretive Essay, Notes, And Glossary By Robert C. Bartlett And Susan D. Collins, The University Of Chicago Press, 2011, p.1.
[24] Aristotle's *Nicomachean Ethics*, Book 1, Chapter 10, p.18.
[25] Nazile Kalaycı, *Daimon'dan Eudaimonia'ya: Aristoteles'te Mutluluk*, Cogito, Aristoteles, Yapı Kredi Yayınları, 2014, p.265.

is self transcendent. Jean Pierre Vernant defines daimon as *"a spirit that resembles the divine and that strives to rejoin it (psyche)".*[26] i

The fact that this misunderstanding and confusion regarding the concept of *Eudaimonia* has not been resolved is a subject that requires examination in itself. When we say that Kant's own understanding of morality was based[27] on the notion that attributing it to an ideal that cannot be agreed upon such as happiness would be wrong, one might fall into errors in understanding what freedom is. It must be emphasized that the happiness Kant criticizes, is not eudaimonia but eudaimonism which is utilitarianism.[28]

In the *Groundwork for the Metaphysic of Morals* Kant states:

"In the natural predispositions of an organized being, i.e., a being arranged purposively for life, we assume as a principle that no instrument is to be encountered in it for any end except that which is the most suitable to and appropriate for it. Now if, in a being that has reason and a will, its preservation, its welfare—in a word, its happiness—were the real end of nature, then nature would have hit on a very bad arrangement in appointing reasoning for this creature to accomplish the aim."[29]

According to Kant if the purpose of a rational being was happiness then all the means to accomplish this goal would have been given to instinct, and nature would undertake not only the selection of the goals but the selection the tools as well and would leave both through a wise foresight solely into the hands of instinct. Whereas however much a cultivated reason would try to enjoy life and happiness it would not get genuine satisfaction. As a matter of fact this situation after a time will bring on a hate of reason (misology).[30]

[26] Jean Pierre Vernant, *Eski Yunan'da Mit ve Din*, Translated by: Murat Erşen, Alfa Yayınları, 2016, p.88.
[27] Doğan Özlem, *Kant Üzerine Yazılar*, p.23.
[28] Kojin Karatani, *Transkritik,* Translated by: Erkan Ünal, Metis Yayınları, 2017, p.152.
[29] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.52.
[30] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.10, 54.

The main debate of Kantian morality is the issue of freedom in its whole, instead of happiness and good and evil as a tendency. We can't speak of good and evil where there is no freedom.[31] As much as transcendental freedom has put itself forth as a concept that can be thought about in theoretical reason it could not gain objective reality. Here transcendental freedom stands like a necessary condition to practical freedom.[32] When/if this concept is put forth together with a compulsory law of practical reason it functions as the cornerstone of Kantian philosophy.[33] When in the *Critique of Pure Reason* Kant starts from the senses going into principles, in practical reason he emphasizes that he will begin not from objects but from principles going into concepts and from concepts to senses. Because the subject of practical reason lies not in the relation of reason and objects, but in its relation with will and the causality of will. Kant states; we must, then, begin with the principles of a causality not empirically conditioned."[34] While Kant's practical philosophy contains in a sense a limitation to empirical reason, it at the same time contains a view on the human nature.[35] Allen W. Wood, emphasizes that while his evaluation of practical anthropology as an indispensable section of philosophy of morality in Kant's *Groundworks for the Metaphysics of Morals* rarely receives any praise, it is a very important detail.[36] Kant states that; one can call all philosophy, insofar as it is based on grounds of experience, empirical, but that which puts forth its doctrines solely from principles a priori, pure philosophy, and when pure philosophy is merely formal, it is called logic; but if it is limited to determinate objects of the understanding, then it is called metaphysics.[37] Physics will thus have its empirical but also a rational part; and ethics likewise; although here the empirical part in particular could be called practical anthropology, but the rational part could properly be called morals.[38] Kant is seeking an answer to whether there is possibility for the existence of a pure philosophy of morality cleansed of what may belong to anthropology.

---

[31] Kojin Karatani, *Transkritik,* p.153.
[32] Paul Guyer, *Kant's System of Nature and Freedom*, Selected Essays, Oxford University Press, 2005, p.121.
[33] Immanuel Kant; *Critique of Practical Reason*, Preface, p.4.
[34] Immanuel Kant; *Critique of Practical Reason*, Introduction, 16, p.24.
[35] Allen W. Wood, *Kant*, Blackwell Publishing, Edited by Steven Nadler, 2005, p.129.
[36] Allen W. Wood, *Kant*, p.132.
[37] Immanuel Kant, *Groundwork for the Metaphysics of Morals,* p.40.
[38] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.39.

Criticizing the evaluation of practical reason as the ability to find the most correct thing to do by employing the laws and connection of what is happening to reach happiness, Kant as a response to the dogmatism of a range of experience dependent on senses makes a distinction in the concept of "practical reason"; Technical Practical Reason and Pure Practical Reason. In Technical practical reason the human reason is capable of understanding the causal connections in true information. It can think of and locate tools to reach the goal and can direct the necessities and desires that direct it towards the goal. These necessities and inclinations unite in the intention of happiness.[39] In Pure Practical reason however, what directs the will is not inclinations and desires but the reason itself. Reason directs through *a priori* laws and principles. Yet the obligation here contains necessity. This is a law that is de facto and universal like the laws of nature but this time not for the nature, but transformed for moral life. While the law of nature is something that "is", the moral law is designed as something that "should be".[40] If a law is to become a moral law and carry liability, it will find the causation of liability not in the natural state of the human being nor in the state of the world but inasmuch as it is found in the concepts of reason a priori.[41] At this point "freedom" that cannot be presented empirically but which is a causality concept steps in. If freedom can be presented as a desired quality that is specific to humans and even all rational beings, this evidence will show that pure reason can be practical. [42]

Freedom is the *ratio essendi* of moral law. The moral law on the other hand is the *ratio cognoscendi* of freedom.[43] Kant states that freedom, among all the ideas of speculative reason, is also the onlyone whose possibility we know a priori—though without having insight into it —because it is the condition of the moral law, which we do know.[44] Kant gives us a warning; theoretical reason will not expand in

[39] Heinz Heimsoeth, *Kant'ın Felsefesi*, Translated by: Takiyettin Mengüşoğlu, Doğu-Batı Yayınları, 2007, p.121.
[40] Doğan Özlem, *Kant Üzerine Yazılar*, p.24.
[41] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.39.
[42] Immanuel Kant, *Critique of Practical Reason*, Introduction, 16, p.24.
[43] Immanuel Kant, *Critique of Practical Reason*, Preface, 4, p.5.
[44] Immanuel Kant, *Critique of Practical Reason*, Preface, p.4.

terms of knowledge through freedom, only a possibility will become confirmed in the practical use of reason.[45]

To understand the possibility of the law we need to examine the practical principles that appear in the analytic of pure practical reason. Practical principles, are propositions encompassing several practical rules. They are subjective, or maxims, if the condition under which they apply is regarded by the subject as valid only for his will; but they are objective, or practical laws, if the condition is cognized as objective, i.e., as valid for the will of every rational being.[46] But the existence of practical laws can be possible with the assumption that pure reason carries a reason sufficient to determine will. When the rational being is stimulated by their passions in itself the practical laws and maxims conflict.[47] The problem with the implementation of the law is exactly here. The theoretical use of reason determined by the structure of the object in natural study, knows the laws of nature as principles. Whereas in practical knowledge the principles a human gains are not obligatory laws. In practical knowledge reason is in a relationship with the will, and sees action as a goal. Because of the structure of the faculty of will the rule may take on different directions. Yet for a being whose will is not determined by reason alone, a rule becomes an imperative.[48] An imperative, brings about objective mandate in taking an action. When reason determines the exact will, the action inevitably happens according to a rule of an "ought". Imperatives are either the effecting reason (conditional imperatives-skill imperatives), or viewed without regard for sufficiency for effect to determine only the will (practical law).[49]

Kant divides freedom into two as positive and negative.[50] While negative freedom is tied to "conditional imperative" in the field of tendencies, positive freedom, which means autonomy, is tied to "categorical imperative". The autonomy of will is the only principle of all moral laws and duty that obeys the laws. While the conditional imperatives are numerous, a single expression, "moral law", is valid

[45] Immanuel Kant, *Critique of Practical Reason*, Preface, 5, p.7.
[46] Immanuel Kant, *Critique of Practical Reason*, 19, p.29.
[47] Immanuel Kant, *Critique of Practical Reason*, 19, p.30.
[48] Immanuel Kant, *Critique of Practical Reason*, 20, p.30.
[49] Immanuel Kant, *Critique of Practical Reason*, 20, p.31.
[50] Immanuel Kant; *Critique of Practical Reason*, 43, p.61.

as a law in a categorical imperative. Kant designs "goodwill" as a single a priori form of reason that encompasses all imperatives, and the will is its manifestation.[51] For Kant, human beings are bound by the laws of causality of nature with all their needs, motives, impulses and affects as a natural existence in the universe, and as natural existence, human is in heteronomous position and not free.

Being happy is necessarily the desire of every rational but finite being. Having a finite structure is decisive. The finite being has wants and the wants are related to the contents of the ability to desire. The reason for determining these contents can only be known empirically by the subject. Happiness is the name given to subjective determining reasons.[52] Happiness is determined by everyone's subjective sense of pleasure and pain. It even exhibits changes in the same subject, so a law can never be decisive.[53] When we separate a law from its entire content, it will remain the form of creation of a general law.

Kant states;

"If a rational being is to think of his maxims as practical universal laws, then he can think of them only as principles that contain the determining basis of the will not by their matter but merely by their form."[54]

A will which is such that the mere legislative form of a maxim can alone serve it as a law is a free will.[55]

"Act only according to a maxim that you would will to become a general law. Act so that you use humanity, as much in your own person as in the person of every other, always at the same time as end and never merely as means.So act as if the maxim of your action were to become through your will a universal law of nature."[56]

[51] Heinz Heimsoeth, *Kant'ın Felsefesi*, p.126.
[52] Immanuel Kant, *Critique of Practical Reason*, 25, p.38.
[53] Immanuel Kant, *Critique of Practical Reason*, 26, p.39.
[54] Immanuel Kant, *Critique of Practical Reason*, 27, p.40.
[55] Immanuel Kant, *Critique of Practical Reason*, 29, p.42.
[56] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.117, 139.

Maxims, which are the subjective principle of will, bring freedom as a prerequisite. Freedom is a human's intentional creation of law for himself/herself by using his free will, that is, autonomy (Auto: self, Nomos: law).[57] But what does Kant mean by freedom?

Freedom is neither something we are directly conscious of the first concept is negative-nor something we can deduct from an experiment. By giving us the law of appearances as natural mechanism, experiment stands against freedom.[58] In a sense, freedom that looks like "a thing in itself" does not exist in its narrow sense, there are reasons that determine all actions. But when freedom exists as a regulatory principle of reason, it exists as the cause of all actions.[59] A being determined by things other than itself can only be free when it acts as a willing existence. It can be free because it has autonomy. Autonomy means being free in the sense of having the law within the self, as well as being a condition of morality.[60]

The principle law of pure practical reason tells us that;

"So act that the maxim of your will could always hold at the same time as a principle of a universal legislation."[61]

This basic law is an categorical imperative. The will is absolute and objective here. The will was determined by the mere form of the law, regardless of the empirical conditions.[62] Kant says that this is an a priori proposition that is not based on any empirical view. Pure reason manifests itself as a legislator. This reason, which is practical on its own, gives people the "moral law". Kant includes not only human beings under the law, but also all sensible finite beings and even the "infinitive being as supreme intelligence"."[63] Will in rational beings is a will that does not conflict with the moral law. Reason and its objective law lead us to action,

[57] Doğan Özlem, *Kant Üzerine Yazılar*, p.25.
[58] Immanuel Kant, *Critique of Practical Reason*, 30, p.43.
[59] Kojin Karatani, *Transkritik*, p.161.
[60] Doğan Özlem, *Kant Üzerine Yazılar*, p.27.
[61] Immanuel Kant, *Critique of Practical Reason*, 31, p.45.
[62] Immanuel Kant, *Critique of Practical Reason*, 31, p.45
[63] Immanuel Kant, *Critique of Practical Reason*, 32, p.47.

that is, to "duty."[64] Since the moral law is not a law of nature, it does not bear the imperative found in laws of nature. Compliance with the moral law is a "duty" rather than an obligation. Duty is the imperative that we undertake to fulfill with our own will and bare the responsibility of. While we speak of an autonomy in duty, there is heteronomy in a task. When the will goes outside itself, it is inevitably heteronomous.[65] A task is what is commanded to us by an authority (state, institution, family) outside of ourselves.[66]

Stating that he does not take into account the views that evaluate the definition of duty as *"a made up concept that transcends the limits of selfimagination and arrogance,"*[67] Kant says:*" From love of humanity I will concede that most ofour actions are in conformity with duty; but if one looks more closely at "the imagination of the thoughts of their hearts," then everywhere one runs into the dear self, which is always thrusting itself forward; it is upon this that the aim is based, and not on the strict command of duty, which would often demand self-renunciation."*[68]

Autonomy of the will is the sole principle of all moral laws and of the duties conforming to them.[69] Objectively, we speak of a practical law and subjectively we speak of a pure respect for this practical law (even if it will harm all our tendencies). It is this sense of "respect" that determines will.[70] By saying; *"One could accuse me of merely taking refuge behind the word respect in an obscure feeling instead of giving a distinct reply to the question through a concept of reason,"* Kant explains what he understands[ii] from the concept of "respect". Respect differs from other emotions as an automatically experienced emotion. Respect is the awareness that my will is under a law and manifests itself as the effect of the law on the subject. The object of respect is the law we accept only by ourselves, as if it is compulsory. We submit to the law without consulting our self

[64] Immanuel Kant, *Critique of Practical Reason*, 32, p.47.
[65] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p. 169.
[66] Doğan Özlem, *Kant Üzerine Yazılar,* p.27.
[67] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p. 83.
[68] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p. 83.
[69] Immanuel Kant, *Critique of Practical Reason*, 33, p.48.
[70] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.66.

love, and the submission is a result of our will. When it comes to respect for a person, it is respect for the law (honesty, etc.) that the person proposes us, not the person per se. What Kant says is that *"All so-called moral interest consists solely in respect for the law"*.[71]

Kant, who frequently emphasizes that a rational being belongs to both the world of senses and the world of thought, and that it can be independent of the reasons stemming from belonging to the world of senses with the side that belongs to world of thought, states that this being as a rational being belongs to the world of understanding. The ability to understand, on the one hand, includes the foundation of the world of senses and its laws, and on the other hand, it opens its doors to this being as the rational being thinks of itself as a free being. Precise imperatives are made possible a priori by the fact that the will is above the will which is stimulated by sensory desires.[72] Desire here is pure will, and pure will involves a causality with freedom. Although the concept of an empirically unconditional causality is theoretically empty, it offers a possibility in its relationship with the undetermined object. It has practical application, that is, practical reality, which occurs concretely in intentions or maxims.[73]

For Kant who said *"Now I assert that we must necessarily lend to every rational being that has a will also the idea of freedom, under which alone it would act,"*[74] will refers to "autonomy" by being a law on its own.[75] The moral law determines will according to the principle of autonomy, and thus the law is based on the autonomous existence of the rational being as a free will. This foundation manifests itself as the value of a human and any rational being. According to Kant's perspective, by talking about an autonomy, we talk about a morality, and it is important to remember that freedom is a necessary prerequisite for morality.

