Chankyu Lee*, Hyeongjoo Kim**

Groundwork of Artificial Intelligence Humanities[1]

SUMMARY

The aim of the Artificial Intelligence Humanities is to present humanistic suggestions for our lives and academia regarding how humans will live in a smart city surrounded by artificial intelligence technology. With that purpose in mind, we introduce the groundwork for the Artificial Intelligence Humanities and share our related concerns. As a concept, "Artificial Intelligence Humanities" comprises three elements: "artificial intelligence," "content of the humanities," and "methodology of the humanities." The "content of the humanities," comprises the derivatives of the traditional humanities, namely, linguistics, literature, history, ethics, sociology, and cultural studies. These five research areas, along with the discipline created through the traditional humanities' acceptance of the changes brought by the age of intelligence, are derived by reflecting on oneself and the world, and they are applied to the world as it exists today in the age of artificial intelligence and the humanities.

Keywords: Artificial Intelligence Humanities, content of the humanities, methodology of the humanities, Posthumanism.

1. Introduction

The goal of this paper is to propose a framework for the "Artificial Intelligence Humanities(AIH)", an area of study that traces those significant changes in our life which occur due to the constant advancement of AI technology and aims to present a constructive image of human and society in the future AI era. Along with this framework, we shall attempt to pave the ways for active participation in technology development in order to prevent AI from threatening humanity.

* Department of Korean linguistic and literature, Chung-Ang University, Seoul, Republic of Korea.

** Humanities Research Institute, Chung-Ang University, Seoul, Republic of Korea

Correspondence Address: Chankyu Lee, Department of Korean linguistic and literature, Chung-Ang University 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea. E-mail: leeck@cau.ac.kr.

In this paper, we shall first observe those key arguments of posthumanism, which are very closely associated with AI phenomena and advocate the negation and reform of the traditional humanities. Results of this observation, however, indicate that the core claims of posthumanism are in practice not quite different from those advocated by the traditional humanities. In other words, the results reveal that posthumanism tries only to reconfigure itself to accommodate the era of AI by positively accepting the achievement of the traditional humanities. Alternatively, we shall argue that AIH takes a path that can be clearly differentiated from that of posthumanism. AIH succeeds in that dichotomous attitude of making a sharp distinction between matter and form as well as between content and method, which has long been maintained since the time of Aristotle. According to this tradition, we understand AIH in two separate concepts, namely the content of AIH, which comprises its academic characteristic, and its method, which consists in its academic attitude. We shall argue that the content of AIH includes the principal branches of the traditional humanities, such as linguistics, history, culture, philosophy as general science which includes anthropology and psychology, ethics, sociology, and culturology, and that its method is a self-reflective attitude. We refer to the former as world-hermeneutics and the latter as the reflective character of humanities. When the focus of reflection in the humanities is on the present-world, AI is absorbed into the above disciplines of the humanities, and is restructured and referenced as AI Technology Criticism, AI Relationships and Communication Studies, AI Socio-Cultural Studies, AI Ethics, and AI Data Hermeneutics.

2. Humanistic Characteristics hidden in the Posthumanism

N. Bostrom, who established the World Transhumanist Association and leads posthumanistic discourses nowadays, firmly states that progress in technology can enable the advent of posthumans and must do so. As with all others, because the discourse on posthumanism has been varied and developed widely, it is difficult to devise a partial definition of it at this point. But this theory undoubtedly begins with the premise that the traditional concept of human that was constructed following the Renaissance must be subjected to some modifications. A brief definition of posthumanism based on existing theories may be suggested as follows.

> Posthumanism is based on the premise that all kinds of dichotomy are violent illusions that must be dismantled, and narrowly dismantles the dualistic opposition of the body and mind, and breaks the boundaries between "material and non-material, animal and human, and even organism and machine"[2] to

---

2   Haraway, Donna (1995). *in Wolfe*, 36. Quoted from *Posthumanism*, 190.

dismantle "dual confrontation between human and non-human"[3]. From this perspective, it seeks to create and develop posthumans who overcome the limitations of the biological form using the achievements of biotechnology and artificial intelligence

As predicted by the fact that posthumanism shares a prefix with postmodernism, which emerged from the philosophical discussion of repetition of 'action and reaction', it was raised from the ideological soil of postmodernism. Many scholars, including S. Herbrechter, S. Sorgner, state that the philosophical starting point of posthumanism is the philosophy of Nietzsche, a pioneer of postmodernism.[4]

> Humans of the future rescue us from past reason. We are rescued from the inevitability to grow further, in other words, intense regurgitation, intent to head towards futility, and nihilism. The bell of correction and great resolution frees our will again and returns objectives to the earth and hope to humans. Antichrist, anti-nihilist, and those that overcome god and nihilism will inevitably appear...[5]

Some traces of Nietzsche's philosophy on the 'Waiting for Übermensch' still remain in the modern spirit: "The noble ideal itself has become incarnated" and "the synthesis of Unmensch and Übermensch is Napoleon."[6] The radical criticism of humanity (human beings) tacitly assumes a better us. The desire to reject oneself and create a better self always depends on the awareness of the self. Self-perfection through self-negation is a transformational imitation of the progressive ideas of the Enlightenment. This negative-positive ambivalence appears more distinctly as the self-destructive nature of posthumanism appears stronger. Let us now examine the arguments of transhumanism, which attempt to present a certain blueprint for maximizing human happiness through the denial of common sense concerning the concept of a human being.

N. Bostrom defines it as "intellectual and cultural movements affirming the possibility and desirability of fundamentally improving human conditions by developing and expanding technologies that eliminate human aging and improve

---

3 Badmington, Neil (2010). Pod Almighty!; Or, Humanism, Posthumanism, and Strange Case of Invasion of the Body Snatchers, *Textual Practice,* 15 (1), 5-22.

