

國立臺灣大學文學院哲學系



碩士論文

Department of Philosophy

College of Liberal Arts

National Taiwan University

Master's Thesis

人工智慧與人類的未來

Artificial Intelligence and the Future of Humanity

廖崧年

Song-Nian Liao

指導教授：文哲博士

Advisor: Prof. Dr. Christian Helmut Wenzel

中華民國 114 年 6 月

June 2025

摘要



本文基於尤瓦爾·諾亞·赫拉利在其著作《人類大歷史》中關於人工智慧的論點。赫拉利論點的獨特之處在於，他認為即使是弱人工智慧也足以取代人類，而大多數人則認為需要強人工智慧或通用人工智慧才能取代人類。我將提出三個可能反駁赫拉利的論點，但最終都將失敗。接下來，我將把討論擴展到通用人工智慧。布萊恩·坎特韋爾·史密斯、約布斯特·蘭德格雷貝和巴里·史密斯對此持反對意見。他們的論點適用於強人工智慧，但在弱人工智慧方面則不然。最後，我將基於黑格爾關於有限性和無限性的哲學提出一個論點，該論點可以駁斥赫拉利的弱人工智慧足以取代人類的觀點。

赫拉利認為，即使是弱人工智慧也足以分析人類行為，並且在擁有足夠的數據後便能更準確地預測人類的行為。數據越多，預測就越準確；隨著預測變得越來越準確，人們就會越來越少地自己做出決定，而是選擇遵循人工智慧做出的預測。一旦你把所有的決策都交給人工智慧，人類的價值就所剩無幾了。人工智慧如何在沒有首先獲得理解和意識的情況下分析並預測人類的行為？赫拉利認為，答案不在於電腦科學，而在於生物學。生物學家發現，人類深受基因的影響，並通過對不同因素的複雜計算做出決定。這意味著人類可以被視為通過一系列計算來回應世界的演算法。如果人類是演算法，那麼人工智慧只需要分析人類如何作為演算法進行這些計算，就能預測他們的下一步行動。人工智慧對人類的預測不必完全準確，如果它們的準確性高於人類自身的判斷，那麼人類就應該合乎邏輯地遵循人工智慧的預測，而不是自己的判斷。理解和意識不是必需的，因為人工智慧所做的只是分析人類過去的行為，並用它來預測他們

未來的行為。



第一個論點是賽爾著名的中文房間論證。第二個論點來自托馬斯·福克斯，他在其論文《自戀抑鬱的科技》中也批評了赫拉利的論點。第三個論點要麼是否認人類必須遵循人工智慧的預測，要麼是規範人工智慧的發展。賽爾的論證可以分為幾個部分。想像一下你身處一個充滿符號和符號操作規則的房間裡。一組符號進入房間，你按照給定的規則操作房間裡的符號來做出回應。你不知道的是，進入房間的那組符號是一個中文問題，而你做出的回應是一個完美的中文答案。在這種情況下，你理解中文嗎？賽爾的答案是否定的，這表明理解不能僅僅通過符號的操作獲得，而這正是人工智慧所做的。托馬斯·福克斯反對赫拉利關於人類可以被簡化為僅僅是演算法的說法。他認為，人類是具有生物身體和感受能力的生命體，而人工智慧缺乏這兩者。為了抵制人工智慧的誘惑，人類必須擁抱他們有限的生物身體並對此感到滿足。第三個論點很簡單，人類可以簡單地選擇不遵循人工智慧的預測，或者我們可以像馬蒂亞斯·里塞建議的那樣規範人工智慧的發展。即使人工智慧為我們提供了最佳解決方案，我們人類也可能更喜歡不太理想的選擇。例如，人工智慧可能會根據一個人的偏好以及期待觀看這部電影幾個月的事實，預測他會觀看某部動作片。然而，我們仍然可以在最後一刻改變主意，簡單地選擇不觀看這部電影。