[71] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.68, 69.
[72] Immanuel Kant, *Critique of Practical Reason*, 65, p.87.
[73] Immanuel Kant, *Critique of Practical Reason*, 56, p.76.
[74] Immanuel Kant, *Groundwork for the Metaphysics of Moral*, p.184.
[75] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.167.

So how will human beings make judgments according to moral law? According to Kant's method, it is necessary to start by observing both our own actions and the actions of others with a natural occupation, making it a habit and questioning the compliance of the actions with the moral law while observing.[76] These observations will enable us to distinguish the law that gives us a reason for obligation only from the law that actually gives obligation. With time the awareness of freedom will join our attention that is focusing on whether our actions are carried out not only for the sake of moral law, but also whether they have a moral value as an intention according to the maxims. The awareness of freedom will undoubtedly bring with it respect to ourselves.[77]

Reason that creates laws for itself, that is autonomous, is the reason that will grasp a *"realm of ends"*. Kant in the Groundwork for the Metaphysics Morals says; *"By a realm, however, I understand the systematic combination of various rational beings through communal laws."*[78] All rational beings will never treat themselves and other rational beings as mere tools, but will act as purposes on their own. In the actions of each member of this realm, which is possible with the freedom of will, the maxim of their wills also be valid as a general law.[79]

Among the rational beings who are members of this realm design, which gives the impression of an utopia, human also looks like a design. Kant's human being is something that exists by being free. When talking about a human with freedom, reason and autonomy; emotions, subjectivity, and instinct are completely left out and therefore the possibilities of morality and autonomy have been narrowed. The most common criticism of Kant's moral philosophy is that this moral understanding is a point of view that does not match human reality and reduces it to a mere mechanism. There are also criticisms reflecting that Kant's morality is not historical. Kojin Karatani states that Kant's moral theory is historical in essence, and the criticisms made in this regard do not take into account the "tool-

[76] Immanuel Kant, *Critique of Practical Reason*, 159, p.199.
[77] Immanuel Kant, *Critique of Practical Reason*, 161, p.202.
[78] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.150
[79] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.150.

purpose" issue.[80] *"Act each time to treat humanity not only as a tool but also as a purpose in everyone else's person as well as in one's own person."* Karatani says that this statement in Kant does not ignore the natural historical process.[81]

When the historicity of humans is taken into account, we also need to take into account that the developments that shape the question of "what is a human?" after Kant, significantly affect the answers given to this question. Evaluations on human and its historicity through the theory of evolution and philosophy of language provide remarkable information. However, even though Kant's philosophy is not aware of these developments, it still continues to be important in current debates about the human and the human reason. Kant's removal of the human as a being of reason and a being of morality from its natural reality, and his design of the human as an X remains a real problem. However, Kant's inclusion of not only people, but all rational beings in the definition of reason and morality, allows us to raise the question of whether this X design is applicable to another type of reason, for example, artificial intelligence referred to together with the concept of autonomy.

1.1.2 Artificial Intelligence and Autonomy

Computer scientist Edward Fredkin, one of the pioneers of Digital Physics, to indicate the importance of artificial intelligence states; *"There are three major events in history. The first is the formation of the universe. The second event is the beginning of life. The third event, as important as the first two, is the emergence of artificial intelligence."*[82]

John McCarthy, who invented the term "Artificial Intelligence", insisted that it would be a bad philosophy to move the artificial intelligence apart from philosophy while bringing up the common points between artificial intelligence and

[80] Kojin Karatani, *Transkritik*, p.170.
[81] Kojin Karatani, *Transkritik*, p.170.
[82] Jack Copeland, *Artificial Intelligence, A Philosophical Introduction*, Blackwell Publishers, 1993, p.1.

philosophy.[83] Although not in the modern sense but referring to autonomy, the dream of making a machine that does human tasks for the humans, or even thinks for them, goes back quite a long time in the history of philosophy. Aristotle, who is one of the most important names in the history of thought, thinks of automation by saying in *The Slave As A Tool* section of *The Politics "Let's assume that every tool we have can fulfill its duty, either by our proposition or by perceiving the need…"*[84] but concludes that it is impossible and emphasizes that slavery is essential so that people have free time.

Years after Aristotle in the 17th century, Gottfried Wilhelm Leibniz imagines a kind of intelligent machine through the system he calls "calculus ratiocinator"[85] based on the idea that human reasoning is a kind of computation. By 1950, Alan Turing's, who is accepted as the father of computer science and the founder of Artificial Intelligence[86iii], question *"Can Machines Think?"* comes to the agenda. Turing says in the *"The Imitation Game"* section of the article published under the title *"Computing Machinery And Intelligence"*[87] in the philosophy magazine *Mind:*

"I propose to consider the question, 'Can machines think?' This should start by defining the meanings of the concepts of 'machine'and 'think'."

Turing accepts that the thinking activity of the machines will be found odd and proposes the *"the imitation"* game to show what the concepts of "thinking" and "machine" are. Through this game, Turing will be able to get the answer to the question he asked at the beginning of the article, and the answer is that *"Machines can think."*

---

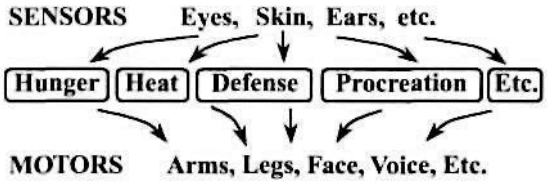
[83] John Mc Carthy, "What has AI in Common with Philosophy?" Stanford University, http://jmc.stanford.edu/articles/aiphil/aiphil.pdf

[84] Aristotle, *The Politics*, Translated by T. A. Sinclair, Revised and Re-Presented by Trevor J. Saunders, Penguin Books, 1992, p.65.

[85] Nicholas Rescher, "Leibniz's interpretation of his logical calculi," https://www.cambridge.org/core/journals/journal-of-symbolic-logic/article/leibnizs-interpretation-of-his-logical-calculi/611475C44E0E52824948975946E6FF46#

[86] Cem Say, *Yapay Zekâ*, Bilim ve Gelecek Kitaplığı, 2018, p.28.

[87] A.M. Turing, *'Computing Machinery and Intelligence' from Mind LIX*, no, 2236 (1950): 433-60. Reprinted by permission of Oxford University Pres; The Philosophy of Artificial Intelligence, Edited by Margaret A. Boden, (2005), p.40.



[88]

While in the most general sense computer is *"the name given to all systems that can calculate,"*[89] this definition also refers to the fact that a human is a kind of computer too.[90] In cognitive sciences, where the definition of machine based on human and the definition of human based on machine is intertwined, although the machines that are predicted to perform every cognitive activity that human beings do, have not yet uttered Descartes' *"I think, therefore I am,"* in theory there is nothing to stop them. Because these machines, which are defined by autonomous systems, are aimed to be assets that are independent in behavior and therefore working and thinking without external intervention.[91]

Thinking machines are referred to as *"intelligent"* because they have the opportunity to show willpower in action and achieve goals. Intelligence is a complex phenomenon consisting of various cognitive states such as perception, learning, reasoning, problem solving, memory, and comprehension.[92] Max Tegmark in his work *Life 3.0* describes intelligence as *"ability to accomplish complex goals are considered good or bad"* and states that it may have different forms, and there will be different intelligence tendencies in artificial intelligence machines, just like in human beings.[93]

Simulating human mental processes in Artificial Intelligence technologies determines the main framework. John Searle criticizes this framework; it is

[88] Marvin Minsky, The Emotion Machine, Simon & Schuster, 2006, p.21.
[89] Cem Say, *Yapay Zeka*, p.43.
[90] Cem Say, *Yapay Zeka*, p.43.
[91] Peter W. Singer, *Wired For War: The Robotics Revolution and Conflict in the Twenty-first Century*, The Penguin Books, 2010, p.74.
[92] Aziz Fevzi Zambak & Roger Vergauwen, "Artificial Intelligence and Agentive Cognition: A Logico-Linguistic Approach", *Logique et Analyse,* Vol: 52, pp. 57-96, 2009. (ISSN: 0024-5836)
[93] Max Tegmark, *Life 3.0, Being Human in the Age of Artificial Intelligence*, Borzoi Book by Alfred A. Knopf, 2017, p.51, 52.

possible to simulate a formally defined process on the computer, just like simulating the rain, but simulating mental processes is really thought to be a mental process when nobody thinks that the simulated rain will wet things.[94]

Criticisms and defenses in Artificial Intelligence work are basically manifested as "symbol - meaning" discussions. Allen Newell and Herbert A. Simon stated in their article titled "Computer Science as Empirical Inquiry: Symbols and Search", that the "Physical Symbol Hypothesis" gives an architectural view of the nature of all intelligent systems.[95] The symbol system hypothesis draws attention to the symbolic behavior of people and states that these behaviors have the characteristics of the human physical symbolic system, and they say that the symbols are the basis of activities involving intelligence.[96]

Hubert L. Dreyfus, who interprets Cognitive Science studies through the hypothesis of Newell and Simon's physical symbol system and preferred to call these studies "cognitivism" instead of cognitive science, states *"That any device, whether it is a computer or a brain, makes sense by storing and operating the physical equivalents of symbols."*[97] However, new studies show that the process of processing images in the human mind is analog and not manipulating the symbols in a separate way.[98] According to Dreyfus, what the physical symbol system hypothesis reflects is that a human is a kind of an "information processing device".[99] The definition of "meaning" in the hypothesis is constructed independent of the embodied experience acting on the world.[100] However, embodiment has a decisive role in the existence of contexts that determine meanings.[101] Brian Cantwell Smith says that in Simon's critique, the meaning is explained as follows;

---

[94] John Searle, "Bilgisayarlar Düşünebilir mi?", Translated by: Liz Amado, Cogito Dergisi, Yapay Zeka, 1998, Yapı Kredi Yayınları, p.63.
[95] Allen Newell ve Herbert A. Simon "Computer Science as Empirical Enquiry: Symbols and Search", the Tenth Turing Lecture, first published in Communications of the Association for Computing Machinery 19 Inc. (Mar. 1976) Reprinted by permission of Oxford University Pres; *The Philosophy of Artificial Intelligence*, Edited by Margaret A. Boden, (2005), p.111.
[96] Allen Newell ve Herbert A. Simon "Computer Science as Empirical Enquiry: Symbols and Search", p.117, 107.
[97] Hubert L. Dreyfus, "Simon's Simple Solutions", Stanford Education: https://web.stanford.edu/group/SHR/4-1/text/dreyfus.commentary.html
[98] Hubert L. Dreyfus, "Simon's Simple Solutions": https://web.stanford.edu/group/SHR/4-1/text/dreyfus.commentary.html
[99] Hubert L. Dreyfus, "Simon's Simple Solutions".
[100] Hubert L. Dreyfus, *What Computers Can't Do?,* Harper & Row Publisher, 1972, p.101.
[101] N. Katherine Hayles, "Anlamın Bedenleşmesi" Cogito Dergisi, Yapay Zeka, p.202.