4 See Herbrechter, Stefan. (2012). 31. See Sorgner, (2009) Nietzsche, the Overhuman, and Transhumanism, *Journal of Evolution & Technology*, 20 (1), 29-42. Volume 20 of the *Journal of Evolution & Technology*, an academic journal on modern technology ethics discourse, was a special volume on Nietzsche's effects on transhumanism. On the contrary, Bostrom stated that Nietzsche's Übermensch concept has only "surface-level similarities" with transhumanism and reduced its effect. Bostrom, Nick (2005). A history of transhumanist thought, *Journal of Evolution & Technology,* 14 (1), 4 4, https://nickbostrom.com/papers/history.pdf

5 Nietzsche, Friedrich (1906), *Zur Genealogie der Moral* in: Nietzsche's Werke B. VIII, Leipzig; C. G. Naumann Verlag, 476.

6 Ibid., 337.

the intellectual, physical, and psychological abilities of humans."[7] Transhumanism supports the enhancement of the human condition through physical prosthesis and orthoses, based on the needlessness of the boundary between the machine and the biological body. In this regard, transhumanists also pursue disembodiment through the uploading of one's consciousness. However, if we contemplate this idea in reverse, it can be said to follow the modern idea of the mind-body dichotomy. In fact, although the blueprint of the future society proposed by Bostrom is the same as that described above, his diagnosis of reality is truly realistic. In his view, "Some authors write as though simply by changing our self-conception, we have become or could become posthuman. This is a confusion or corruption of the original meaning of the term. The changes required to make us posthuman are *too profound to be achievable* by merely altering some aspect of psychological theory or the way we think about ourselves. *Radical* technological modifications to our brains and bodies are needed."[8]

As Bostrom clearly states that "transhumanism, (which pursues the fulfillment of the ideal of posthuman: add. From Kim), can be viewed as an extension of humanism, from which it is partially derived"[9]. While the autonomy of the individuals can be taken as a concept perfecting the modern spirit and is indeed the highest value of that spirit, the combination of AI and interface technology produces a tool that may be used to this end.[10] In this way, posthumanism, which is based on anti-humanism, returns to humanism as it shows its true colors. Although those "humans of the future" of which posthumanism speaks may not be "us," they are still human. And although posthumanism is not the humanism of the past, it is still a humanism.

In the next chapter, we will argue that artificial intelligence humanities shares many resemblances with posthumanism. However, there are crucial differences, as follows:

> The posthumanism, which is actually strongly baptized by humanism, advocates anti-humanism. However, on the other hand, AIH embraces the achievements and traditions of humanistic traditional humanities. And it makes social and cultural phenomena due to artificial intelligence the objects for its academic discussions.

---

7   Bostrom, Nick (2003). The transhumanist FAQ. In Bostrom, N, *Nick Bostrom's Home Page*, https://www.nickbostrom.com/views/transhumanist.pdf, (accessed: 3 July 2019).

8   Ibid.

9   Ibid.

10   See. Ibid.

3. Contents and Methodology of AIH

Firstly, once we attempt a simple analysis of the words AIH, two components are drawn: "artificial intelligence" and "humanities." Secondly, "artificial intelligence" is regarded as an object confronting the "Humanities," which is the reflective inquiry into the era. Finally, we employ the dichotomous attitude of the traditional humanities, which began with Aristotle, and again divide the elements that constitute the humanities into two kinds. Specifically, by means of an analysis of the academic attitude throughout the whole period of the modern Western humanism, during which the academic questioning of the humanities became full-scale and subsequently intensified, the content of the humanities is separated from its method; as such, the contents of the humanities discipline are defined as world-hermeneutics and its method as reflection. Artificial intelligence has three components: "artificial intelligence, world-hermeneutics, and the reflective character of humanities."

3.1 Deduction of Contents of AIH

It is common to refer to "Studia Humanitas," which is known to be first used by Cicero, as the origin of today's commonly used word "humanities." When it was first uttered, Humanitas meant "Favorable human behavior patterns (habits)." In this context, it may also be noted that it was he who first introduced the term 'philosophia moralis", which will be dealt with later.[11]   Here it was dissolved in ancient Greek traditions to view Humanitas as a virtue(arete) of a liberal, that is, "someone who can use Logos." In fact, that discipline inherited the essence of paideia, which takes teaching humanity and excellence as its objectives. Simply put, "Studia Humanitas" is a Latin translation of the Greek word "paideia".[12] The latter word was often taken to be synonymous with culture and education. Considering that these two words of the latter have a similar meaning in the Greek tradition, we can understand 'paideia' as the meaning of 'self-cultivation' to improve human-virtue.[13] Cicero defined it as "a term that includes all liberal arts, in other words, topics appropriate for the education of free people." This became the root of the

---

11 Conze, Werner and Koselleck, Reinhart eds., (1982), Menschheit, Humanität, Humanismus in *Geschichtliche Grundbegriffe: Historischers Lexikon zur politisch-sozialen Sprache in Deutschland,* Bd. 3, Stuttgart; Klett-Cotta, 1064. Cicero coined the word philosophia moralis as a translation of the Greek word ἠθική(ēthikós). Mos (or Morse) is the root word of the moralis. Ethik, Moral, and Sitten in modern German have the same etymology with Humanitas, which means "the Favorable Human behavior patterns" mentioned in the text. See Diemer, Alwin, Historische Wörterbuch der Philosophie online. (https://www.schwabeonline.ch/schwabe-xaveropp/ elibrary/start.xav#__elibrary__%2F%2F*%5B%40attr_id%3D%27verw.moral.moralisch.moralphilosophie%27 %5D__1585115242588)

12 However, regarding the approach that the meaning of humanitas does not necessarily coincide with paideia, see Conze, Werner and Koselleck, Reinhart, eds., (1982), 1066.