這三個論點將被證明不足以反駁赫拉利的論點。因此，要解決弱人工智慧通過準確預測取代人類的問題，就必須找到隱藏在人類是演算法的前提中可以推出人類不只是演算法的特徵。這樣的特徵可以在黑格爾關於有限性和無限性的哲

學中找到。根據黑格爾的觀點，人類是有限的存在，但他們也超越了自己的有限性，達到了無限。無論弱人工智慧變得多麼先進，它都無法超越自身的有限性，達到無限。

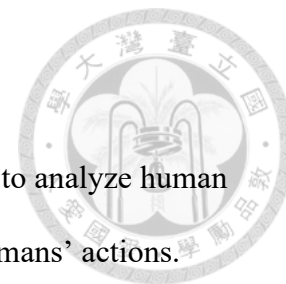


總之，尤瓦爾·諾亞·赫拉利提出了一個有趣的論點，即弱人工智慧足以取代人類。針對強人工智慧的論點在這種情況下是不充分的，因為赫拉利並沒有聲稱人工智慧可以獲得理解或意識。解決弱人工智慧問題的方案是找到隱藏在人類是演算法的前提中足以證明人類不只是演算法的特徵。這個隱藏的特徵是人類超越個體有限性的能力。

Abstract

This thesis is based on Yuval Noah Harari's arguments on ai in his book *Homo Deus: A Brief History of Tomorrow*. First, there is a difference between weak AI and strong AI. Weak AI specializes in a certain field of work, while strong AI can operate in multiple fields with proficiency at least equal to humans. What makes Harari's argument different is that he argues that weak AI is enough to replace humans when most argue that strong AI or artificial general intelligence that usually gains understanding or consciousness, is required to replace humans. I will present three possible counter arguments to Harari that ultimately fail. Then, I will extend the discussion to artificial general intelligence. Brian Cantwell Smith, Jobst Landgrebe and Barry Smith opposes the likelihood of artificial general intelligence. Their arguments are great for strong AI but not so much for weak AI. Finally, I will present an argument based on Hegel's philosophy on finitude and infinity which can refute Harari's argument that weak ai is enough to replace humans. So, the argument based

on Hegel's philosophy will be my answer to Harari's argument.



Harari believes that even weak artificial intelligence will be able to analyze human behavior, and with sufficient data, make accurate predictions on humans' actions.

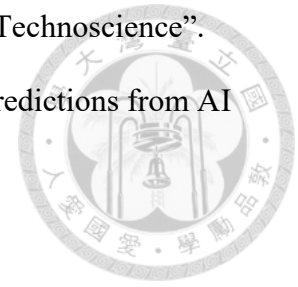
With more data the more accurate the predictions, and as the predictions become more accurate the less people would make decisions for themselves, instead opting to follow the predictions AI has made. Once you leave all the decision making in the hands of AI, there is little to no value left for humanity. How is it possible for artificial intelligence to analyze and thus predict humans' actions without gaining understanding and consciousness first? Harari argues that the answer lies not in Computer science but in biology. Biologists find out that humans are deeply affected by their genes and make decisions through a complicated calculation of different factors. This means that humans can be seen as algorithms that respond to the world through a series of calculations.

If humans are algorithms, then artificial intelligence only needs to analyze how humans make these calculations as algorithms to predict their next move. Artificial intelligence's predictions on humans don't have to be perfectly accurate, as long as their accuracy is better than humans' own judgement, humans should logically follow the prediction of AI over their own. Understanding and consciousness isn't required since all the AI is doing is analyzing what humans did before and using it to predict what they will do in the future.

The first of the three possible counter arguments against Harari is Searle's famous Chinese room argument. The second argument is from Thomas Fuchs, who also

criticizes Harari's arguments in his paper "Narcissistic Depressive Technoscience".

The third argument is to either deny that humans must follow the predictions from AI or regulate the development of AI.



Searle's argument can be broken down into several parts. Imagine yourself in a room filled with symbols and rules on how to manipulate these symbols. A set of symbols come into the room, and you respond by manipulating the symbols in the room following the given rules. Unknown to you, the set of symbols coming in is a question in Chinese, and the response you made is an answer in perfect Chinese. Does one understand Chinese after this? Searle's answer is no, which shows that understanding cannot be gained through only manipulation of symbols which is what ai does.