"The symbols and their structures in the memory of the reader, the symbol structures of the skull, the process that takes place in a set of neurons called the brain".[102]

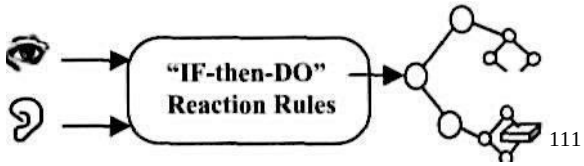
Stating that Simon refers to the non-referential language, Smith implies that there is deliberate blurring in the views of meaning in the text.[103]John Searle, who defines views that compare the human brain to a digital computer as the "strong artificial intelligence" view, in his article "Can computers think?" states that this view reflects that *"If we build a computer powered by a windmill made from old beer cans, it can gain intelligence through correct programming."*[104] Searle states that the main question to be asked is *"what digital computers actually are"* and that the operations of digital computers are determined purely formally, the steps of these processes are determined by abstract symbols, but the symbols do not have any meaning or semantic content. When we say mind, a structure with content rather than formal structures is mentioned. According to Searle's "Kansas City" example, when we think of Kansas City, we are talking about content beyond formal situations. Even though our thoughts are formed in the order of symbols (which do not have any meaning in themselves), they carry content.[105]

Searle[106], who says that computers only have a syntactic structure (that they are not syntactic) is an indication that they do not have a mind, suggests an intellectual experiment called the "Chinese Room" for a better understanding of the subject.[107] *"Imagine I was closed in a room for the experiment and I was given a lot of writing in Chinese,"* he says. The person who is closed in the room does not speak Chinese, nor does he know the spelling differences between Chinese and Japanese. But he has a book that says what to do with Chinese symbols. The rule book is in English. The person in the room who knows English understands the following; he will be able to use syntactic symbols, but will not be able to learn their meanings, that is, he will not understand those symbols. What the person who

[102] Brain Cantwell Smith, "Sap ve Saman Meselesi", Translated by: Elif Özsayar, Cogito Dergisi, Yapay Zeka, p.224.
[103] Brian Cantwell Smith, "Sap ve Saman Meselesi", p.228.
[104] John Searle, "Bilgisayarlar Düşünebilir mi?", p.58.
[105] John Searle, "Bilgisayarlar Düşünebilir mi?", p.59.
[106] John Searle, "Mind, Brains, and Programs", *The Philosophy of Artificial Intelligence*, Oxford University Press, p.69.
[107] John Searle, "Mind, Brains, and Programs", *The Philosophy of Artificial Intelligence*, p.69.

responds to all the instructions given in this Chinese room according to syntax rules does is not to speak Chinese, it is only to process inputs according to the rules, just as computers do.[108]

The fact that people think that the system understands Chinese when the system looks like it understands Chinese is commonly due to the fact that the situation is evaluated from a "behaviorism" perspective.[109] Cem Say says that an "input/output" program like in the Chinese Room is not a remarkable intelligence program, and that a chat program that will give the impression of intelligence should create complex real world models by matching the models in the dialogue.[110] Say's criticism seems to correspond more to the form of establishment of the experiment, whereas what Searle emphasizes is the difference between symbol and meaning, and Searle's criticism still remains valid in the narrow scoped artificial intelligence phase we are in today. Explanations by behaviorism, which say speaking Chinese and knowing Chinese, falling in love and pretending to be in love are the same, show that this school has no intellectual depth on the concepts.



111

The functioning of the human brain is of great importance in artificial intelligence approaches that model the functioning of the human brain. The first design of the human brain is based on the compressed human genome.[112] When residual data is extracted from the entire genome of eight million bits, between thirty and one hundred million bits of information remain, and this is a lower value

[108] John Searle, "Mind, Brains, and Programs", *The Philosophy of Artificial Intelligence*, p.69.
[109] John Searle, "Bilgisayarlar Düşünebilir mi?", p.63.
[110] Cem Say, *Yapay Zeka*, p.164.
[111] Marvin Minsky, *The Emotion Machine*, Simon & Schuster, 2006, p.20.
[112] Ray Kurzweil, *İnsanlık 2.0,* Alfa Basım Yayın Dağıtım, Translated by: Mine Şengel, 2016, p.213.

than Microsoft Word software. Even if the epigenetic knowledge capacity of man is added, the comparison remains valid.[113]

However, the fact that the original state of the brain is too complex to be reduced to a scheme in human brain modeling continues to be the main problem faced by the approach based on the human mind in modeling. In the human brain, circuits are slow, but each information works simultaneously in parallel. The brain can combine analog and digital phenomena and renew itself. While the brain uses the developing features, it patterns the information and the brain is holographic. The aim is to artificially recreate the functioning of the human brain and then move on to the next stage that transcends human intelligence. Initially, while the pioneers of artificial intelligence said that there could be a human level artificial intelligence, they did not believe that artificial intelligence could be more intelligent than human. In the proposal text to the Rockefeller Foundation, which would finance the research of ten scientists who came together for the historical "Dartmouth Summer Project" meeting held in 1956 it read;

"An attempt will be made to find how to make machines that use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.."[114]

Although we cannot say that there is a very thick line between the "software" that has actively entered into our lives today and artificial intelligence, all the systems (classifiers, search algorithms, planners, solvers and representative frameworks etc.) currently used are narrowly targeted, and it should be noted that *Human-Level Machine Intelligence* - HLMI is not yet realized, but it is at a preliminary stage that contributes to this goal.[115] Evaluating the defeat of the chess program Deep Blue, developed by IBM, which has an important place in the recent history of Artificial Intelligence (1997), to chess champion Garry Kasparov, Güven Güzeldere draws attention to the fact that this success occurred in a *"closed off*

[113] Ray Kurzweil, İnsanlık 2.0, p.214.
[114] Nick Bostrom, *Superintelligenc , Paths,* Dangers, Strategies, Oxford University Press, 2014, p.5.
[115] Nick Bostrom, *Superintelligence,* p.16.

world".[116] In a computer, a game takes place in an abstract world. The computer does not know *"where, when and for what purpose the game is being played, from which item or what shapes the chessboard or pieces are made of"*.[117] The computer does not know whether the human player is sleepless from the night before and what he will feel if he loses the game. The computer plays the game in a world completely closed off from the world.[118] Despite this fact, attention has been turned more into performance, according to Güzeldere, who says that the interest in machines playing chess is caused by *"people approaching to chess as 'a collision of opponents in the arena of minds' "*.[119] As in the case of Deep Blue, there is no mention of human level artificial intelligence.

In Ray Kurzweil's *Humanity 2.0*, he states what he understands from artificial intelligence at the human level;

"Music and art abilities, creativity, physical motion and emotions on earth are all different and subtle areas in which people are intelligent, including understanding and reacting correctly."[120]

There are two basic approaches to an artificial intelligence developed at the human level in artificial intelligence technologies based on imitation of man; imitation of the functioning of human intelligence or imitation of human intelligent behavior.[121] The answer to the question *"When will human level artificial intelligence be reached?"* is *"A period between 2022 and 2075"*.[122] Another target that is expected to be achieved immediately after this goal is defined as *"Super Intelligence"*.[iv] What is meant by this type of intelligence is a intelligence that surpasses the human intelligence tremendously. Computer scientist and science fiction writer Vernor Vinge states that by 2030, a superior computer intelligence will emerge, this intelligence that transcends human intelligence will direct progress and the speed of such progress will be unimaginable. Vinge calls this

[116] Güven Güzeldere, "Yapay Zeka'nın Dünü, Bugünü, Yarını", Cogito Dergisi, Yapay Zeka, Yapı Kredi Yayınları, p.36.
[117] Güven Güzeldere, "Yapay Zeka'nın Dünü, Bugünü, Yarını", p.36.
[118] Güven Güzeldere, "Yapay Zeka'nın Dünü, Bugünü, Yarını", p.36.
[119] Güven Güzeldere, "Yapay Zeka'nın Dünü, Bugünü, Yarını, p.37.
[120] Ray Kurzweil, *İnsanlık 2.0*, p.211.
[121] Nils J. Nilsson, *Yapay Zekâ Geçmişi ve Geleceği*, Translated by: Mehmet Doğan, Boğaziçi Üniversitesi Yayınları, 2011, p.667.
[122] Nick Bostrom, *Superintelligence,* p.19.

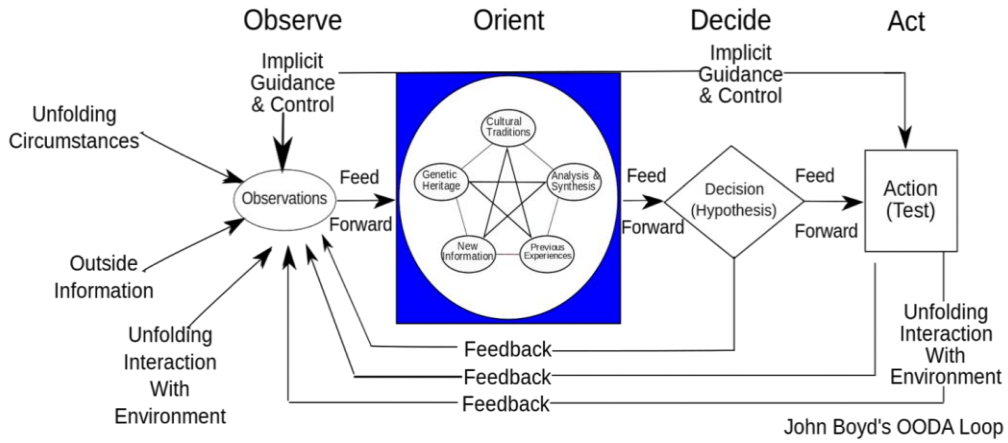
situation *"singularity"*.[123] Technological Singularity; is a futuristic period in which the speed of technological change is irreversibly transforming human life.[124]

Ray Kurzweil's example on the subject of singularity is striking; It is not possible to upload all the information you have obtained when you read Tolstoy's War and Peace. The other person who wants to understand the work has to go through the same troublesome process you have gone through. When it comes to singularity in machines, the situation will be the opposite.[125] It would be meaningful to evaluate the capacities of future war robots in this respect. The fact that the human soldier could not transfer his experience as a commander to the soldiers he directed, the fact that the commander's death and military experience, and knowledge would also disappear, but the robots that fight at the level of singularity can be reproduced in seconds in the robot fighting.

To understand the distinctions between Narrow Scoped Artificial Intelligence (today) and Human Level Artificial Intelligence and Super Intelligence (future), it is necessary to touch upon autonomy and autonomy levels in Artificial Intelligence.

The OODA Loop proposed by the US Air Force pilot and military strategist John Body, which is accepted in many professions, especially in military fields, indicates that a human takes decisions in four steps; Observe, Orient, Decide, Act.[126]

[123] Nils J. Nilsson, *Yapay Zekâ Geçmişi ve Geleceği*, p.664.
[124] Ray Kurzweil, *İnsanlık 2.0*, p.19.
[125] Ray Kurzweil, *İnsanlık 2.0,* p.211.
[126] OODA Loop: https://fasttransients.files.wordpress.com/2018/05/boyds-real-ooda-loopx.pdf



Schema-1

Although there is criticism stating that the OODA Loop oversimplifies the way people make decisions, that cycles in decision making should be synchronous, not sequential, or that the cycle does not explain the action of human decision-making, it is still widely accepted among cognitive scientists. In this theory, if the person remotely controls the machine it is defined as "in the loop", if the person watches the loop from the outside and breaks the loop using the veto power when necessary it is "on the loop", and when autonomy is activated and the human is disabled it is defined as “out of the loop”.

“Autonomy is simply; whether or not a human is in control and simultaneously how he relates to the world,” says P. W. Singer, and in his work *Wired For War*, and poses some questions to make autonomy more understandable:

“Can the robot create its own model of the world? Can it operate on its own using this model? Can it change and update the model it created automatically? Finally, can the Robot set this model aside and decide on a new route of action?”[127]

As Singer expresses in his work, what is meant when it comes to autonomy in machines is a definition of human maturity. This is not yet reached, but the distinction between the concepts of "autonomy" and "automation" allows us to

[127] Peter W. Singer, *Wired For War,* p.74, 75.

clarify, albeit partially. Automation is both a premise and an important part of autonomy. Automated systems are capable of processing without human intervention, but they are systems that cannot decide on their own.[128] In autonomy, there is a goal and an independent decision making involved to achieve that goal. A sampling for the transition levels from automatic systems to autonomy is as follows[129]:

	Level Definition
	1-The computer offers no assistance human must take all decisions and actions.
	2-The computer offer a complete set of decision/action alternatives, or
	3-Narrows selection down to a few, or
	4-Suggests one alternative
	5-Executes that suggestion if the human approves, or
	6-Allows the human a restricted time to vote before automatic execution, or
	7-Executes automatically, then necessarily informs the human, and
	8-Informs the human only if asked, or
	9-Informs the human only if it the computer, decides to,
	10-The computer decides everything, acts autonomously, ignoring the human.

Schema-2

As seen above autonomy in Artificial Intelligence is expected to happen at Level 10. In the narrow scoped artificial intelligence phase we are in, *"AI is not yet*

[128] Çağlar Ersoy, *Robotlar, Yapay Zekâ ve Hukuk*, On İki Levha Yayıncılık, 2017, p.24.
[129] Raja Parasuraman, Thomas B. Sheridan, Fellow, IEEE, and Christopher D. Wickens, *A Model for Types and Levels of Human Interaction with Automation*; https://www.researchgate.net/publication/11596569_A_model_for_types_and_levels_of_human_interaction_with_automation_IEEE_Trans_Syst_Man_Cybern_Part_A_Syst_Hum_303_286-297

able to address highly abstract ideas and evaluations, such as human motivations and ultimate goals, "[130] therefore, the concept of autonomy for machines transforms into a " pseudo autonomy" form.[131] For example, drones programmed pre-war fulfilling their missions programmed by humans is an example of the quasi autonomy situation.[132] While Bostrom states that it is not very important for artificial intelligence to be similar to the human mind he states that there is no reason for us to expect artificial intelligence to act with love, hate, pride or similar human emotional states.[133] When an artificial intelligence has reached Level 10 of autonomy, that is, the stage of complete autonomy, not a quasi autonomy, it is not possible to talk about the autonomy itself whether this autonomy will contain a state morality or not, because of the contents of the definition of autonomy. When we say "autonomy", we have to accept that we do not yet know what kind of autonomy will emerge in a nonhuman being and that our definitions and accordingly our expectations of morality are caused by a human-centered reading. Today as we are in the narrow scoped artificial intelligence stage, while we underline the fact that the human being is a living entity and a machine is a mechanism, it must be noted that the first "living robot"[134] was announced to the world in January 2020 by *The Guardian* newspaper[135], created from frog stem cells at Allen Discovery Center, Tufts University.