13 See Jaeger, Werner (1973). *PAIDEIA*, Berlin; De Gruyter, 6, 384.

seven preliminary learnings: Trivium, which includes grammar, logic, and rhetoric, and Quadrivium, which includes geometry, algebra, astronomy, and harmonics. In the Middle Ages, Trivium was applied in the teaching of theology and philosophy. However, medieval philosophy could not serve as anything more than an explanatory instrument for theology. In this sense, Trivium, which can be narrowly defined as a medieval version of the humanities, was used in a negative sense. As living beings, humans must seek Studia Divinitatis, the study of the divine, while humanitas was regarded as something that should be denied. Thus, trivium became trivial, and humanitas came only to concern something corporeal (Fleischliches). However, as is commonly known, the humanities revived the spirit of the ancient humanitas during the Renaissance and established itself as a field of study. It comprehended literature, grammar, moral philosophy, linguistics and rhetoric, poetics, history, and politics, and became a principal source upon which a great portion of the university curricula is now based.[14] Over time, the ideology of human progress, which appeared as a reaction to the absolute pursuit of divinity, gradually evolved into the pursuit of the "good present world", which moved humanities' focus towards practicality.[15] As this tradition was continued, practical anthropology was settled, for instance, by I. Kant in the eighteenth century as a part of moral philosophy.[16] According to him, anthropology should be regarded as an area of humanities.

Meanwhile, as medicine and architecture were excluded from the nine liberal arts in ancient Rome, the seven subjects survived in the medieval colleges; of these, grammar, rhetoric, and logic became trivium. Some disciplines have been added since then, and history and process analogous to the modern re-establishment of the humanities continue to take place today. In the German tradition, Humanitas is translated into mental science (Geistwissenschaft(en)). This came from J. Schiel's translation of J. S. Mill's *Moral Science*, according to the testimony of a well-known modern philosopher, W. Dilthey. However, when traced to its origin, the first instance of the word appeared in the 18th century. This concept is referenced as being distinct from the study of God in the paper "Wer sind die Aufklärer?," which was published anonymously. At that time, it was part of the Weltweisheit, and was used as a synonym for philosophy; it comprised rationalism, naturalism, law, and politics. In the early 19th century, E. A. E. Calinich later published a book entitled "*The Method of Natural Science and Mental Science*." At this point in time,

14   See Kimball, Bruce A. (1986). *Orators & Philosophers: A History of the Idea of Liberal Education*, New York; Teachers College Press, 78, and Hiorth, Finngeir (1996) *Humanismus – Genau Betrachtet*, Neu-Isenburg; Angelika Lenz Verlag, 57.

15   Grafton, Anthony and Jardine, Lisa (1986). *From Humanism to the Humanities*: *Education and the Liberal Arts in Fifteenth- and Sixteenth Century Europe*, Cambridge; Harvard University Press.

16   See Kant, Immanuel (1900). *Grundlegung zur Metaphysik der Sitten(AA4)* in: Kants gesammelte Schriften (Sog. Akademie-Ausgabe), Berlin; Walter de Gruyter, 388.

the distinction between the two made by Dilthey was fixed, and natural science was naturally excluded from the scope of mental science. However, as Goethe said, "it has already been a century since liberal humanity has no longer moved human feelings;" the humanities tends to be positivist, in that all actions and emotional behaviors of humans can be reduced to physical and physiological realities by the influence of the newly developed natural sciences. This movement continues to this day, and the posthumanism discussed above is also under its influence. Moreover, as N. Hartmann's work in the 20th century can be applied here, sociology has also fallen into the realm of the humanities through the introduction of the term, "social psychology" in Marxian philosophy.[17] In addition, rational psychology, which was a part of metaphysics along with cosmology and theology since Christian Wolf's time, was established as an essential part of philosophy by I. Kant. W. Wundt critically inherited this tradition in the name of experimental psychology. In this process, psychology was separated from philosophy. Although it may be conceived as social science in that it performs empirical research through objective experimentation. However, it can still be taken as an area of humanities in that its subject-matter is human psychology and spirit. Thus we regard psychology here as an area of humanities.

The history of the humanities is like a game of alternating between self-determination and self-denial. For example, Heidegger reads the death of God described by Nietzsche as the end of metaphysics. Foucault emphasizes "the end of man" in *The Order of Things*. E. Cassirer assessed Kant, who is commonly argued as having perfected the enlightenment, as a destructor of metaphysics. It is not radical or new to state that post-humanism, which is alleged to represent the humanities in the era of AI, claims to support the birth of a completely new discipline that denies its ideological foundation. This is in line with the slogan and legacy used throughout the history of the humanities. In Section 2, we have analyzed the essence of posthumanism, which has been developed by means of abolishing the traditional disciplines of the humanities and argued that it nevertheless maintains the essence of the traditional humanities. AIH positively embraces the spirit of the traditional humanities but takes a different path that posthumanism has followed. Therefore, AIH suggests that the existing disciplines of the humanities that we extracted above, such as linguistics, literature, history, ethics, sociology, anthropology, and psychology are its essential contents that should put on a new outfit in order to prevent themselves from being discarded with the spread of AI. The grounds for this new appearance and reinterpretation of the humanities can be found in the reflective attitude, which is a

---

17   Diemer Alwin, Historische Wörterbuch der Philosophie online. (https://www.schwabeonline.ch/ schwabe-xaveropp/elibrary/start.xav?start=%2F%2F*%5B%40attr_id%3D%27verw.geisteswissenschaften%27%5D#__ elibrary__%2F%2F*%5B%40attr_id%3D%27verw.geisteswissenschaften%27%5D__1550933851671).

symbol of modernity. If the humanities target the era, with which it is confronted, as a subject for reflection, it will reconstruct itself as a result. In this sense, the study of AI has a trait that allows the humanities to reconstruct itself in keeping with the era of AI.