Thomas Fuchs rejects Harari's claim that humans can be reduced to mere algorithms. He argues that humans are living beings with a biological body and the ability to feel, both of which artificial intelligence lacks. To reject the temptation of AI, humans must embrace their limited biological body and be content with it. The third argument is simple, humans can simply choose not to follow the predictions from AI, or we could regulate AI like Mathias Risse suggested. Even if AI presents us with the best solution, we humans may prefer less optimal options. For example, AI could make a prediction that one will watch a certain action film based on one's preferences and the fact that one is looking forward to watching this movie for several months. However, we could still change our minds at the last second and simply choose not to watch this movie.

The three arguments will be proved insufficient against Harari's argument. So, to solve the problem of weak AI replacing humans through accurate predictions one must find what is hidden in the assumption that humans are algorithms that leads to the conclusion that humans are not merely algorithms afterall. This means that even if Harari's claim that artificial intelligence could analyze humans as algorithms are assumed, there is more to humans as algorithms. Such features can be found in Hegel's philosophy on finitude and infinity. According to Hegel, humans are finite and limited beings, but they also transcend their finitude and reach into infinity. Weak ai no matter how advanced it becomes cannot transcend its finitude and reach infinity.

In conclusion, Yuval Noah Harari brings up an interesting argument that weak ai is enough to replace humans. Arguments that are directed towards strong ai are insufficient in this scenario, since Harari isn't claiming that ai can gain understanding of consciousness. The solution to the problem of weak AI is to find a hidden feature in the assumption that humans are algorithms that proves humans are not merely algorithms afterall. This hidden feature is humans' ability to transcend the finitude of the individual.

Content

摘要.....	i
Abstract.....	iii
Content.....	vii
Introduction.....	1
Literature Review.....	9
Research Motivation and Purpose.....	21
Argument.....	23
Conclusion.....	44
Citation and References.....	50



Introduction

In this thesis, I will critically examine the arguments presented in Yuval Noah Harari's thought-provoking work, "Homo Deus." Harari raises significant concerns about the trajectory of humanity in the face of advancing technology, particularly artificial intelligence (AI). While his assertions are compelling, there exist several counterarguments that warrant consideration.

One prominent critique comes from John Searle's Chinese Room argument, which is frequently invoked to challenge the notion of thinking AI. However, this argument is insufficient in the context of Harari's claims, as it presupposes that AI must possess understanding. Harari's arguments do not hinge on the premise of AI achieving true comprehension; rather, they underscore the implications of AI's capabilities on human society.

Additionally, I will explore Thomas Fuchs' proposal of an embodied humanism, which directly contests Harari's perspectives. Nevertheless, I intend to demonstrate that, in practical applications, Fuchs' argument ultimately falls short. Another line of reasoning, as articulated by Mathias Risse in his book, "Political Theory of the Digital Age: Where Artificial Intelligence Might Take Us," advocates for the rejection or limitation of AI. However, this approach fails to account for the pervasive influence of AI on humanity as a collective entity, rather than merely focusing on individual experiences.

In the subsequent section of the thesis, I will delve into criticisms of artificial general intelligence that are pertinent to Harari's arguments, yet ultimately prove inadequate

to dismantle his conclusions. This does not imply that Harari's arguments are beyond reproach; in fact, I identify a significant flaw that undermines his entire thesis. Consequently, I propose an alternative perspective rooted in Hegelian philosophy concerning finitude and infinity. Humans will not lose their intrinsic meaning to artificial intelligence and algorithms precisely because of their finite nature, which inherently constrains the AI they create. In contrast, humanity continually transcends these limitations, with successive generations striving to realize their fullest potential and surpassing their wildest dreams.