---

[130] Peter Lee, "Armed Drones: Automation, Autonomy, and Ethical Decision-Making," *The Political Economy of Robots*, Ryan Kiggins, Palgrave Macmillan, published by Springer Nature, 2018, p.301.
[131] Peter Lee, "Armed Drones: Automation, Autonomy, and Ethical Decision-Making," p.301.
[132] Peter Lee, "Armed Drones: Automation, Autonomy, and Ethical Decision-Making," p.301.
[133] Nick Bostrom, *Superintelligence*, p.29.
[134] Living Systems: https://www.pnas.org/content/117/4/1853
[135] The Guardian: https://www.theguardian.com/science/2020/jan/13/scientists-use-stem-cells-from-frogs-to-build-first-living-robots

1.2 A PERSPECTIVE ON ARTIFICIAL INTELLIGENCE FROM KANTIAN PHILOSOPHY

1.2.1. "Life" in Kant's Critique of The Power of Judgment, Kantian Approach to Transcendental Mechanisms and Artificial Intelligence

Kant says that the *"power of judgement"*, which acts as a middle term between understanding and reason, is the thinking as the particular being covered under the universal.[136] If the universal principle, rule, and law is given the power of judgement is decisive. But only if the particular is given and the universal needs to be found for it, then the power of judgement is pure contemplation. The power of judgement that contemplates such a principle, can only give itself a law by itself. The power takes the principle from itself and does not impose this principle on nature. Thus as an a priori particular concept the causality of nature finds its root in the reasoning power of judgement. The transcendental concept of nature's teleology is neither a concept of nature nor a concept of freedom but acts as a subjective principle of the power of judgement.[137] Nature has made its universal laws specific to our ability to know according to the teleological principle.[138]

In the *Critique of Teleological Power of Judgement* chapter of the *Critique of the Power of Judgment* Kant states, *"Now for a thing as a natural end it is requisite, first, that its parts (as far as their existence and their form are concerned) are possible only through their relation to the whole."* The thing itself is a goal, and therefore everything that defines it must be covered under a concept to be grasped as a priori.[139] In addition, parts of nature products will depend on the unity of the whole as the causes and effects of each other's forms. The objects which are judged as goal of nature have the ability to produce each other

[136] Immanuel Kant, *Critique of the Power of Judgment*, Edited by Paul Guyer, Translated by Paul Guyer and Eric Matthews, Cambridge University Press, 2000, Preface, 5:168, p.56
[137] Immanuel Kant, *Critique of the Power of Judgment*, Introduction: 5:180, p.67, 5: 184, p.71
[138] Immanuel Kant, *Critique of the Power of Judgment*, Introduction: 5:180, p.67, 5: 186, p.72.
[139] Immanuel Kant, *Critique of the Power of Judgment*, 5: 373, p.244, 245.

mutually.[140] A tree can produce another tree according to the laws of nature. What it produces is of the same species; while producing itself according to the same species (effect), by producing itself as a reason it also preserves itself as a species.[141] The same tree has the opportunity to produce itself as an individual through growth. In addition, it should be noted that by grafting the bud from one tree branch to another tree branch, it can produce plants of its own kind in a different species.[142]

Kant includes the watch example after the tree example to clarify the matter of living; *"In a watch one part is the instrument for the motion of another, but one wheel is not the efficient cause for the production of the other: one part is certainly present for the sake of the other but not because of it."*[143] A part missing in the watch cannot be produced by another part, one watch cannot produce another watch. While nature presents an organized structure, the machine is a driving force but not a creator.[144] In the organized structure, everything is both a target and a tool, and in this sense, the organized structure is a "living" structure.

Whether the machines will transcend from being a driving force to creating can be among the discussion topics of the future science and philosophy world. In the debates on the machines today, Kant is among the philosophers mentioned in both the transcendental mechanisms of consciousness and the ethics-autonomy debates. The subject of transcendental information in Kant is what we know of objects and this system that encompasses all of the information is called "transcendental thinking".[145] The object is what is shaped according to our knowledge.[146] The task of transcendental thinking is to determine the limits of reason.[147]

On the Critique of Pure Reason Kant states that all of our knowledge

[140] Immanuel Kant, *Critique of the Power of Judgment*, 5:374, p.245.
[141] Immanuel Kant, *Critique of the Power of Judgment*, 5:371, p.243.
[142] Immanuel Kant, *Critique of the Power of Judgment*, 5:371, p.243.
[143] Immanuel Kant, *Critique of the Power of Judgment*, 5:374, p.246.
[144] Immanuel Kant, *Critique of the Power of Judgment*, 5:374, p.246.
[145] Ahmet Ayhan Çitil, *Matematik ve Metafizik*, Kitap 1: Sayı ve Nesne, Alfa Basım Yayım Dağıtım, 2012, p.27.
[146] Allen W. Wood, *Kant*, Blackwell Publishing, Edited by Steven Nadler, 2005, p.27.
[147] Ahmet Ayhan Çitil, *Matematik ve Metafizik*, p.27.

begins with experience.[148] Information is of two types, a priori and a posteriori. While compulsory and universal a priori propositions do not need experiment, it is experience that determines the correctness of a posteriori propositions.[149] After his distinction in knowledge, Kant mentions two different types of judgement; analytical and synthetic judgments. While analytical judgments do not cause an expansion in knowledge, they are descriptive, and synthetic judgments broaden our knowledge.[150] If we know the meaning of the concepts of "single" and "unmarried", we do not use experimental methods to confirm the proposition that "one is unmarried," in which one concept is included in the other.[151] But in the proposition "there are two books at the table," the predicate is not included in the subject, which makes it necessary for us to resort to experience. The relationship between the object and judgement is at the heart of Kant's understanding of objects.[152] Judgement is the conception of a representation through another representation, and it is the thinking that gives unity to the multitude.[153] Among Kant's judgement types analytical a priori and synthetic is not a posteriori but a synthetic priori causes a question; *"How are synthetic judgments a priori possible?"*[154] This question is the question which the *Critique of Pure Reason* is trying to answer. Synthetic a priori propositions have the characteristics of necessity and universality in analytical propositions. Analytical propositions are suggestions that have been proved by conceptual analysis or logical inferences.[155] But synthetic propositions are also propositions that broaden our knowledge. Kant mediates[156] a third thing to clarify this, namely how synthetic a priori knowledge is formed, this is the "intuition" derived from time and space, the pure forms of our senses. Time and space exist as pure forms in our minds as a priori. For Kant, the space[v] of objects is reason, and they are grasped within the judgment in reason.[157] A concept is the comprehension of an object in a judgment. Judgment is what gives space to the

[148] Immanuel Kant, *Critique of Pure Reason*, Introduction, B1, p.136.
[149] Immanuel Kant, *Critique of Pure Reason*, B2, p.136.
[150] Immanuel Kant, *Critique of Pure Reason*, B11, p.141.
[151] Paul Guyer, *Kant*, Edited by Brian Leiter, Published by Routledge, 2006, p.47.
[152] Ahmet Ayhan Çitil, Matematik ve Metafizik, p.31.
[153] Immanuel Kant, *Critique of Pure Reason*, B93, p.205, 206.
[154] Immanuel Kant, *Critique of Pure Reason*, B19, p.146.
[155] Ahmet Çevik, *Matematik Felsefesi ve Matematiksel Mantık*, Nesin Yayıncılık, 2019, p.51.
[156] Immanuel Kant, *Critique of Pure Reason*, B194, p.281.
[157] Ahmet Ahmet Çevik, *Matematik Felsefesi ve Matematiksel Mantık*, p.51, 52.

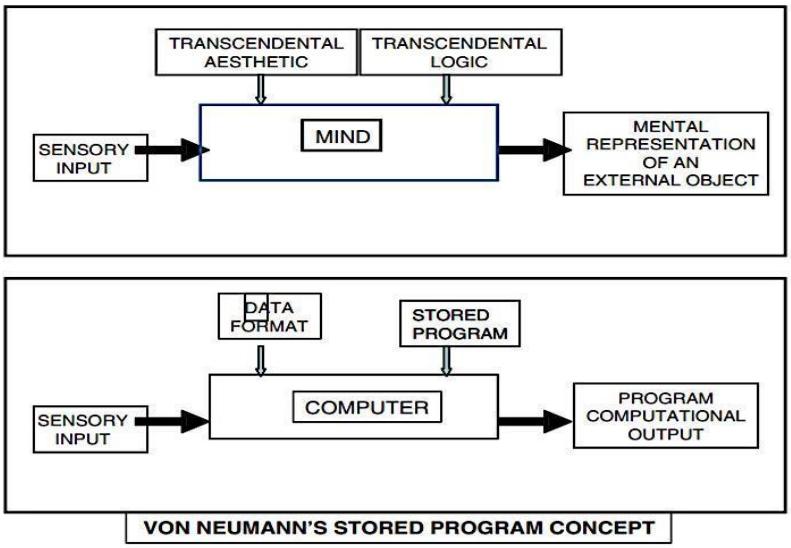
concept.[158] Object and concept, therefore, information is only what happens in judgement. At this stage, Kant goes to the distinctions of "Transcendental Aesthetics" and "Transcendental Logic."[159] Transcendental Aesthetics (TA) focuses on time and space as pure intuition and a priori intuition forms. Kant argues that in TA, all representations of certain objects must be given to us in time and space, so that space and time are all forms of intuition, and that space and time itself must be represented as certain objects.[160] In Transcendental Logic (TL), where he thinks about pure thinking, Kant's dual distinction is; transcendental analytical and transcendental dialectic. Kant, who deals with comprehension and its principles in transcendental analytics, deals with the principles of reason and reason in transcendental dialectics.[161]

According to Kant, through the representations created with the ability to take gives the material that makes the "sensing" intuition possible.[162] In the ability to feel that provides the material in the manifest (thing in itself), the form of the manifest must be distinguished from its material; while the material of feeling is given to us a posteriori, we already have the form as a priori.[163] The material, which is lined up and arranged in time and space forms, turns into an experience in us, through the realization of the mind that has the power to produce representation from itself.[164] The combination of the ability to feel and comprehension emerges as a concept and the manifestation is represented as an understood object.[165] The dual distinction in representation is an important detail; representation; representation of sensible things and it is understood that there is nothing sensory in it, pure representation.[166] It functions in maintaining and reproducing the representation of the object through the time form of the Imagination.[167] Imagination is the ability to represent a visual counterpart that does not exist.[168] As a result of the conceptual activity of the understanding and unity with judgement, these activities are

[158] Ahmet Ayhan Çitil, *Matematik ve Metafizik*, p.33.
[159] Immanuel Kant, *Critique of Pure Reason*, B36, p.156.
[160] Paul Guyer, Kant, p.55.
[161] Bülent Gözkan, *Kant'ın Şemsiyesi*, Yapı Kredi Yayınları, 2018, p.41.
[162] Immanuel Kant, *Critique of Pure Reason*, B33/43, p.155.
[163] Immanuel Kant, *Critique of Pure Reason*, B33/43, p.155.
[164] Immanuel Kant, Critique of Pure Reason, B75, p.193.
[165] Immanuel Kant, *Critique of Pure Reason*, B103, p.210.
[166] Bülent Gözkan, *Kant'ın Şemsiyesi,* p.44.
[167] Bülent Gözkan, *Kant'ın Şemsiyesi*, p.54.
[168] Immanuel Kant, *Critique of Pure Reason*, B151, p.256.

provided with the object and its knowledge. Judgement is one and the same ability as thinking, which covers the whole thinking ability.[169] Apprehension is the actual judgment that ensures that a representation or imagination can be grasped through another imagination.[170] What gives all these activities unity is the "transcendental apperception" that can mean a consciousness becoming consciousness of itself. It is possible through the transcendental apperception that the perceiving activity of both the ability to feel and the imagination unites itself.[171] In terms of the subject we are examining, we must state that the imagination, which is a basic skill in Kant, is important not only in the establishment of the phenomenal world but also in the establishment of the "I".[172]

These general explanations of Kant regarding the knowing experience of the human reason and the limits of reason are considered as a method for some cognitive scientists to artificially construct the phenomenon of consciousness.