3.2 Methodology of AIH

The Cogito proposition of R. Descartes, who wide opened the door to modern trends of thought, is still doing its "attractive magic"[18] in the Western history of thought. Human beings and other living beings called nature.[19] This took a monumental legacy of ancient philosophy—that is, the distinction between form(eidos) and matter(hyle) from Aristoteles[20]—as its nourishment and developed it into that methodological division of form and content, which is the foundation of modern academic construction. Remaining true to this tradition, we cite Kant as a philosopher committed to completing the humanities spirit in the age of Enlightenment.[21] For example, in *Grundlegung zur Metaphsysik der Sitten*, in order to build a new discipline called "moral metaphysics," he brought in natural science, ethics (moral philosophy), and logic which are three areas of research in ancient Greek philosophy[22], and by introducing the diagram of content and form into these areas, logic is divided into pure formal learning (natural science and ethics) and thereby becomes a discipline containing empirical elements. In other words, the application of the form-content diagram to ethics established "practical philosophy," which includes anthropology and extracts "moral metaphysics" from purely formal ethics.[23] By applying this attitude, we can reinforce the universal value of the research areas constituting the essential content of the humanities. In the question "what is learning?," the focus is laid on the "attitude of reflection," the foundational principle of academic work, according to Fichte, a self-styled faithful successor of Kant, who was preoccupied with the work of building a foundation for learning. His initial works, including *Ueber den Begriff der Wissenschaftslehre*, *Grundlage der gesammten Wissenschaftslehre*, etc.,[24] are an attempt

---

18   Brands, Hartmut (1982). *Cogito ergo sum*, Karl Alber, Freiburg, 10.

19   More detail, see Aristotle (1954). *The Works of Aristotle v. VIII Metaphisica (trans. from Ross)*, the Clarendon Press, Oxford, 980ᵃ-981bᵇ·

20   Aristotle introduces shape and material to explain substance as a substratum. It can be described in three different ways, namely matter, form, or a compound of them. In the process of argument, he points out the incorrectness of the view that substance is made up only of matter and implicitly supports the view that it is a compound of matter and form. As can be seen here, the dichotomy of form and matter becomes increasingly apparent as he reaches the conclusion of his argument. See Ibid., 1027ᵇ-1029ᵇ.

21   See Wundt, Mohr (1945). *die deutsche Schulphilosophie im Zeitalter der Aufklärung*, Tübingen; Olms, 1.

22   The fact that these three disciplines were a solid branch of Greek philosophy can be found in Heidegger's book, "Über den Humanismus" as well; Heidegger, M. (2000) *Über den Humanismus*, Klostermann, 8.

23   See Kant, Immanuel (1900).; 387-390.

24   Hereinafter will be referenced as *UBW* and *GLW*, respectively.

to justify the principle of self-reflection represented by the catchphrase "I am I" (Ich bin Ich). We take this principle of reflection as the methodology for establishing the concept of AIH.

Investigating the question "what is science and what conditions must it contain," Fichte suggested "I am" as its first absolute condition. He performed an inquiry into the fundamental and self-fulfilling conditions of science, and the principle he found was more commonsensical than it might appear at first. Similarly to the mindset that every being in the world depends for its existence on another being, he asserted that "one academic discipline has a systematic format and all the propositions that comprise it relate to one fundamental proposition."[25] Individual sciences, each of which has its own underlying proposition, depend on one another. In the end, Fichte's concern may be reduced to the science of science, i.e., the determination of its essence. Furthermore, it boils down to identifying the fundamental proposition that supports learning at its most basic level. He said at last that this fundamental proposition is "a clear proposition independently from it prior to the various dependencies it has,"[26] and that as a study of the learning in general, "the theory of learning is based on logic, a purely formal study."[27]

Based on this premise, he emphasized the principle of identity, the first proposition of logic, and the fundamental principle of all thoughts. Suggesting, for example, that the signification layer of the principle of identity "A is A" does not lie in either "A" or "there is A," but in the conditional sentence "if there is A, there is A," Fichte took note of the necessary relation between the antecedent and the consequent of this conditional statement. He then referred to the necessity of this relationship as X, and deduced that as long as the judge is the self, this necessity should at least be given within "me." In "A is A," then, the former A and the latter A have a necessary relation through X, and this relation is expressed by: "if A is established within the I, A is established."[28] If A is established in the I, A has been derived from the establishment

---

25 Fichte, Gottlieb W. (1971a), *UBW in: Fichtes Werk I*, Berlin; Walter du Gruyter, 38; "A science has systematic form. All propositions in it are connected in one single fundamental proposition" (Fichte, Gottlieb W. (1889a), Introduction: Concerning the conception of the science of knowledge generally, in: *The Science of Knowledge*, Kroeger, A.E trs., London; Trübner & Co,.Ludgate Hill, 58.)

26 See Ibid., 41; "Such a proposition, which has certainty before and independent of all connection, is a fundamental principle. Every science must have a fundamental principle; nay, it might consist of simply such one principle, which in that case could not be called fundamental, however, since it would not be the foundation of others. But a science also cannot have more than one fundamental principle, for else it would result in many sciences" (Fichte, G.W. (1889a). 62.)

27 Ibid., 46; "From this results the determined relation of logic to the science of knowledge" (Fichte, G.W. (1889a). 46.)

28 Fichte, Gottlieb W. (1971b). *GLW* in: *Fichtes Werk* I, Berlin; Walter de Gruyter, 91; "If A is posited in the Ego, then it is posited, or then it is."(Fichte, Gottlieb W., (1889b). Fundamental principles of the whole science of knowledge, in: *The Science of Knowledge*, Kroeger, A.E, trs., London; Trübner & Co,.Ludgate Hill, 67.)

of the proposition that "A exists for itself in the I general through the establishment of the I that judges," and this gives rise to the proposition "I am I." Further, "I am" is implied in "I am I." This "I am" is the "first definitive unfounded fundamental proposition" of all sciences.[29]

It may be noticed at a glance that this conditional proposition, which is inferred from the principle of identity (A=A), a fundamental principle of thought, as well as the proposition "I am," is originated from the famous proposition by Descartes, "I think, therefore I am."[30] In short, "I am" is Fichte's interpretation of Descartes's thesis. Descartes's statement features two different Is: the former "I" is the subject of "thought" and the latter "I" the object of thought. In other words, the "I" in "I think" is the subject of thought, and "I" in "I exist" is the "thought of I." Also, because these two sentences are linked by the connective "ergo," they are combined by the relationship of cause and effect. In other words, the self's ability to think is the cause of the self's existence. In yet another word, the former is a necessary condition for the latter. However, Fichte, who considered "I am" to be the fundamental proposition of academic theory, tried to reverse Descartes's argument. His argument can be summed up as "I am, therefore, I think." Seen in this light, "thought" is a special characteristic that I, who exists, has."[31] From this premise, "I am therefore I am" is derived, and since "The I in the former and the I in the latter significance are to be absolutely equal,"[32] the final conclusion must be "I am I."