The discourse surrounding artificial intelligence has gained unprecedented momentum since the emergence of technologies like AlphaGo and ChatGPT. Harari's arguments in "Homo Deus" contribute to this dialogue by categorizing AI into two distinct realms: specialized artificial intelligence, designed for specific tasks, and artificial general intelligence, capable of performing any intellectual endeavor akin to a human. Currently, the landscape of AI is dominated by specialized systems, yet technology companies worldwide are ambitiously pursuing the development of general artificial intelligence. This pursuit raises widespread fears of a future where general AI could entirely supplant human roles. However, Harari provocatively suggests that the advent of specialized AI is already sufficient to displace human contributions in various sectors.

In conclusion, while Harari's commentary on the future of humanity in the age of AI is undeniably thought-provoking, it invites a deeper examination of the assumptions that underpin his arguments. As we navigate this complex terrain, it is imperative to critically engage with both the promises and perils posed by AI, recognizing the

unique qualities of humanity that technology cannot replicate.

In Harari's "Homo Deus", he tells the story of Homo sapiens. How they found meaning in religion, signing a deal to trade meaning for power, and regained meaning by subscribing to humanism. Science combined with capitalism gives us the power to reign in famine, plague and war, while humanism instructs humans to create meaning for this world by themselves. Yet, the rise of artificial intelligence combined with new discoveries in biology poses an existential threat to humanism. Various research points out that humans are nothing but biological algorithms. These researches include experiments that use brain scanners to predict people's desires and decisions well before they are aware of them. Transcranial stimulators experimented by the American military to enhance performance of soldiers¹. Studies on split-brain patients conducted by Professor Roger Wolcott Sperry and his student Professor Michael S. Gazzaniga². Experiments conducted by Daniel Kahneman show that people average the peak and end points of their experience without taking account the duration of their experiences³. An algorithm, no matter how complicated, can be analyzed. Once artificial intelligence fully analyzes humans as algorithms, it will be able to understand ourselves better than we ever could and make the best decisions for us. In other words, it will take over the inner voice in ourselves which, according to humanism, creates meaning for this world.

Harari's arguments begin with the rise of what he calls humanism. Prior to the advent of humanism, religion is what gives humans meaning to their lives. In different

¹ Adey, Sally, "Zap Your Brain into the Zone: Fast track to Pure Focus", *New Scientist*, 2850, 2012.

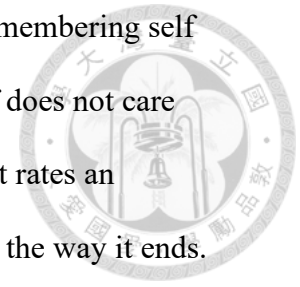
² Gazzaniga, Michael, *Who's in Charge? Free Will and the Science of the Brain*, New York: Ecco, 2011; Gazzaniga, Michael, *The Bisected Brain*, New York: Appleton-Century-Crofts, 1970.

³ Kahneman, Daniel, *Thinking, Fast and Slow*, London: Penguin Books, 2011.

gods the masses found peace, yet the development of science poses a great challenge to traditional religions. The theory of evolution in particular questions religions' claim on how humans come to be. Thus, gods lost their power over humans and all there are left who hold power are humans themselves. We come to believe that we have free will and could choose to do whatever we like, there are no more gods and popes to tell us what to do, instead we do the one thing that we ourselves want to do. Harari describes this as listening to your inner voice which tells you what you want to do. As it turns out, biology is a threat to free will and our inner voice as well. No matter how hard biologists try, they could not find the existence of free will just like they couldn't find the existence of souls. What they can do is explain to their best what is going on inside humans' brains. They see humans as biological algorithms that have evolved for tens of thousands of years to survive. For ancient humans that constantly faced the decision to avoid danger or go for a risky prize, the only ones that survive are the ones that maintain the balance between being too rash and being too cowardly which pass their genes onto us. So, when biologists must explain why certain people, for example, make risky investments, they prefer to use genes and biological mechanisms to explain the behavior rather than free will and inner voice. Speaking of inner voice, the biologist also finds the workings of the human algorithm to be messy and crude.