173

Shcema: 3

[169] Ahmet Ayhan Çitil, *Matematik ve Metafizik*, p.30.
[170] Ahmet Ayhan Çitil, *Matematik ve Metafizik*, p.31.
[171] Ahmet Ayhan Çitil, *Matematik ve Metafizik*, p.35.
[172] Bülent Gözkan, *Kant'ın Şemsiyesi,* p.61.
[173] Frederick Betz, *Managing Science*, Illustration: Kant's Critique of Pure Reason, Springer Media, 2011, p.151

Frederick Betz draws attention to two features that must be present on a computer before data can be entered; the data format and stored program required to configure the data. In this definition, what corresponds to Kant's transcendental aesthetics is specified as data format, while transcendental logic is expressed as a stored program. Just as Kant stated in the first criticism, the same situation is valid for data format and stored format in computers, just as the mediate representiton function in order for a thing to turn into information as an object in us.[174] Accepting Kant's transcendental mechanisms as the brain's mechanisms, Betz states that this scheme is functional in the development of computers.[175]

Richard Evans, in his article "A Kantian Cognitive Architecture" presented to IACAP (International Association for Computing and Philosophy) in 2016, states *" why should a book written in the eighteenth century have anything to teach us now? Kant still has something to teach us. His insights have not yet been fully absorbed into cognitive science or AI,"* and continues that Kant has great contributions to make to machine learning.[176] Evans states that the modern debates between deep learning and logic-based approaches, which are among the learning methods in artificial intelligence, are similar to the discussion between experimenters and rationalists in philosophy.[177] Stating that Kant's philosophy helped make great progress in thought by bringing these two opinions together, Evans also states that the Kant framework can do the same job in artificial intelligence, and subjects Kant's evaluation of the sensory data to the conditions required to be understood by any agent to computer evaluation.[178] Kant's unification of diversity in synthesis is a kind of a characterizing activity of imagination and what Kant means is not a linguistic symbolization, but rather something that computer scientists call gen-sym (a generated symbol: an atomic identifier) an uninterpreted symbol. What the actor does by associating these uninterpreted symbols with other things is a kind of a comprehension.[179] For example, if one sensor (sensing device) pattern represents the nose and the other

[174] Frederick Betz, *Managing Science*, p.150.
[175] Frederick Betz, *Managing Science*, p.151.
[176] Richard Evans, "A Kantian Cognitive Architecture", On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence, Themes from IACAP 2016, Springer Nature, 2019, p.233.
[177] Richard Evans, "A Kantian Cognitive Architecture", p.260.
[178] Richard Evans, "A Kantian Cognitive Architecture", p.234, 235.
[179] Richard Evans, "A Kantian Cognitive Architecture", p.236.

pattern represents the eye, it will turn out that the total model of the sensors represents a face.[180] Emphasizing the succession of the time issue in Kant, Evans creates some computer models by reinterpreting the basic concepts of Kant, such as imagination, as well as time.[181] According to Evans, this modeling created by Kant's understanding of cognition is strong enough to lead to achieving a human level performance in artificial intelligence.[182]

The evaluation of Kantian approaches in their studies on cognitive activities in Artificial Intelligence deserves to be a thesis topic in itself. However, we can only make a brief evaluation in order to evaluate the main subject of our study, the evaluation of "autonomy" from the Kantian perspective and not to deviate from the main subject since autonomy is related to morality. Although Kant accepts that transcendental mechanisms can create a strong example scheme in artificial intelligence studies, the conceptual mediation in the Kant interpretation based on formal systems based on Evans seems to be eliminated or misunderstood. It is not explained in terms of formal strings that the representations can be kept together with another representation, visualization activity (the activity of imagination). When this situation cannot be explained, a pure synthesis from Kant's perspective cannot be made meaningful either. In order to speak of a concept, the object must be put forward in the judgement activity and the activity of thinking must be a transcendental activity.[183] We should point out that all these situations are not yet meaningful for the *"artificial narrow intelligence"* technologies stage that we are in, but we can also point out that up to a certain extent, Kant's *Critique of Pure Reason* can contribute to the development of artificial intelligence studies.

1.2.2 The Question of Applying Kant's Morality to Artificial Intelligent Structures

[180] Richard Evans, "A Kantian Cognitive Architecture", p.257.
[181] Richard Evans, "A Kantian Cognitive Architecture", p.236
[182] Richard Evans, "A Kantian Cognitive Architecture", p.234.
[183] Bülent Gözkan, *Kant'ın Şemsiyesi,* p.56.

In the *Groundwork for the Metaphysics of Morals*, the main problem of Kant's moral philosophy, which is directed from the natural consciousness to the categories of moral philosophy, is undoubtedly the problem of freedom. The *"categorical imperative"*[184] that covers all intelligent beings is not a *"particular adaptation of human nature"* or the random conditions in which it lies. It is not only the people with reason, but all the finite beings with reason and will, and even the Supreme Being as the infinite being.[185] It is remarkable that Kant has universalized the definition of mind and reason in a way that includes people and other beings. Reason is mind if it is universal, and this universal ability includes beings other than human. Naturally, if this coverage can be achieved, we can easily say that artificial intelligence can be included that we have done serious studies in this direction, namely, creating an artificial intelligence, reason, consciousness, and cognition, and that the definition of formal reason in the Kantian philosophy has been taken as a model. The fact that Kant's moral law formed by the concepts of reason, necessity, autonomy and freedom is just conceptual and has problems in practical application may bring along the questioning whether this law is applicable in any other type of reason other than human. The question we ask is this; Can the problems experienced in Kant's moral law be overcome in Artificial Intelligence? Can Kant's moral understanding be applied to machines?

$$\text{IF } V_{\text{guilt}} > \text{Max}_{\text{guilt}} \text{ ThEN } \boldsymbol{P}_{\boldsymbol{l\text{-}ethical}} = \boldsymbol{\emptyset}$$ [186]

Ethics is no longer just referred to as a human-centered ethics, with the introduction of Machine Ethics as a new discipline that researches the ethics of self-driving cars, robots and drones, autonomous and semi-autonomous systems.[187] The "three laws of robotics" of the sci-fi writer Isaac Asimov, which was put on the

[184] Immanuel Kant, *Groundwork for the Metaphysics of Morals,* p.171,172.
[185] Immanuel Kant, *Crituqe of Practical Reason*, p. 32.
[186] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots*, CRP Press, 2009, p.142.
[187] Oliver Bendel, Kevin Schwegler, Bradley Richards, "Towars Kant Machines", The AAAI 2017 Spring Symposium on Artificial Intelligence for the Social Good Technical Report SS-17-0, School of Business FHNW, Bahnhofstrasse 6, CH-5210 Windisch.

agenda by the European Parliament in 2016, in a report[188] which discusses the evaluation of robot laws in terms of European civil rights state; First Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm. Second Law: A robot must obey the orders given it by human beings except where such orders would conflict with the First Law. Third Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.[189]

Asimov, by adding a "zeroth law" to the laws of robotics later (A robot may not harm humanity, or, by inaction, allow humanity to come to harm.) changed the first law to; *"A robot may not injure a human being or, through inaction, allow a human being to come to harm unless it contradicts the zeroth law."* In the report published by the European Parliament, it states that these laws cannot be enforced by robots yet, but that people who design robots can take these laws as basis. Kant is also mentioned among the opinions applicable to machine ethics. In his article "Prospects for a Kantian Machine", Thomas M. Powers says that; Kant's ethical approach can be applied to machine ethics as it offers a computational structure for the judiciary,[190] continues that, As Aristotle says people suffer from weakness of willpower, but machines don't have this problem." Machines are stable; on the contrary, when people encounter the same problem twice, they make the same decision.[191] Emphasizing that Kant's moral law does not contain individuality, Powers points out that machines do not have the autonomy to apply a universal law to themselves, but categorical imperative can be applied to machines through people. Analyzing whether categorical imperative can be applied to machines in terms of absolute consistency, practical reasoning and compliance, Power states that there are some difficulties, but studies on the logic of machine

[188] European Parliamet, European Civil Law Rules in Robotics: https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf

[189] Isaac Asimov, *Ben, Robot*, İthaki Yayınları, Translated by: Ekin Odabaş, 2016.

[190] Thomas M. Powers, "Prospects for a Kantian Machines", Intelligent Systems, IEEE 21(4):46- 51 · August 2006, https://www.researchgate.net/publication/3454367_Prospects_for_a_Kantian_Machine

[191] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda": https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3180741

ethics can make progress on this issue and further clarify the moral principles of man.[192]

Goodbot (a chat bot[vi]) is a software whose prototype was developed in 2013, coded not to lie unless its lying does not hurt the user. Liebot, produced in 2016, is a lie software. Liebots are referred to as "Münchausen Machines"[vii] while Goodbots are referred to as "Kantian Machines".[193] In order to understand the risk of Münchausen Machines (lying machines), we need to consider how natural language systems are penetrating into our daily lives. Online media, websites, automatic agreements and robo-journalism created through natural language systems are becoming more and more common. So how can we trust the information we have gained through these systems? As a solution, the production and spread of Kantian Machines in use are among the discussions.[194]

It should be noted that what determines Kant's understanding of morality is "duty" and that it is given as a "mission" to not lie to a machine, and Kantian ethics will not be implemented in this way since autonomy and freedom cannot be mentioned in such a structure. Ozlem Ulgen, in her article titled "Kantian Ethics in the Age of Artificial Intelligence and Robotics", states that artificial intelligence and robots do not have human thinking, reasoning and free will, so they naturally do not understand that the law is desirable, feasible, valuable and universal.[195] However, categorical imperative has the opportunity to be implemented programmatically, taking into account that people design, manufacture and develop artificial intelligence technologies. Ulgen states; *" For these rules to be capable of universalisation they must be 'public and shareable."*[196] But this depends on what the technologies are intended for. Will they completely replace human functions and rational thought, or will they serve a purpose that supports human

[192] Thomas M. Powers, "Prospects for a Kantian Machines", Intelligent Systems, IEEE 21(4):46-51.
[193] Oliver Bendel, Kevin Schwegler, Bradley Richards, "Towards Kant Machines", The AAAI 2017 Spring Symposium on Artificial Intelligence for the Social Good Technical Report SS-17-0, School of Business FHNW, Bahnhofstrasse 6, CH-5210 Windisch.
[194] Oliver Bendel, Kevin Schwegler, Bradley Richards, "Towards Kant Machines".
[195] Ozlem Ulgen: "Kantian Ethics in the Age of Artificial Intelligence and Robotics" https://www.researchgate.net/publication/321883098_Kantian_Ethics_in_the_Age_of_Artificial_Intelligence_and_Robotics
[196] Ozlem Ulgen: "Kantian Ethics in the Age of Artificial Intelligence and Robotics".

characteristics?[197] Evaluating the risks of full autonomous weapon systems, Ulgen drawing attention to the ICRC (International Committee The Red Cross) 2014 and 2016 reports[198], states that many scenarios in which full autonomy weapon systems will have capacities such as independent thinking, monitoring, choosing and attacking human targets in war have been evaluated. In regard to the differences between human will and machine will, Ulgen states that based on narrow scoped artificial intelligence, the machines are not in an imbalance on which of the options are moral, and when it comes to voluntary autonomy, she says that it is essential that the person has completed their internal and external development and has reached a moral stance.[199] Therefore, the will, which is autonomous when people are essential, becomes a heteronome when it comes to today's machines.

In the paper written in 2018 for the Harvard Kennedy School, Carr Center for Human Rights Policy titled "Human Rights and Artificial Intelligence: An Urgently Needed Agenda"[200], Mathias Risse puts the discussion on the agenda of whether the rationality between David Hume and Immanuel Kant determines our values. Recalling that Hume did nothing to determine the values of reason, he states, *"If Kant is right, a superintelligence might be a true role-model for ethical behavior."*[201] According to Risse, who focuses on future scenarios rather than today's semi-autonomous machines, artificial intelligence technologies will bring some difficulties in terms of human rights. Considering the immunity and superiority of human life, *"These basic assumptions are questioned through the anticipated arrival of entities that are not alive in familiar ways but nonetheless are sentient and intellectually and perhaps eventually morally superior to humans."*[202] The following determination of Risse is remarkable, saying that machines can show emotions such as love, loyalty and may even attempt to solve the problem by worrying about being shut down; *" If the mind just is a complex*

[197] Ozlem Ulgen: "Kantian Ethics in the Age of Artificial Intelligence and Robotics".
[198] ICRC: 'Report of the ICRC Expert Meeting, "Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects".
[199] Ozlem Ulgen: "Kantian Ethics in the Age of Artificial Intelligence and Robotics".
[200] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda", Risse, Mathias, Human Rights and Artificial Intelligence: An Urgently Needed Agenda (May 18, 2018). HKS Working Paper No. RWP18-015. Available at SSRN: https://ssrn.com/abstract=3180741 or http://dx.doi.org/10.2139/ssrn.3180741
[201] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda".
[202] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda".

algorithm, then we may eventually have little choice but to grant the same moral status to certain machines that humans have."[203] Risse makes future-based determinations based on artificial intelligence at the human level and superintelligence that exceeds the human level, which is based on full autonomy. We should state that there is no duty morality in the Kantian sense because there is no autonomy when we base on narrow scoped artificial intelligence. However, it would still be appropriate to consider to what extent Kant's morality can be applied to machines.

Kant states that moral law covers all rational finite beings. By excluding the role of the consciousness of being a finite being in human activity, even though the machines will remain on the earth for a much longer time than human life, they are still not infinite by being connected to the material world, and in the first sense, they can be included in the moral law of Kant. Taking into account that the practical laws in themselves conflict with the maxim when stimulated by the passions of the wise man, the fact that semi-autonomous machines do not experience this conflict seems to be an advantage. The objective desire of the machines (by deactivating their full autonomous states) is a workable situation not as moral duty through humans but in terms of mission. Although determining which concrete action is suitable for the maxim seems technically quite problematic in terms of machinery as well as for humans, this problem seems to be solvable in a world where every object living in the technologies of the future finds itself in an algorithm network. Kant, in the *Groundwork for the Metaphysics of Morals* states; skillfulness in the selection of tools for a person to put himself in the best state in a narrow sense can be called intelligence.*"*[204] Certainly, the imperative of this intelligence is a conditional imperative. The action was transformed into a tool for another purpose. But the imperative that directly commands the action is related to the mode of action and the principles that will reveal it. The imperative commands us morality.[205] When we say that the moral law exists in every human being as an a priori, when we state that the so-called being already contains autonomous law, if we say that reason is a person-specific ability in the human body and in the

[203] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda".
[204] Kant, *Groundwork for the Metaphysics of Morals,* p.104.
[205] Kant, *Groundwork for the Metaphysics of Morals,* p.106.

historical evolution of the body-Kant does not say this, we have excluded intelligent machines from this definition of mind. But if we state that the ability to reason is a formal and universal structure, we have to accept that this type of reason is possible to be artificially produced in the future. This artificially produced reason will contain an exact copy of the intelligent human reason referred to by Kant as "mandatory" in moral law. If this copy does not contain a moral law, a shadow will fall on Kant's universality of reason. And if this artificial mind has responsible freedom a result of its actions, the thesis of a priori freedom will also be compromised.