AIH focuses on the reflective attitude of this nature. We have previously argued that linguistics, literature, history, philosophy, ethics, sociology, and cultural studies, anthropology, psychology are still viable in the era of AI that is represented by the term "posthumanism." If you apply the principle of self-reflection, the reflective character of humanities as a method for establishing an academic foundation may be said to be an exploration of the essence and function of the humanities. However, this is only a reflection on the existing fields of humanities research but is still insufficient in view of the research objective of establishing AIH with a new matter of "AI." Although Fichte's proposition "I am I" is based on the principle of thorough reflection, the range of reflection is still limited to one entity and is not commodious enough to

---

29 Ibid., 91; "the absolute, first, and unconditioned fundamental principle of human knowledge"(Fichte, G.W, (1889b). 63.)

30 See Ibid., 99; "we thus obtain as a fundamental principle of logic the proposition A=A, which can only be proven and determined through the science of knowledge. Proven: for A is A because the Ego, which has posited that A is the same as the Ego in which A is posited. Determined: for whatever is, is only in so far as it is posited in the Ego, and there is nothing outside of the Ego" (Fichte, G.W, (1889b). 72.)

31 See Ibid., 100; "Thinking is not the essence, but merely a particular determination of the Ego" (Fichte, G.W, (1889b). 73.)

32 Ibid., 96; "The Ego in the former and the Ego in the latter significance are to be absolutely equal"(Fichte, G.W., (1889b). 69.)

establish a new humanities department that extends beyond the range of I and learns about the present world. Therefore, let us take that attitude of reflection only as the principle of academic establishment which followed from Descartes, and then concentrate on the position of interpreting the Cartesian reflection from a different perspective.

Concerning Fichte's attention to the Cartesian reflection, Kant offers a similar yet different interpretation. Descartes' argument is incomplete in many ways, according to Kant. First, "I think therefore I am" is incomplete in form because it lacks such a major premise as "everything that thinks exists." Second, even if we deliberately supplement this premise, this I that derives its existence from its thinking is an isolated I who can only dwell on its thoughts, in Kant's view. Behind this idea lies Kant's unique interpretation of Descartes's thesis, "I think." According to Kant, Descartes' "I think" is an incomplete sentence, as in "I claim", from which the object is absent. It is possible for "I" to take the place of this object to form a sentence like "I think of I," but then the meaning would be no different from that of Fichte's subjective idealism. Kant also regarded the latter, as an incomplete sentence because the object of one's thinking must always be outside one's consciousness. In the place of the object, there must be an empirical I that perceives the world and not an empty reflection only of the thinking I itself without real content.[33] Now, the reflection extends to the living world we are facing here and now, not to consciousness alone. Applied to our study, the reflection on the relation between the world and the humanities that we experience is derived as a fundamental proposition for the groundwork of science. This opens the way for "AI," the first component of the "AIH," to enter into the scope of our discussion. The subject of AIH is the humanities, and the object experienced by the subject is artificial intelligence. Humanities, the subject of AIH, always performs self-reflection to clearly recognize its permanent existence, while trying to redefine its transforming shape through its relationship with AI, which is already related to the subject of its own experience.

4. The System of AIH

First, as the Humanities is a study of reflection when the point of reflection is directed inwardly or toward oneself, it constantly repeats its proper role. As such, we have argued that the innate power in the humanities, which can maintain its essential content while reconfiguring itself in keeping with the changing times, lies in its reflective ability. As mentioned previously, this backs up the claim that the contents

---

33 See Kim, Hyeongjoo (2016). *Zur Empirizität des "Ich denke" in Kants Kritik der reinen Vernunft*, Könighausen & Neumann, 136f.

of the traditional humanities are still valuable today. Second, if the scope of reflection is directed at the inner changes that occur when the humanities exposes itself to the world it confronts today, the road to "humanistic reflection on AI" will be opened. Based on the above, the process which AIH conforms to establish itself is as follows.

Stage 1: Contents of "AIH:" Artificial intelligence.

Stage 2: Contents of "humanities" - humanities as world-hermeneutics: Linguistics, literature, history, anthropology, psychology, ethics, sociology, and cultural studies. Method of "humanities" - humanities as world-hermeneutics: Self-establishment of linguistics, literature, history, anthropology, psychology, ethics, sociology, and cultural studies.

Stage 3: Method of AIH - Humanistic reflection on "artificial intelligence".

Stage 4: Contents of "AIH:" Linguistics, literature, history, philosophy (anthropology, psychology), ethics, sociology, and culturology with artificial intelligence as the object of reflection.

As seen above, AIH has been analyzed and managed based on the framework of content, and form, and object and method. It pulls off a reform by which it can cope with a recently developed scientific object, namely "artificial intelligence," while retaining as its own content such essential disciplines of the traditional humanities as linguistics, literature, history, philosophy, ethics, sociology, and culturology. The research scope of AIH reveals both aspects of self-preservation and self-renewal.