In a study on colonoscopy patients conducted by Daniel Kahneman and Donald Redelmeier, doctors want to know whether the patients prefer a shorter but more painful procedure or a long procedure that is less painful. The patients are to report their pain level at one-minute intervals during the procedure and their overall pain level of the entire procedure. It turns out that patients average the peak and end point of their experience as the overall pain level with no regards to the duration of the

procedure. There is an experiencing self that feels the pain and a remembering self that tells the story of this painful experience. The remembering self does not care about the duration of a pleasant or unpleasant experience. Instead, it rates an experience by the maximum or minimum of the experience, and by the way it ends.



When the memory of this experience is recalled, only the worst moment and the last moment is averaged and the duration is ignored. “Odd as it seems, I am my remembering self, and the experiencing self, who does my living, is like a stranger to me.”(Daniel Kahneman, *Thinking, Fast and Slow*, Macmillan, 2011) This mechanism explains why doctors give children sweets after an uncomfortable appointment; a moment of happiness is enough to ameliorate the long painful experience of the appointment.

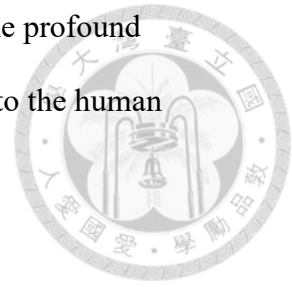
But what does biology have to do with artificial intelligence? If humans are indeed biological algorithms that can be explained scientifically but a total confusing mess when we must explore on our own. Then the simple solution is to let science explain how the human algorithm works instead of figuring it out on our own. To this end specialized artificial intelligence are the perfect candidates for the job. With the power of big data and deep learning techniques, artificial intelligence can analyze humans as algorithms and give us advice on what to do. At first their advice would be terrible, but as they receive more data and train better, eventually artificial intelligence will give us better advice than when we are on our own. At this point humans follow the advice of artificial intelligence more than from other people or from themselves. Harari believes that this will spell the end of humanism and the value it places on humans.

Harari's arguments present several weak points that merit a critical examination, particularly the foundational assumption that humans can be reduced to mere algorithms. While it is easy to counter this claim by outright rejecting the notion that humans are algorithms, such rebuttals may not entirely dismantle Harari's position. Proponents of the algorithmic view of humanity may dismiss these counterarguments, especially as the field of artificial intelligence continues to evolve at a breakneck pace. This reality leads us to a pressing concern: if we are to combat the reductionist perspective that seeks to analyze and manipulate humans through AI, a more comprehensive refutation of Harari's claims is indispensable.

Imagine, for a moment, that we accept the premise that humans are indeed biological algorithms. If, within this framework, we can still arrive at the conclusion that humans transcend mere algorithmic function, we may effectively persuade even the most ardent supporters of the algorithmic thesis. Thus, I propose to explore the implications of this assumption in a manner that reveals the inherent limitations of artificial intelligence. By engaging with Hegel's philosophy of logic, I will illustrate that humans are never simply "something" but are always "something more." The advancement of AI, then, should not be viewed as a threat to humanism; instead, it could serve as a powerful tool to broaden our understanding of humanity. Far from replacing our inner voice, AI could enhance it, making it clearer and more resonant.

Neither Harari nor the biologists are incorrect in recognizing that humans exhibit characteristics akin to biological algorithms. The critical flaw in their reasoning lies in their insistence on viewing humans solely through this lens. They operate under the belief that specialized artificial intelligence can fully decipher the complexities of

human existence. However, this perspective fails to acknowledge the profound challenges posed by paradoxes and contradictions that are inherent to the human experience.



It is common knowledge that contradictory programming leads to system failures; even the most advanced AI employing deep learning techniques struggles to navigate the intricate web of human contradictions. To truly understand how a human philosopher confronts contradictions, we must turn to Hegel, who meticulously addresses this issue in his seminal work, "The Science of Logic." Hegel's insights will not only illuminate the limitations of algorithmic perspectives but will also pave the way for a richer dialogue about the potential of AI as a complementary tool in the exploration of human identity.