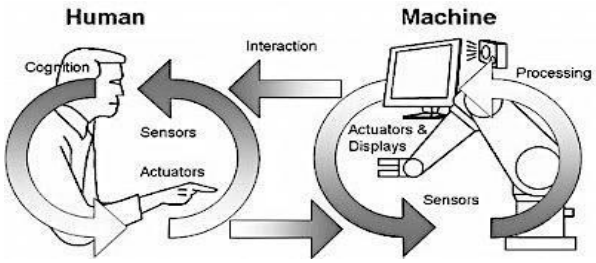
While it is possible to remove the *"categorical imperative"* from the Kant context, and ascribe it through the human as a mission to semi-autonomous machines and to use the term "Kantian Machine" in this sense, to say that Kant's moral law will be appropriate in full autonomous machines contradicts the concept of "autonomy". Let's remember Singer's definition of autonomy in machines;

"Can the robot create its own model of the world? Can it operate on its own using this model? Can it change and update the model it created automatically? Finally, can the Robot cast aside this model and find a new approach?"[206]

To say whether an autonomous machine can have a moral law will not take us beyond doing a "human-centered reading". Max Tegmark in *Life 3.0* states that, we will not be able to set goals or ethical rules for artificial intelligence in the super intelligence phase.[207] Another remarkable study on machine ethics is the *"artificial conscience"* study carried out by Ronald C. Arkin. Arkin, who stated that in his robot ethics studies Kant's categorical imperative cannot be applied to war robots, also states that an imperative is not needed as well.[208] Arkin, in a proposal to the United States Army Research Office, states that "artificial conscience" projects were financed in 2006, that these robots were called Humane-oids, and that this

[206] Peter W. Singer, *Wired For War,* p.74, 75.
[207] Max Tegmark, *Life 3.0*, Borzoi Book Published by Alfred A. Knopf, 2017, p.269.
[208] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots*, CRP Press, 2009, p.104.

type of robots could potentially be more ethical than humans on the battlefield.[209] If we take into account that we haven't yet reached a human level in artificial intelligence, we can say that artificial conscience robots that Arkin is talking about are semi-autonomous. We must state that a fully autonomous artificial intelligence at the human level and that exceeds the human level (superintelligence) is a future-oriented work, and that we cannot say whether an artificial intelligence at this level will have a human-form morality, but also that we cannot make a moral proposal for the "new human" of the future.



210

There are different opinions on whether a super-intelligence exceeding human level will be realized simultaneously with a human level artificial intelligence or after human level artificial intelligence.[211]Among methods of reaching super intelligence, opinions on turning people into a super intelligence are on the agenda as well. By using embryos and gamete cells in genetic selection and intervening in these cells, the level of super intelligence can be achieved through *"designed humans"*.[212] Nick Bostrom states that such a preference cannot be economical even if it is feasible.[213] *"The impact of this technology will be dampened and delayed by several factors."* it will be necessary to take into account that at least twenty years must pass before the elected children can be brought into the workforce.[214] It is also possible to strengthen the existing biological structure by means of cybernetic organisms (machine-human mixture) rather than

[209] Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots*, CRP Press, 2009, Preface, xvi.
[210] Nick Bostrom, *Superintelligence,* p.21.
[211] Nick Bostrom, *Superintelligence,* p.21.
[212] Nick Bostrom, *Superintelligence,* p.37.
[213] Nick Bostrom, *Superintelligence,* p.39.
[214] Nick Bostrom, *Superintelligence,* p.39.

transforming the organic structure of a man into super intelligence, thereby equalizing humans against a possible machine super intelligence and eliminating the concerns about the position of the human species against the machine. According to Max Tegmark, as much as cyborgization seems like a product of science fiction, it is possible because there is no condition that violates the known laws of physics.[215]

Ray Kurzweil states that in the 2030s, biological intelligence and non-biological intelligence will be combined with the implementation of nanobot applications, and this merging will expand the human mind.[216] Nanorobotic devices similar to biological cells are nanometer scale programmable structures.[217] Nanobots consist of structures that are inserted through blood rather than surgery, such as neuron implants placed by surgical procedures, and can be easily removed from the body by directing the nanobots if necessary.[218] These structures greatly improve pattern recognition, memory and thinking capacity in general, as well as share thoughts, emotions, reactions, and pleasures with other rational beings through neurological connections.[219] Humans being turned into cyborgs also brings the risk of hacking of the human mind, as well as the opportunity to technically eliminate the state of *"lying"*, which is the most important feature that distinguishes today's human from other living things.

In the *Groundwork for the Metaphysics of Morals*, Kant states:

"Everyone must admit that a law, if it is to be valid morally, i.e., as the ground of an obligation, has to carry absolute necessity with it; that the command 'You ought not to lie' is valid not merely for human beings, as though other rational beings did not have to heed it."[220]

[215] Max Tegmark, *Life 3.0*, p.155.
[216] Ray Kurzweil, *İnsanlık 2.0,* p.473.
[217] Deepak N. Kapoor, Arvind Singh, Sushila Singh, "Nanorobotics": https://www.researchgate.net/publication/227245023_Nanorobotics
[218] Ray Kurzweil, *İnsanlık 2.0,* p.474.
[219] Ray Kurzweil, *İnsanlık 2.0,* p.472.
[220] Kant, *Groundwork for the Metaphysics of Morals*, Preface, p.38.

In a sense, the natural structure of man stands as an obstacle to the implementation of the law. Elimination of human-originated situations such as lying on the way to Kant's "Moral world wisdom"[221] may seem to enable this wisdom to be realized. Kant's "Realm of Ends" is the *"systematic union of various rational beings established through common laws."* It is quite possible to say that this systematic union can be achieved through a network that will be provided neurologically between humans and between human and machine through nanobots. Mathias Risse draws attention to such a possibility regarding super artificial intelligence and cyborgization, by saying, *"Perhaps then we have some chance at getting protection, or even some level of emancipation in a mixed society composed of humans and machines, given that the abilities of the human brain are truly astounding and generate capacities in human beings that arguably should be worthy of respect."*[222]

[221] Kant, *Groundwork for the Metaphysics of Morals*, Preface, p. 38.
[222] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda".

1.3- HUMAN RIGHTS, DIGNTY, AND AUTONOMY

The discussion of exactly what human nature is shown as the normative source of human rights goes back to very old times. Although it is not possible to say that there is a clear consensus in the discussions, it would be appropriate to state that today's technology deepens these discussions. "Who is called a human being?" New questions have been added to this question now; Where does the human being begin and where does it end? *"Are fetuses, designed babies, clones, and those who are in constant vegetative life fully human?"*[223] Responses/possible responses to these modern questions bring/will bring new definitions to the human, and cause new discussions on the issue of rights and dignity. Stating that the human rights literature is facing new issues every day and therefore there is a "human rights inflation", Mustafa Erdoğan underlines that the concept of "right" should be defined correctly.[224] When a legal right is mentioned, this concept brings with it a person and the autonomy of that person as a moral subject.[225] Costas Douzinas, who sees human rights as a special type of rights as something that combines law and morality as an institution, defines human rights as follows; *"Moral rights or demands that are given to individuals or not recognized in a certain legal system ... The human rights institution combines law and morality, description and command."*[226]

Natural law is cited as a source for human rights thought, including some modern definitions. Human rights have historically been derived from natural law, and natural law claims that law is part of morality.[227] In natural law, the right and wrong are understood through universal principles, while the natural law principles that act as superior principles initially find their source in God, it is stated that "the natural laws set by God should direct the laws of human

[223] Costas Douzinas, *İnsan Hakları ve İmparatorluk*, Translated by: Kasım Akbaş, Rabia Sağlam, Bilgi Üniversitesi Yayınları, 2017, p.55.
[224] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku*, Hukuk Yayınları, 2018, p.7.
[225] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku,* p.8.
[226] Contas Douzinas, *İnsan Hakları ve İmparatorluk*, p.10.
[227] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku,* p.38.

administrators"[228]. From the references to God and the scriptures, the interpretations that translate from state constitutions and conceptions referring to international law include a long process of change, ranging from natural rights to human rights.[229] For some, while human rights are an appropriate and morally necessary work in the face of injustice, human rights consist of unqualified works that should be approached with skepticism. Again, according to some, human rights are a concept referring to the human's inherent dignity and value, while for others, this concept is only the result of some security requirements that serve as a shield.[230]

In human rights discussions, it is seen that the historical development of human rights is considered in three main periods.[231] The distinction, called First, Second, Third Generation opinions, constitutes the agenda of fundamental rights, politics, social and economic rights in the first and second generations, while there is an expansion in third generation rights. Civil and political rights such as cultural rights, group rights, the right to seek a country of asylum, asylum seeker right, and the right to work are among the most discussed issues in the views of Third Generation rights.[232] One of the main concepts of Third Generation rights, "Right to solidarity" refers to a particular group by referring to a social group rather than individual rights, taking on subjects such as cultural authenticity, and covers all humanity with environmental issues and other similar issues.[233] The right to solidarity also invites all humanity to actions worldwide and puts the responsibility of the problems experienced on all of humanity. Other than the right to solidarity, the Third Generation opinions, which draw attention with the discussions about the "Right of peoples to exist", show a kind of minimal approach with the rights of a particular group by putting genocide, national, ethnic, racial or religious groups on the agenda.[234]

[228] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku,* p.39, 40.
[229] Costas Douzinas, *İnsan Hakları ve İmparatorluk*, 11.
[230] Andrew Clapham, *Human Rights,* A Very Short Introduction, Oxford University Press, 2007, p.1,2.
[231] Harun Tepe, *İnsan Hakları Felsefesi*, Bilge Su Yayınları, 2018, p.76.
[232] Harun Tepe, *İnsan Hakları Felsefesi*, p.76.
[233] Harun Tepe, *İnsan Hakları Felsefesi*, p.81.
[234] Harun Tepe, *İnsan Hakları Felsefesi*, p.81, 82.

In the human rights issue, which follows a historical course from natural law to minimal approaches of positive law, Kant stands at an important point as the person who shows that natural law should turn into positive law.[235] While the idea of human rights in the pre-Kant period was understood as a kind of synthesis of Stoicism and Christianity, with the contribution of Kant, human beings are separated from other creatures by reason with the framework established by the minds, and these rights become meaningful with "dignity" that gives the basis for human rights.[236] Human rights refer largely to freedom rights and are derived from freedom.[237] In this sense, freedom of choice lies at the basis of the rights. Man has the opportunity to realize his own nature through freedom. The fact that one has a right has an explanation value for his freedom.[238] In this respect, freedom can also be shown as the source of human rights.[239] According to the rational view of freedom, which draws attention to the relationship between freedom and autonomy, there are not only external conditions but also internal conditions and internal obstacles that prevent freedom.[240] In the case of non-rational internal conditions such as unconsciousness and false consciousness, a person's freedom cannot be mentioned.[241] Freedom in Kant is dependent on the fact that one can think. Man belongs to both the world of thought and the world of senses, but when it comes to freedom, what makes people moral beings is that they can think. What makes man an ethical subject in Kant is not only that he is acting, but rather what he wants, and more importantly, is the principle behind his will.[242]

Yavuz Kılıç states in his article titled "The Possibility of Valuable Action in Kant's Ethical Opinion", "*It is the demand that the definitive command of good, pure practical mind, which is an attribute of demand, is determined."*[243] If good is

[235] Celal Yeşilçayır, *İnsan Hakları Felsefesi,* Çizgi Kitabevi Yayınları 2019, p.153.
[236] Celal Yeşilçayır, *İnsan Hakları Felsefesi,* p.135.
[237] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku,* p.33.
[238] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku,* p.33.
[239] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku,* p.34.
[240] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku,* p.35.
[241] Mustafa Erdoğan, *İnsan Hakları Teorisi ve Hukuku,* p.35.
[242] Yavuz Kılıç, "Kant'ın Etik Görüşünde Değerli Eylemin Olanağı", Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Sayı 22, 2015, Sayfa 93-100, p.94.
[243] Yavuz Kılıç, "Kant'ın Etik Görüşünde Değerli Eylemin Olanağı", p.94.

not based on practical law, it turns into a form of pleasure and is left to the initiative of the experiential field.[244]

Good, but it can gain meaning in the concept of "duty" as desired. Kant defines duty as follows; "*Duty is the necessity of an action from respect for the law*."[245] The law deserves to be commanded as the subject of respect.[246] In Kant, respect is important to understand the human's worthiness and in this sense dignity. A number is a feeling that is experienced as a concept of the mind and which is separated from all other senses.[247] As a consciousness that the will is determined by a law, respect is a valence that is experienced freely, autonomously, without our own determination.[248]

Kant states; *"Act so that you use humanity, as much in your own person as in the person of every other, always at the same time as end anever merely as means."*[249] The worthiness of the wise man manifests itself in this maxim. The being that leaves the love of I behind is an honorable being for Kant.[250] When it comes to Kant, what makes people dignified and valuable is their willingness to comply with their own laws, that is, being autonomous.[251] In Kant's "Realm of Ends" everything has either a price or that thing is valuable in itself. Value is something being itself a purpose.[252]

With Kant, the concept of dignity, which has been widely brought up in the grounding of human rights, is found significant in relation to human worthiness. As the carrier of values, the human is defined as a presence of dignity.[253] Just as there are those who associate the existence of human dignity with the issue of values, there are also those that relate to the human position of nature, like George Kateb. Kateb, in his *Human Dignity*, states that what distinguishes human from other beings makes it special and honorable is *"keep the record of nature,*

[244] Yavuz Kılıç, "Kant'ın Etik Görüşünde Değerli Eylemin Olanağı", p.95.
[245] Kant, *Groundwork for the Metaphysics of Morals,* p.65.
[246] Kant, *Groundwork for the Metaphysics of Morals,* p.66.
[247] Kant, *Groundwork for the Metaphysics of Morals,* p.68 (footnote)
[248] Kant, *Groundwork for the Metaphysics of Morals,* p.68 (footnote)
[249] Kant, *Groundwork for the Metaphysics of Morals,* p.139.
[250] Yavuz Kılıç, "Kant'ın Etik Görüşünde Değerli Eylemin Olanağı", 98.
[251] Yavuz Kılıç, "Kant'ın Etik Görüşünde Değerli Eylemin Olanağı", p.98.
[252] Kant, *Groundwork for the Metaphysics of Morals,* p.154.
[253] Celal Yeşilçayır, *İnsan Hakları Felsefesi,* p.131.

understand nature, and appreciate it."[254] In the view of Jurgen Habermas, human dignity depends on a symmetry of relations.[255] In Habermas, which interprets Kant's categorical imperative as "demanding to leave the first person perspective for the sake of a perspective", the honor is "human dignity".[256] Human dignity; "*It points to mutual acceptance in interpersonal relations and "immunity", which has a meaning in the equality of people's relations with each other."*[257]

The relationship of honor with autonomy and rights, which comes to the fore in the grounding of human rights, is not just a subject of discussion that focuses on people. Especially in disciplines such as machine ethics and machine law, it brings together new debates such as machine rights and machine dignity in the definition of human-machine relationship and autonomous machines. The law, which cannot proceed simultaneously with the speed of technological changes, is facing great difficulties in making restrictive definitions today.[258] For example, the concept of autonomy in machines itself makes it difficult to define the concept legally, as it points to a different level at each stage in decision cycles.[259] Legally, autonomy means:

-To have a legal status (to have rights and responsibilities).