First, as AI Technology Criticism, we classify those studies of literature and history-oriented humanities that target AI. The unique humanistic imagination of literature is now steering the posthumanistic discourse. AI-themed sci-fi movies are also part of the work of advanced literature. It should be among the main tasks of contemporary literature to project the self-image created by us into the description of superhumans and non- or semi-humans, to present a new picture of society upon the emergence of those beings, and to depict the inner world of biological humans faced with them. The automaton (automated machine) Talos, featured in Greek mythology, Frankenstein, the epitome of "scientific myths," and many AI characters that appeared in the recently released films are all potential reality based on the literary imagination. In addition to the micro-history centered on the histories of persons, times, and regions, now an integrated study of history starts to raise certain serious questions concerning the status of the human species in nature, as exemplified by the Big History. Similar to historical speculation about prehistory, the future of AI also constitutes an axis of historical research. By providing this humanistic imagination with a ground for realistic reflection on the very nature of technology and a positivistic analysis of AI

technology, AI Technology Criticism strives to attach the weight of "realistic" to the "literature work and historical speculation as a likely story."

In this regard, much research must first be carried out first on the characteristics of technological development according to human nature in the age of AI and on the change patterns in human nature that might be brought about by AI technology. To this end, while performing an analysis of the history of AI development and attempting to disclose the essence of human intelligence that cannot be replaced by machines, the possibility of AI technology at a cultural aspect that extends beyond the frame of industrial production should be examined. And then a due inquiry into the possibilities of a super-intelligence era and technological research for the preservation of human dignity[34] must be made if the era does arrive.

Second, a humanistic study centered on linguistics that targets AI is termed as AI Relationships & Communication Studies. It also investigates the reestablishment of relationships among all entities in the world and mode change in communication among them, which are highly likely to occur owing to the emergence and development of AI. As described in the previous chapter, the predominance of Descartes' Cogito determines the direction of our understanding of human nature, as well as the metaphysical status of humans. If its influence grows, the status of humans in nature ascends, and the existential characteristic of humans defined by reason, and consciousness becomes more determinate. In the opposite case, the distinction between humans and non-humans becomes faint, and reason and consciousness will be considered no more as the essence of human nature. While the Cartesian Cogito forms the basis of foundationalist humanism, posthumanism views it as the main culprit needed to be dismantled.[35]

AI Relationships and Communication Studies diagnose first the possibility of a universal grammar relying on foundationalism, the ultimate aporia of linguistics. Such an attempt can be made possible because of the development of AI technology aggregating a huge amount of data. We make use of these data to ascertain the theoretical possibility of constructing a universal grammar and language typology. Second, it searches for an appropriate communication model for each channel, noting the relationships resulting from the development of AI technology and the

---

34   Discussion about human dignity in the age of artificial intelligence expands its area beyond the end-means relationship among humans to the relationship between artificial intelligence and humans. We regard the preservation of self-objectivity as an important criterion for securing human dignity in the age of artificial intelligence, as shown in the second form of Kant's categorical imperative. "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." See Kant, I. (1900), 429.

35   For the discussion on the relationship between the deductive system of universal grammar and consciousness, see Chomsky, Noam (2009). *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*, Cambridge; Cambridge University Press.

diversification of communication channels. The emergence of chatbots, for example, has rendered it no longer possible to dismiss computer messengers as "messengers" just for human-to-human communication. If we recognize the companion robot or chatbot as a communication player, the geography comprised of the innumerable channels of communication and relationships will change dramatically. The channels of communication will be multiplied from human-to-human communication to human-to-machine. Furthermore, the communication models of humans who are accustomed to communicating with machines will also be an important subject. For example, to examine the emotional communication between humans and AI, research has been conducted on the output sensitivity of AI and the patterns of human interaction; and to explore the communication between AI and human, studies will be conducted on the status and future of our language acquisition and automatic translation using the deep learning method, as well as the possibility of communication among AI's themselves. This is accompanied by a genealogical study on the changes in the identity of communication subject according to the changes of the communication method.

The third research domain of AIH is AI Socio-Cultural Studies. The discipline of AI Socio-Cultural Studies analyzes social and cultural phenomena that arise due to the emergence of AI and reviews its effects and issues related to the current social culture. As AI advances, whereas the focus of AI Technology Criticism is placed on the changes in the personality of the individual subject, AI Socio-Cultural Studies attempt to form a discourse from a macroscopic perspective regarding the patterns of change in individual cultures, social structures, groups, and systems, in addition to maintaining an open forum of discussion on comparative technology. To this end, first, relying on the research on the differences between the changes in segmented social sectors prompted by AI and those changes brought in by the previous industrial revolution, we seek to find new research methods that can address the issues of society and culture following the universalization and commercialization of AI technologies. If the emergence and commercialization of sex robots can reveal changes in the internal and behavioral patterns of individuals as sexual actors through Technology Criticism, Socio-Cultural Studies predict the resulting changes in the relationship between men and women, family relationships and concepts, and further changes and cultural conflicts between those who can own and produce the technology capital and those who cannot. Accordingly, we then accumulate basic data on the human response to the negative consequences of changes in society and culture in the age of AI and examine some methods for preserving dignity. This study leads to the fourth area of research, which is AI Ethics.

AI Ethics, in conjunction with the results presented by the above research, ultimately aims to provide practical norms suitable for the age of AI. First, in order to achieve this goal, we make a clear distinction between the philosophical concept of AI and AI robot or machine, and between various layers and categories contained in the concept of AI. As the AI industry develops, the frequency of using the word "artificial intelligence" is rapidly increasing. For example, the products we use in everyday life were simple home appliances many years ago, but recently they have been advertised as using AI technology; as the number of novels, movies, and dramas based on AI has grown exponentially, we no longer view such materials as new. This raises questions concerning the substance of AI and the conceptual prescription of its category. AI Ethics that falls into a broader scope of philosophical research carry out these conceptual studies first. This leads to a study on the rules governing the existential status of AI beings. The blending of such concepts as The Strong AI, Super AI, and General AI (Generative AI), and the discussion on the reality of the advent of technological Singularity may seem to be an analytical exploration of related concepts. However, from a broader perspective, they lead to a question on the very nature of AI, and further the nature of intelligence itself. In the posthuman era, with the aim of building AIH to elevate the value of the humanities, we establish ethical theories on AI to confine its use to a single tool. If the study on the definition of AI's existential status is based on the philosophy of psychology and metaphysical research relevant to, for example, identifying the essentials of "consciousness," the establishment of ethical theory focuses on AI treatment issues based on this establishment of its existential status. The former is concerned with developing explainable AI and the latter with solving the problem of the ethical vacuum caused by AI. As a result, AI ethics relates to the creation of practical and specific norms pertaining to AI governance. We will participate in the formation of sound discourse to establish a social safety net by presenting ethical guidelines for producers, consumers, and sellers for the use of each of the AI robots produced, including those specialized in health care, combat, and for treating those robots as pets, along with the creation of norms at a pan-ethics level, such as the Robot Ethics Charter.