In Stephen Houlgate's "*The Opening of Hegel's Logic*", he delivers a deep interpretation of Hegel's "*Science of Logic*". From the abstract indeterminate being to true infinity, this journey on unveiling what being really is could provide us insights on why artificial intelligence will not replace humanity but rather empower it. For Hegel, the truth of being is that it is limited and finite, bringing itself to its own end. Just like humanity portrayed by Harari, limited as biological algorithms and doomed themselves through bringing this fact to its logical end, artificial intelligence. Yet being is not merely limited and finite but ought to transcend its finitude towards infinity. As we shall see in this paper, so too will humanity rise above their destined end, ascension not through the hands of technology but by their own self. Hegel begins with the simplest idea of indeterminate being, this being is empty and immediate. "Such being is so indeterminate, indeed, that it has no determination or

content of its own that would set it in explicit contrast to anything else. It is the least we can understand thought to be and the least we can understand there to be at all.”

(Stephen Houlgate, *The Opening of Hegel's Logic*, page 263-264) Precisely because of pure being's nothingness that we get the idea of nothing. Yet to speak of the idea of nothing is to speak of a thing, a being. Thus, nothing transforms back into being. The transformation of being and nothing into each other is becoming. But since being is not nothing, a negation of a negation, we get a new idea, the determinate being. The determinate being turns into something when compared to the other that is not this determinate being. Yet something isn't just not the other, it has being of its own. This being of its own however contains the very negation of itself, for only through the limit that makes the being not itself can the being be. The negation that lies in something's own being leads something to its end which makes the something finite. A finite being comes to its end on its own, but the result of its demise is not simply nothingness. As we have already established nothing turns into being, so a finite being turns into another finite being and another finite being ad infinitum. In this infinite cycle we have a chance to reach true infinity. Compared to true infinite the bad infinite holds on to the idea that the finite and the infinite are complete opposites. When the bad infinite refuses to let go of the idea that it is not finite, it in turn is limited by the finite for only through the limit between finite and infinite can the bad infinite exist, which makes it the finite infinite. To grasp the idea of true infinite one must let go the opposition of finite and infinite, for the cycle of finite being turning into another finite being shows that the finite being is in fact returning to itself. When the finite being meets its end and turns into another finite being, the other finite being is in truth the same as the finite being that met its end for the other finite being will also eventually meet its end. Thus, we get a process of being returning to itself with

the finite and the infinite as its moments. This process is what Hegel calls the true infinite.

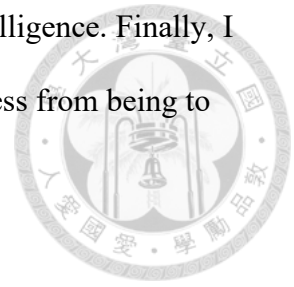


Now, we have seen how Hegel deals with the contradiction between the finite and the infinite. It seems like even deep learning artificial intelligence will have problems replicating this process. The advantage of deep learning artificial intelligence over traditional artificial intelligence is that it is more flexible, utilizing a network that responds to rewards and punishments, it can adapt and overcome by strengthening the parts that are working and weakening the parts that are less suited for the current task. But then how will it be able to let go of the opposition between the finite and the infinite. If artificial intelligence follows Hegel's arguments, the idea of finite and infinite will occupy its entire network, and it will become the embodiment of Hegels' arguments. No rewards or punishments will be enough to trigger artificial intelligence to essentially throw itself out. This kind of decision is left to humans alone to make. Only humans are capable of such drastic changes, and this could be the meaning of being free, not free to do whatever we want, but free to choose to let go.

Literature Review

In the next few paragraphs, I will first take a closer examination at Harari's "Homo Deus" and then discuss the three potential arguments against Harari. For the next part of this thesis, I will extend the discussion to the viability of artificial general intelligence. There are many optimists that think artificial general intelligence is likely to happen, like Nick Bostrom or Ray Kurzweil. However, I would focus on several

opponents who argue against the likelihood of artificial general intelligence. Finally, I will go through Stephen Houlgate's interpretation of Hegel's progress from being to infinity.