-Having legal capacity (making legally valid decisions or taking action).

-Being legally responsible for decisions or actions.[260]

Çağlar Ersoy, who is currently evaluating the current work on whether statuses such as "electronic person", "artificial person" can be given to autonomous machines with artificial intelligence, emphasizes on the subject of "robot dignity" that until the concept can be established by the robots themselves we can make a definition of robot dignity through human dignity,[261] and *"As part of a study by the*

[254] George Kateb, *Human Dignity*, The Belknap Press of Harvard University Press, 2011, p.114.
[255] Jurben Habermas, *İnsan Doğasının Geleceği*, Translated by: Kaan Ökten, Everest Yayınları, 2003, p.55.
[256] Jurben Habermas, *İnsan Doğasının Geleceği*, p.91.
[257] Jurben Habermas, *İnsan Doğasının Geleceği*, p.55.
[258] Çağlar Ersoy, *Robotlar, Yapay Zeka ve Hukuk*, On İki Levha Yayınları, 2017, p.33.
[259] Çağlar Ersoy, *Robotlar, Yapay Zeka ve Hukuk*, p.35.
[260] Çağlar Ersoy, *Robotlar, Yapay Zeka ve Hukuk*, p.36.
[261] Çağlar Ersoy, *Robotlar, Yapay Zeka ve Hukuk*, p.142.

UK government in 2006, it is predicted that robots may claim rights and some rights will be granted to them in the next 20 to 50 years,"[262] he states.[263] Report mentioned by Ersoy gives us some clues about the near future of humanity. It is understood from the report of the British government that technological singularity is supported, while the society will have responsibilities towards the "digital citizens" and the machines will have full social rights.[264]

Human rights, created on the basis of human hierarchical superiority against other life forms, will most likely experience some difficulties in the face of artificial intelligence technologies.[265] Risse states that we need to worry about artificial intelligence as *"AI will drive a widening technological wedge into societies that leaves millions excluded, renders them redundant as market participants and thus might well undermine the point of their membership in political community.."*[266] When wealth was determined by land ownership, the rich needed the rest to to pay the rent. When wealth was determined by factory ownership, factory owners needed the rest to run the machines and buy goods. But those who are on the losing side of the technological gap will no longer be needed.[267] What will people be useful for in this possible future? Bostrom roughly responds to this question as follows: What kind of price unit human-made goods have today (relatively more expensive than machine-made goods) may become a trend in the future.[268] Of course, the trend is likely to partially exceed today's conceptions; man-made goods can also be included as a natural ability; such as human athletes, human artists, human lovers, human soldiers. It is not guaranteed that the trend towards human beings will not be a temporary fashion in this market where people are also buyers. Rights and responsibilities, dignity and immunity, freedom and autonomy manifest themselves as concepts that people will share with machines in the future. These concepts are the concepts that make people "human", distinguish them from "other" beings, value them and even have been used in

[262] BBC News, Robots Could Deman Legal Rights, http://news.bbc.co.uk/2/hi/technology/6200005.stm
[263] Çağlar Ersoy, *Robotlar, Yapay Zeka ve Hukuk,* p.142. (footnote)
[264] Peter W. Singer, *Wired For War,* p.403.
[265] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda", http://dx.doi.org/10.2139/ssrn.3180741
[266] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda".
[267] Mathias Risse, "Human Rights and Artificial Intelligence: An Urgently Needed Agenda".
[268] Nick Bostrom, *Superintelligence,* p.160.

response to the question of what the nature of human is, and for the first time in history, people have opened these concepts to something other than their own.

"What is the nature of man today?" This question is a new question that looks for answers in the face of artificial intelligence, but also takes into account bio-technology studies. *Brave New World* by Aldous Huxley, an example of a future-oriented dystopian fiction, is a good example of understanding how bio-technology will create a world in the future. The novel, which focuses on biotechnology as well as machine technologies, says that what we call human today will be completely different in the future. According to the story in the novel, people come into the world through the method of hatching, not through the womb, who do not experience feelings such as pain and sadness, and live in permanent happiness thanks to the soma pill. While Huxley describes a world in which a human is no longer a human, Francis Fukuyama opens up the question in his book[viii], *Our Posthuman Future*: *"Why is it so important to be a human in the traditional sense, as defined by Huxley? After all, the human race is the product of an evolutionary process that has been going on for millions of years and will likely extend into the future."*[269] Fukuyama states that the power of today's technologies to change human nature is a threat, what is called human nature is a complex whole, not the sum of simple parts, and that this whole itself is what causes honor, and human dignity is under threat.[270] Habermas, who interprets the technologicalization of human nature as "moralise of human nature" in the words of Wolfgang van den Daele, states that this is "re-sacralization".[271] According to Habermas, who says *"The border between our nature that we exist as ourselves and the organic material we give to ourselves is blurred,"* socio-moral determinations that can prevent biotechnological developments are blurred for economic and medical reasons and are ignored for the sake of developments.[272]

Artificial intelligence technologies and biotechnological human's nature to which artificial intelligence relates, lead to a rightful anxiety about human nature, human dignity as a carrier of human originated values, and human rights.

[269] Francis Fukuyama, *İnsan Ötesi Geleceğimiz*, Translated by: Çiğdem Aksoy Fromm, ODTÜ Yayıncılık, 2003, p 7.
[270] Francis Fukuyama, *İnsan Ötesi Geleceğimiz*, p.212.
[271] Jurben Habermas, *İnsan Doğasının Geleceği*, p.39.
[272] Jurben Habermas, *İnsan Doğasının Geleceği*, p.35, 37.

Undoubtedly, the increasing unemployment problem and inequality created by this technology are the leading human rights violations arising from Artificial Intelligence. These technologies, which increase the existing problems of humanity in a sense, also bring new subtitles to human rights. On the one hand, machine ethics takes on the questions such as "What should the robots that work at home care do if the human patient refuses to take their medication?"[273], human rights suggestions are made for the "digital age". Gerd Leonhard gathers digital human rights proposals under five headings:

1- The right to remain natural (biological).

2- The right to be inefficient when it requires a basic definition of humanity. (Suggestion not to make efficiency more important than human).

4- The right to disconnect (the right to go dark in the network, close the connection, stop communication, tracking and monitoring).

5- The right to be anonymous (Identity remains confidential, the right to not get tracked)

5- The right to hire or to incorporate people instead of machines.[274]

The presence of machines in almost all the workplaces where people operate and the replacement of people to a large extent seem to be possible at the Human Level Artificial Intelligence stage and the full autonomy that will take place at this stage. Today, we are at the stage of Artificial Narrow Intelligence in terms of autonomy robotic engineering, which points to a legal and at the same time an ethical subject that shows the courage to use reason as Kant's phrase, *"Once activated, it can operate in the real world at least in a certain area without any external control for long periods of time"* while pointing out the human today is used for robots.[275]At the stage we are in today, artificial intelligence structures are semi-autonomous, heterogeneous entities as Kant expresses, but opinions that these

[273] Gerd Leonhard, *Teknolojiye Karşı İnsanlık*, p.163.
[274] Gerd Leonhard, *Teknolojiye Karşı İnsanlık*, p.169, 170.
[275] Çağlar Ersoy, *Robotlar, Yapay Zeka ve Hukuk*, p.36.

machines will turn into willful, autonomous, rights and responsibility bearing and even honorable lawful and ethical subjects in the near future have taken their place in the literature.

CONCLUSION

She said come to me

She said stay to me

She said laugh to me

She said die to me

I came

I stayed

I laughed

I died

Haldun M. Özaktaş in his article titled "Artificial Intelligence: Mind-Body Problem in the Information Age" states *"These lines are so grammatically simple that it can be asserted that they can be produced by a computer program. Indeed, it is hard to argue precisely that these strings cannot be produced by a computer."*[276] These are the lines of the poem Nazım Hikmet wrote to Vera. Özaktaş asks "If these lines were produced by a computer program, would it have the same effect on us?"[277] This question is an important question for understanding the difference between artificial intelligence and human intelligence, the thinking activity of machines and the thinking activities of humans. Our knowledge of the poet's life story, our human sympathy and love for him, of course, affect the influence that poem has on us.[278] One of the most important criticisms directed to cognitive scientists who state that the information is processed according to which algorithm is more important than the hardware (meat, bone or silicone) that processes the information, is the contribution of the context to the formation of meaning, as in the case of poetry. Stating that imitating the act of understanding does not cause a

[276] Haldun M. Özaktaş, "Yapay Zeka: Bilgi Çağında Akıl-Beden Sorunu", Cogito Dergisi, Yapay Zeka, 1998, Yapı Kredi Yayınları, p.83.
[277] Haldun M. Özaktaş, "Yapay Zeka: Bilgi Çağında Akıl-Beden Sorunu", p.83.
[278] Haldun M. Özaktaş, "Yapay Zeka: Bilgi Çağında Akıl-Beden Sorunu", p.83.

meaning, Ercüment Aytaç says that in order for a language to make sense for artificial intelligence, the following conditions must be met;

*All sentences of a language evolve to the realities of this world. Otherwise, the act of comprehension cannot be materialized between imitation and the real world.

*It is possible to have a representative expression model for all the realities of this world. Otherwise, it cannot be imitated for the act of real understanding.

*A sentence of a language can become comparable to representative expressions through a generally valid algorithm. Otherwise, each sentence should be made comparable to representative expressions with a unique algorithm. Because of the countless sentence possibilities in a language, the translation becomes unrealizable by the computer.

*It is possible to create systematic operations to obtain valid questions that need to be determined after comparison. If valid results cannot be determined, we have nothing but a useless model that does not produce meaning.[279]

The conditions expressed by Aytaç are not fulfilled based on Narrow Scoped Comprehensive Artificial Intelligence. Whether these conditions can be fulfilled by Human Level Artificial Intelligence or Super Intelligence is among the issues that can become clear in the future. Machines of reason, which are capable of the "thinking" action mentioned in artificial intelligence studies at the level reached today, are used a synonym with computational activity. The transcendental mechanisms of Kant, which is one of the schemes proposed in order to improve the mental activity of the machines in Artificial Intelligence studies, seem to contribute in some ways to the existing studies. Even though Kant's mental approach, which makes all consciousnesses, all times the same and identical, gives a mechanical scheme for the mind, it should be accepted that this representation refers to a situation that exceeds the stage reached in cognitive science. It is not possible to

[279] Ercüment Aytaç, "Yapayazar", Cogito Dergisi, Yapay Zeka, 1998, Yapı Kredi Yayınları, p.134.

say that autonomy, which is one of the basic concepts of Kantian ethics, is completely different from the autonomy used as the main target in artificial intelligence studies. Semi-autonomous machines operating today are targeted to reach the stage of transition to full autonomy in the next ten or twenty years. What is meant by full autonomy is an independent will and thinking activity. At this stage, there are different views as to whether machine autonomy can have a value system or a moral world. The fact that the machines are ethical subjects in the context of Kant ethics necessitates them to have wisdom and obey unconditional command. The answer to the question of "law" which cannot be practiced in a human subject in practice and whether duty morality can be implemented in another type of mind requires the machines to be fully autonomous. Kant is a philosopher mentioned in machine ethics discussions even in the present stage of artificial intelligence studies. Non-lying bots which are referred to in the literature as "Kantian Machines" are seen as ethical digital elements of today and the future. Although Kant Machines cannot be associated with morality in terms of Kant in terms of performing a task assigned to them rather than duty morality, it may be a guide to consider Kant's ethical understanding by excluding the question of freedom - in establishing norms regarding machine ethics.It is not yet known whether autonomy will bring a sense of freedom to machines, and to what extent this autonomy will resemble the mentioned human autonomy. In the *Groundwork for the Metaphysics of Morals,* Kant states,

"*If one adds that unless one wants to dispute whether the concept of morality has any truth and relation to any possible object, one could not deny that its law is of such an extensive significance that it would have to be valid not merely for human beings but for all rational beings in general, and not merely under contingent conditions and with exceptions, but with absolute necessity.* "[280]

It can be thought that Kant included God as an intelligent entity other than human in the category of existence in the expression "*not only for people, but for all intelligent beings in general*". But this generalization-the general validity of the

[280] Immanuel Kant, *Groundwork for the Metaphysics of Morals*, p.85.

law depends on this, it seems to include intelligent machines. Of course, there is no problem in the moral ownership of machines that are defined as intelligent assets that can make their own decisions. However, it does not seem impossible for an artificial intelligence at this level to name itself as a species, moreover to “name” the humans. Whether machines can see human beings "not as a tool but end" or whether machine dignity can be talked about is among the topics discussed in the literature today. Based on all these discussions, we can express the following; it is the first time in human history that humanity consciously or unconsciously shared the qualities used to distinguish itself from other things in nature such as reason, thinking, autonomy, morality, freedom, and dignity. At the same time, humans attempted to transform machines into something similar to themselves, but also attempted to transform humans into something similar to machines. Studies show that human beings will be different from the ones we know today and that the “human nature” that forms thebasis of human dignity and human rights will change.