Finally, if we reapply the frame of "Contents and Methods" to "AIH," in addition to the humanities described earlier, which takes AI as its content, the humanities that uses AI technology as a method will fall into that category. If the reflection as methodology takes "AI"(technology) for its object, the humanities that uses AI technology as its research method, that is, the digital humanities which contains humanistic content in a bowl of AI technology will be established under the

name of AI Data Hermeneutics.[36] First, AI Data Hermeneutics will establish "humanity" big data, which utilizes the achievements of the digital humanities that are based on the computerization, analysis, and visualization of humanities data. As a first step, it attempts to build an emotional ontology that creates data and implements the patterns of human emotional expression reproduced in various humanities contents in a three-dimensional manner. While AI Ethics radiates the achievements of AIH as an integrated achievement of the AI norms, AI Data Hermeneutics provides a methodical basis for the systematic data set of humanistic content that is accessible in every research area. Second, the subject of research is to analyze the "modus of representation of AI in literature and art works" using specific digital technology and to develop a modern interpretation of humanities classics. Conventional humanistic research has consisted of the transitions between character data through such methodologies as "intuition, experience, expression, understanding and interpretation." However, the results of AI Data Hermeneutics using such methodologies as computerization, data mining, and data visualization transcend the range of character data. In short, AI Data Hermeneutics consists in the modern interpretation of the classical literature featuring AI (which is humanistic data in a broad sense), the modern interpretation of humanities classics by using the methodology of digital humanities, and the methodological foundation for the aggregation of human data using digital technology to rediscover human nature in the age of AI.

To recapitulate, the contents of AIH are composed of AI Technology Criticism, AI Relationships & Communication Studies, AI Socio-Cultural Studies, and AI Ethics. Furthermore, its methodology comprises AI Data Hermeneutics. The four research areas comprising the contents of AIH have an organic structure. Tracking AI technology development by AI Technology Criticism and adjusting individual subjects' personality and behavior patterns due to technological development will constitute the cornerstone for the other three research areas. In addition, while AI Relationships & Communication Studies involve research on AI technology and personality of subjects confined to language and communication, AI Socio-Cultural studies, paying attention to the changing patterns of collective culture in society, set foot in the foundation of AI Technology Criticism but form a counterpoint with it in research. AI Ethics collects the achievements from the above three areas of research to establish a philosophy that elevates human dignity in the age of AI. In its role as

---

36   Many Humanists consider the Digital Humanities as a tool that engages with the traditional work of the humanities as a type of methodology. (See Gardiner, E. and Musto, R. (2015). *The Digital Humanities*, Cambridge; Cambridge University Press, 3-4) As such, the content of the digital humanities comprises the subject of the traditional humanities, and the "methodology" is the Humanities computing equipped with digital tools. In this sense, we understand that artificial intelligence humanities data analysis, defined as the methodology of the study of humanities, implies the digital humanities.

a methodology for AIH, AI Data Hermeneutics uses digital technology to build its own academic system, simultaneously leading the research in the above four areas. The relationship between the five areas of study described above is illustrated as follows.
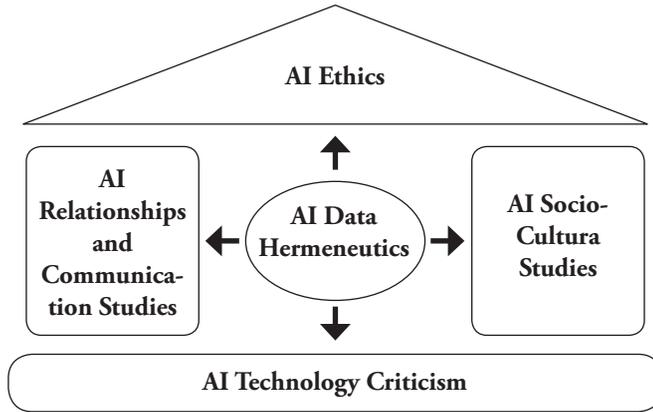

Fig. 1. the System of AIH

5. Conclusion

We employ the methodology of separating content and form, a methodology unique to the traditional humanities, in order to establish the academic identity of a new discipline, i.e., AIH. As such, we extracted a few disciplines from the traditional humanities by specifying the distinction between the academic attitude of posthumanism and AIH, which leads to the AI-related discourse. From this, we deduced five research areas of AIH and described the organic structure of these five areas. Through this, we reorganized the traditional humanities disciplines into five research areas and named them "AIH" to establish a multidisciplinary and convergent research model focused on AI. The ultimate goal is to determine the function humans will perform in the age of AI and how human values should be defined. Amid the prospect that AI will replace most human thought and activity, some changes in the basic human identity have held for thousands of years, "human as a producer" and "working man," have already emerged, albeit to a limited degree. We want to continue our research in this period of change, in which we practically and persuasively determine the functions humans will perform and the actions they should take, as well as how the process will proceed and maintain its desirable qualities.

References

Aristotle (1954), *The Works of Aristotle v. VIII Metaphisica(trans. form Ross)*, the Clarendon Press, Oxford, 980ᵃ-981ᵇ, 1027ᵇ-1029ᵇ.