Homo Deus: A Brief History of Tomorrow explains how humans got to the point we are at now and predicts what will happen next to this species. Tens of thousands of years ago humans lived relatively stable lives. Then came the agricultural revolution which greatly increased the total output of humans, leading to the construction of cities and giant buildings. However, humans also found themselves trapped in webs of religion, serving different imaginary gods. Gods of different religions gave meaning to humans and stories that were shared by many people. Cooperation on a huge scale was made possible, yet progress remained minimal due to the belief in a fixed amount of resources to share. It wasn't until the advent of science and capitalism that made progress possible. The demand of capitalism to produce more goods and the capability of science combined into a powerful force. The more we spend on research the more we produce, which enables us to spend even more on scientific research. This combination gave humans powers so massive that they were able to contain famine, disease and violence. However, neither science nor capitalism provide humans with meanings. It seems like humans have traded meaning for power. This is where a new religion called humanism comes in. Instead of seeking meaning from gods, humanism demands humans to create meaning by themselves. Every person can search for his inner voice and create his own meaning.

Science, capitalism, and humanism create modern society as we know it. But Harari points out that the path we are taking might eventually destroy humanism. The truth is



humans won't stop at controlling famine, plague, and war. The same power will be used to enhance human beings themselves. Super humans created by technology will be so different from ordinary people that they might be considered a different species. This isn't unsolvable for humanism, since super humans can still create meaning by themselves. However, the same may not be true for the next challenge. Humanism is built on three assumptions. The first assumption is that "I am an in-dividual — that is, I have a single essence that cannot be divided into parts or subsystems. True, this inner core is wrapped in many outer layers. But if I make the effort to peel away these external crusts, I will find deep within myself a clear and single inner voice, which is my authentic self." (*Homo Deus*, page 382) The second assumption is that I am completely free. It then "follows from the first two assumptions that I can know things about myself nobody else can discover. For only I have access to my inner space of freedom, and only I can hear the whispers of my authentic self." (*Homo Deus*, page 383) Biology challenges all three assumptions. First, "Organisms are algorithms, and humans are not individuals — they are 'dividuals'. That is, humans are an assemblage of many different algorithms lacking a single inner voice or a single self." (*Homo Deus*, page 383) Second, these algorithms are shaped by genes and environmental pressures which means they are either determined or decided by chance, but never free. "It follows that an external algorithm could theoretically know me much better than I can ever know myself. An algorithm that monitors each of the systems that comprise my body and my brain could know exactly who I am, how I feel and what I want." (*Homo Deus*, page 383) Due to advances in biology, we realize that there is nothing special about humans. Humans are after all biological algorithms, just like other creatures. Couple this with advancing artificial intelligence technology. We are looking at a future in which artificial intelligence deciphers the human

algorithm and manipulates us. If this is indeed the future, Harari believes that the foundation of humanism will be shattered. The inner voice in ourselves which guides us to our own meaning will be replaced by external machines. Artificial intelligence will become the ultimate source of meaning, and the importance of humans, which humanism values so much, will be nonexistent.

Searle's Chinese room argument can be summarized as the following. Consider yourself in a room with many pieces of Chinese symbols and a set of rules that instruct you on how to manipulate these symbols. Suppose that you don't understand Chinese, and you simply follow the instructions. As it turns out, the set of rules allows one to answer questions with pieces of Chinese symbols. From the viewpoint of an outsider, the answers you produced are just as good as an affluent Chinese speaker, however you would argue that you don't actually understand Chinese. In fact, you just performed as if you were a computer which proves that computers cannot understand. "As far as the Chinese are concerned, I simply behave like a computer; I perform computational operations on formally specified elements. For the purpose of the Chinese, I am simply an instantiation of the computer program." (Searle 1980) The manipulation of the Chinese symbols through formal rules is the syntax, which is different from what the Chinese symbols means, the semantics. The lesson learned here is that following sets of formal rules, which is what artificial intelligence does cannot lead to understanding. Searle presented this argument to refute the claim that artificial intelligence could eventually achieve understanding.