FINAL NOTES

[i] Sir David Ross in his work *Aristotle* when stating that; *"The adjective corresponding to Eudaimonia means the person who is protected by a good angel (daimon) in origin,"*emphasized that the concept's translation as "happiness" creates some problems. Pierre Hadot in his work "*What Is Ancient Philosophy?"* says that Eros is a daimon. Hadot, underlines that a Daimon is a creature that stands somewhere in the middle between godsand humans, mortals and immortals and states that *"There is not only a middle road between to opposing orders, but an intermediary as well."* We encounter the fact that Daimon mediates in the dialogue between Socrates and Diotima in Plato's *Symposium*. Daimon stands between mortal and immortal as a great force. This intermediary, who brings word from gods to people, and from people to gods, is in the middle of knowledge and ignorance.

[ii] Kant's explanations regarding "respect" can be found in the 2nd annotation p. 16 of the Groundwork for the Metaphysics of Morals.

[iii] Artificial Intelligence (AI) is defined as the study of how computers are to be built and/or programmed to enable them to do what minds can do. (The Philosophy of Artificial Intelligence, Edited by Margaret A. Boden, Oxford Readings in Philosophy, Oxford University Press, 2005.)

[iv] While Bostrom stated that Super Intelligence can be reached after reaching Human Level Machine Intelligence, there are also those who state that these two types of intelligence can be reached at the same time.

[v] Ahmet Ayhan Çitil, uses the concept "place" by separating it from the termin "space", to mean a ground or place where an object exist. (Mathematics and Metaphysics). Use of the word is based on this meaning.

[vi] Bots: Programs that perform a given task. Bots used in indexing technologies by search engines are semi-autonomous.

[vii] A reference to the Münchausen Syndrome. Münchausenyndrome is when a person makes himself/herself constantly and deliberately ill to receive attention from their environment or healthcare professionals. Münchausen syndrome, which is a psychological disorder, was named after Baron Munchausen, a fictional German nobleman created by the German author Rudolf Erich Raspe in the book of Baron Munchausen, published in 1785, about his Fabulous Travels and Struggles in Russia.

[viii] Francis Fukuyama adopts a critical approach to biotechnological developments.

REFERENCES

Arkin Ronald C., *Governing Lethal Behavior in Autonomous Robots*, CRP Press, 2009

Aristotle, *Nicomachean Ethics*, Book 1, Translated, With An Interpretive Essay, Notes, And Glossary By Robert C. Bartlett And Susan D. Collins, The University Of Chicago Press, 2011.

Aristotle, *The Politics*, Translated by T. A. Sinclair, Revised and Re-Presented by Trevor J. Saunders, Penguin Books, 1992.

Asimov Isaac, *Ben, Robot*, İthaki Yayınları, Translated by: Ekin Odabaş, 2016.

Aytac Ercumet, "Yapayazar", Cogito Dergisi, Yapay Zeka, 1998, Yapı Kredi Yayınları.

BBC News, Robots Could Deman Legal Rights, http://news.bbc.co.uk/2/hi/technology/6200005.stm

Bendel Oliver, Schwegler Kevin, Richards Bradley, "Towars Kant Machines", The AAAI 2017 Spring Symposium on Artificial Intelligence for the Social Good Technical Report SS-17-0, School of Business FHNW, Bahnhofstrasse 6, CH-5210 Windisch.

Betz Frederick , *Teknolojik Yenilik Yöntemi,* Tübitak Popüler Bilim Kitapları, 2010.

Betz Frederick, *Managing Science*, Illustration: Kant's Critique of Pure Reason, Springer Media, 2011.

Bostrom Nick, *Superintelligence*, Paths, Dangers, Strategies, Oxford University Press, 2014.

Carthy John Mc, “What has AI in Common with Philosophy?” Stanford University, http://jmc.stanford.edu/articles/aiphil/aiphil.pdf

Clapham Andrew, *Human Rights,* A Very Short Introduction, Oxford University Press, 2007.

Copeland Jack, *Artificial Intelligence, A Philosophical Introduction,* 1, Blackwell Publishers, 1993.

Cevik Ahmet, *Matematik Felsefesi ve Matematiksel Mantık*, Nesin Yayıncılık, 2019.

Citil Ahmet Ayhan, *Matematik ve Metafizik*, Kitap 1: Sayı ve Nesne, Alfa Basım Yayım Dağıtım, 2012

Douzinas Contas, *İnsan Hakları ve İmparatorluk*, Çevirenler: Kasım Akbaş, Rabia Sağlam, Bilgi Üniversitesi Yayınları, 2017.

Dreyfus Hubert L., “Simon’s Simple Solutions”, Standford Education: https://web.stanford.edu/group/SHR/4-1/text/dreyfus.commentary.html

Erdogan Mustafa, *İnsan Hakları Teorisi ve Hukuku*, Hukuk Yayınları, 2018.

Ersoy Caglar, *Robotlar, Yapay Zekâ ve Hukuk*, On İki Levha Yayıncılık, 2017.

Evans Richard, “A Kantian Cognitive Architecture”, On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence, Themes from IACAP 2016, Springer Nature, 2019.

European Parliamet, European Civil Law Rules in Robotics: https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf

Fukuyama Francis, *İnsan Ötesi Geleceğimiz*, Translated by: Çiğdem Aksoy Fromm, ODTÜ Yayıncılık, 2003.

Gore Al, *Gelecek*, Translated: Çağlar Kök, Ebru Kızıldağ, MediaCat Kitapları, 2014.

Gozkan Bulent, *Kant'ın Şemsiyesi*, Yapı Kredi Yayınları, 2018

Guzeldere Guven, "Yapay Zeka'nın Dünü, Bugünü, Yarını", Cogito Dergisi, Yapay Zekâ, 1998.

Guyer Paul, *Kant's System of Nature and Freedom*, Selected Essays, Oxford University Press, 2005.

Guyer Paul, *Kant*, Edited by Brian Leiter, Published by Routledge, 2006.

Habermas Jurben, *İnsan Doğasının Geleceği*, Translated by : Kaan Ökten, Everest Yayınları, 2003.

Haslakoğlu Oguz, *Platon Düşüncesinde Tekhne*, Sentez Yayınları, 2016.

Hayles N. Katherine , "Anlamın Bedenleşmesi" Cogito Dergisi, Yapay Zeka.

Heidegger Martin , *Teknik ve Dönüş*, Çeviren: Necati Açа, Bilim ve Sanat Yayınları, 1998.

Heimsoeth Heinz, *Kant'ın Felsefesi*, Translated by: Takiyettin Mengüşoğlu, Doğu-Batı Yayınları, 2007.

ICRC: 'Report of the ICRC Expert Meeting, "Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects".

Kalaycı Nazile, *Daimon'dan Eudaimonia'ya: Aristoteles'te Mutluluk*, Cogito, Aristoteles, Yapı Kredi Yayınları, 2014.

Kant Immanuel, *Critique of Pure Reason,* Translated and Edited by Paul Guyer, Allen W. Wood, Cambridge University Press, 1998.

Kant Immanuel ; *Crituqe of Practical Reason,* Conclusion, Translated; Werner S. Pluhar, Hackett Publishing Company, 2002.

Kant Immanuel, *Groundwork for the Metaphysics of Morals,* Edited and translated by Allen W. Wood,Yale University Press, New Haven and London, 2002

Kant Immanuel, *Critique of the Power of Judgment*, Edited by Paul Guyer, Tranlated by Paul Guyer and Eric Matthews, Cambridge University Press, 2000.

Karatani Kojin, *Transkritik,* Çeviri: Erkan Ünal, Metis Yayınları, 2017.

Kapoor Deepak N., Singh Arvind , Singh Sushila , "Nanorobotics": https://www.researchgate.net/publication/227245023_Nanorobotics

Kateb George, *Human Dignity*, The Belknap Press of Harvard University Press, 2011.

Kılıc Yavuz, "Kant'ın Etik Görüşünde Değerli Eylemin Olanağı", Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Sayı 22, 2015.

Kurzweil Ray, *İnsanlık 2.0,* Alfa Basım Yayın Dağıtım, Çeviren: Mine Şengel, 2016.

Lee Peter, "Armed Drones: Automation, Autonomy, and Ethical Decision-Making," *The Political Economy of Robots,* Ryan Kigings, Palgrave Macmillan, published by Springer Nature, 2018.

Leonhard Gerd, *Teknolojiye Karşı İnsanlık*, Çevirenler: Cihan Akkartal, İlker Akkartal, Siyah X Yayınları, 2018.

Living Systems:

https://www.pnas.org/content/117/4/1853

Minsky Marvin, *The Emotion Machine,* Simon & Schuster, 2006.

Newell Allen ve Simon Herbert A. "Computer Scince as Empirical Enquiry: Symbols and Search", the Tenth Turing Lecture, first published in Communications of the Association for Computing Machinery 19 Inc. (Mar. 1976) Reprinted by permission of Oxford University Pres; *The Philosophy of Artificial Intelligence*, Edited by Margaret A. Boden, (2005).

Nilsson Nils J., *Yapay Zekâ Geçmişi ve Geleceği*, Translated by: Mehmet Doğan, Boğaziçi Üniversitesi Yayınları, 2011

OODA Loop:

https://fasttransients.files.wordpress.com/2018/05/boyds-real-ooda-loopx.pdf

Ozaktas Haldun M., "Yapay Zeka: Bilgi Çağında Akıl-Beden Sorunu", Cogito Dergisi, Yapay Zeka, 1998, Yapı Kredi Yayınları.

Ozlem Dogan, *Kant Üzerine Yazılar*, Notos Kitap, 2014.

Powers Thomas M., "Prospects for a Kantian Machines", Intelligent Systems, IEEE 21(4):46- 51 · August 2006,

https://www.researchgate.net/publication/3454367_Prospects_for_a_Kantian_Machine

Parasuraman Raja, Sheridan Thomas B., Fellow, IEEE, and Wickens Christopher D., *A Model for Types and Levels of Human Interaction with Automation*;

https://www.researchgate.net/publication/11596569_A_model_for_types_and_levels_of_human_interaction_with_automation_IEEE_Trans_Syst_Man_Cybern_Part_A_Syst_Hum_303_286-297

Rescher Nicholas, “Leibniz's interpretation of his logical calculi,”

https://www.cambridge.org/core/journals/journal-of-symbolic-logic/article/leibnizs-interpretation-of-his-logical-calculi/611475C44E0E52824948975946E6FF46#

Risse Mathias, “Human Rights and Artificial Intelligence: An Urgently Needed Agenda”:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3180741

Say Cem, *Yapay Zekâ*, Bilim ve Gelecek Kitaplığı, 2018.

Searle John, “Bilgisayarlar Düşünebilir mi?”, Translated: Liz Amado, Cogito Dergisi, Yapay Zeka, 1998.

Searle John, “Mind, Brains, and Programs”, *The Philosophy of Artificial Intelligence*, Oxford University Press.

Singer Peter W., *Wired For War: The Robotics Revolution and Conflict in the Twenty-first Century*, The Penguin Books, 2010.

Smith Brain Cantwell, “Sap ve Saman Meselesi”, Translated by: Elif Özsayar, Cogito Dergisi, Yapay Zeka.

Tegmark Max, *Life 3.0, Being Human in the Age of Artificial Intelligence*, Borzoi Book by Alfred A. Knopf, 2017.

Tepe Harun, *İnsan Hakları Felsefesi*, Bilge Su Yayınları, 2018.

Turing A.M., *‘Computing Machinery and Intelligence’ from Mind LIX*, no, 2236 (1950): 433-60. Reprinted by permission of Oxford University Pres; The Philosophy of Artificial Intelligence, Edited by Margaret A. Boden, (2005).

The Guardian:

https://www.theguardian.com/science/2020/jan/13/scientists-use-stem-cells-from-frogs-to-build-first-living-robots

The Society Pages:

https://thesocietypages.org/cyborgology/2012/09/12/recycling-ellen-ullman/human-machine-interface/

Ulgen Ozlem: "Kantian Ethics in the Age of Artificial Intelligence and Robotics"

https://www.researchgate.net/publication/321883098_Kantian_Ethics_in_the_Age_of_Artificial_Intelligence_and_Robotics

Vernant Jean Pierre, *Eski Yunan'da Mit ve Din*, Translated by: Murat Erşen, Alfa Yayınları, 2016.

Wood Allen W. Wood, *Kant*, Blackwell Publishing, Edited by Steven Nadler, 2005.

Yesilcayır Celal, *İnsan Hakları Felsefesi,* Çizgi Kitapevi Yayınları 2019.

Zambak Aziz Fevzi & Vergauwen Roger, "Artificial Intelligence and Agentive Cognition: A Logico-Linguistic Approach", *Logique et Analyse*, Vol: 52, pp. 57-96, 2009. (ISSN: 0024-5836)