Badmington, Neil. (2001), Pod Almighty!; Or, Humanism, Posthumanism, and Strange Case of Invasion of the Body Snatchers, *Textual Practice,* 15 (1), 5-22.

Bostrom, Nick (2003), The transhumanist FAQ, in: Bostrom, N, *Nick Bostrom's Home Page*, https://www.nickbostrom.com/views/transhumanist.pdf (accessed: 3 July 2019)

Bostrom, Nick (2005), A history of transhumanist thought, Journal of Evolution & Technology, 14 (1), 4, https://nickbostrom.com/papers/history.pdf

Brands, Hartmut (1982), *Cogito ergo sum*, Freiburg; Karl Alber. 10.

Chomsky, Noam (2009), *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*, Cambridge; Cambridge University Press.

Conze, Werner and Koselleck, Reinhart eds. (1982), Menschheit, Humanität, Humanismus, *Geschichtliche Grundbegriffe: Historischers Lexikon zur politisch-sozialen Sprache in Deutschland,* Bd. 3, Stuttgart; Klett-Cotta, 1064-66.

Diemer, Alwin, Historische Wörterbuch der Philosophie online, https://www.schwabeonline.ch/schwabe-xaveropp/elibrary/start.xav#__elibrary__%2F%2F*%5B%40attr_id%3D%27verw.moral.moralisch.moralphilosophie%27%5D__1585115242588 (accessed: 25 March 2019)

Fichte, Gottlieb W. (1889a), Introduction: Concerning the conception of the science of knowledge generally, in: *The Science of Knowledge*, Kroeger, Adolph E. trs., London; Trübner & Co,.Ludgate Hill, 46-58

Fichte, Gottlieb W., (1889b), Fundamental principles of the whole science of knowledge, in: *The Science of Knowledge*, Kroeger, Adolph E, trs., London; Trübner & Co,.Ludgate Hill, 63-73

Fichte, Gottlieb W. (1971a), *UBW in: Fichtes Werk I*, Berlin; Walter du Gruyter. 38-46.

Fichte, Gottlieb W. (1971b), *GLW in: Fichtes Werk I*, Berlin: Walter du Gruyter, 91-99.

Fukuyama, Francis (2004), Transhumanism, *Foreign Policy,* 144, 42-3

Gardiner, Eileen and Musto, Ronald. (2015), *The Digital Humanities*, Cambridge; Cambridge University Press, 3-4

Grafton, Anthony and Jardine, Lisa (1986), *From Humanism to the Humanities: Education and the Liberal Arts in Fifteenth- and Sixteenth Century Europe*, Boston; Harvard University Press.

Haraway, Donna (1995), *in Wolfe*, 36. Quoted from Posthumanism, 190.

Heidegger, Martin (2000), *Über den Humanismus*, Frankfurt; Klostermann. 8.

Herbrechter, Stefan (2012), *Posthumanism A Critical Analysis*, London; Bloomsbury, vii-31.

Hiorth, Finngeir (1996), *Humanismus – Genau Betrachtet*, Neu-Isenburg; Angelika Lenz Verlag. 57.

Jaeger, Werner (1973), *PAIDEIA*, Berlin; De Gruyter.

Kant, Immanuel (1900), *Grundlegung zur Metaphysik der Sitten(AA4)* in: *Kants gesammelte Schriften (Sog. Akademie-Ausgabe)*, Berlin; Walter de Gruyter. 387-390.

Kimball, Bruce A. (1986), *Orators & Philosophers: A History of the Idea of Liberal Education*, New York; Teachers College Press. 78.

Kim, Hyeongjoo (2016), *Zur Empirizität des "Ich denke" in Kants Kritik der reinen Vernunft*, Würzburg; Könighausen & Neumann. 136f.

Kim, Hyeongjoo and Lee, Chankyu (2019), Beyond the Posthumanism – Try to Laying a Conceptual Foundation Artificial Intelligence Humanities, *Philosophical Investigation*, 53 (3), 51-80

Nietzsche, Friedrich (1906), Zur Genealogie der Moral in: Nietzsche's Werke B. VIII, Leipzig; C. G. Naumann Verlag, 476

Sorgner, Stefan (2009), Nietzsche, the Overhuman, and Transhumanism, *Journal of Evolution & Technology*, 20 (1), 29-42.

Tertullianus. (2001), De Praescriptione Haereticorum,

http://www.tertullian.org/works/ de_praescriptione_haereticorum.htm, (accessed: 7 August 2019)

Wundt, Mohr (1945), *die deutsche Schulphilosophie im Zeitalter der Aufklärung*, Tübingen; Olms, 1

Temelj humanističke umjetne inteligencije

SAŽETAK

Cilj humanističke umjetne inteligencije je predstaviti humanističke prijedloge za naše živote i život akademske zajednice u vezi s tim kako će ljudi živjeti u pametnom gradu, okruženi tehnologijom umjetne inteligencije. U tu svrhu predstavljamo temelje za humanističku umjetnu inteligenciju i dijelimo s tim povezane zabrinutosti. Kao koncept, „humanistička umjetna inteligencija" sadrži tri elementa: „umjetnu inteligenciju", „sadržaj humanističkih znanosti" i „metodologiju humanističkih znanosti". „Sadržaj humanističkih znanosti" obuhvaća derivate tradicionalnih humanističkih znanosti, naime, lingvistiku, književnost, povijest, etiku, sociologiju i kulturalne studije. Ovih pet područja istraživanja, zajedno s disciplinom stvorenom tradicionalnim prihvaćanjem promjena u humanističkim znanostima koje donosi doba inteligencije, proizvod su razmišljanja o sebi i svijetu, a primjenjuju se na svijet kakav danas postoji u doba umjetne inteligencije i humanističkih znanosti.

Ključne riječi: humanistička umjetna inteligencija, sadržaj humanističkih znanosti, metodologija humanističkih znanosti, posthumanizam.