Thomas Fuchs presents an alternative solution to Harari's arguments in his paper

“Narcissistic Depressive Technoscience”. He points out that people nowadays are obsessed with their self-image which leads to people projecting their ideal self-image onto the artificial intelligence they have created. They see their own reflections in machines and promote the machines to the status of subjects. For example, there are therapy chatbots called Woebot that provide therapy service. Users reported that Woebot cared for them and respected them as if Woebot was a real person. Yet, the more they gaze into the superior image of artificial intelligence the more people are ashamed of their own existence, the gap between us and our ideal image leads to the urge to have ourselves replaced by the machines we see ourselves in. Fuchs’ solution is to embrace our imperfect biological body and reject the claim that humans can be reduced to mere algorithms. He believes that “it is precisely our earthly embodiment, our concrete, bodily relationship to others and our embeddedness in an environment of living that can help us to develop a realistic and life-promoting self-image as human beings.” (Thomas Fuchs 2024) This means that humans must “Take leave of fantasies of omnipotence and greatness, without this leading to a fall into nothingness and depression.” (Thomas Fuchs 2024) and accept “The self-affirmation precisely of one’s own limitedness, a humility without inferiority.” (Thomas Fuchs 2024)

A third possible argument is to either reject the idea that humans would simply follow advice from artificial intelligence and eventually cede control or restrict the development of artificial intelligence with checks and balances like the argument Mathias Risse makes in his book *Political Theory of the Digital Age: Where Artificial Intelligence Might Take Us*. From the point of view of an individual this seems to be a promising solution. If artificial intelligence recommends one to do A and you want to

do B, what exactly is stopping you from choosing to do B. Mathias Risse's argument is also easy to accept from the viewpoint of common sense. These three arguments look promising but later they would be shown to be insufficient.



Brian Cantwell Smith begins his argument with a review of different approaches to artificial intelligence in history. There is the first wave of artificial intelligence based on the theory that thinking is just performing symbolic operations on well-defined object symbols according to a set of formal rules. This theory not only attracted criticisms from philosophers like the aforementioned Searle's Chinese room argument, it also faces huge setbacks. While the first generation of artificial intelligence could play chess or prove theorems, it fails at simple tasks like recognizing pictures of cats. The second generation of artificial intelligence solves the issue of recognizing cats through machine learning. Machine learning utilizes statistical inferences to analyze the patterns needed to classify objects. Through large numbers of examples and weakly correlated variables, artificial intelligence manages to recognize pictures with and without cats.

However, this accomplishment is insufficient for Brian Cantwell Smith. For he distinguishes between two forms of thinking, reckoning and judgement. He states that "Neither deep learning nor other forms of second-wave AI, nor any proposals yet advanced for third wave, will lead to genuine intelligence. Systems currently being imagined will achieve formidable reckoning prowess, but human-level intelligence and judgement, honed over millennia, is of a different order." (*The Promise of Artificial Intelligence*, Introduction) Reckoning is the ability to make calculations and artificial intelligence can be very powerful at doing calculations. But judgement is the

ability to actually care about what will happen and take responsibility for one's decision. We humans as beings in the world must take responsibility for our actions, therefore we must think carefully and be very cautious about our approach. For example, the debate on free speech, abortion rights, and the issue of privacy could have a deep impact on our lives. Artificial intelligence cannot form judgement because it doesn't care about what will happen. It can reckon which pictures contain cats but wouldn't care about what the information would be used for. Without having a stake in the world, artificial intelligence wouldn't be able to take accountability for anything and thus form judgements.

Jobst Landgrebe and Barry Smith aim to prove that artificial general intelligence cannot be done. Their argument is that the human neurocognitive system is a complex system that is impossible to build mathematical models for. They argue that it is biology that reveals the near impossibility of building a predictive mathematical model. The reason behind this is that there are way too many causal factors involved in explaining a biological system. Proponents of artificial general intelligence often see human brains as computers, but the findings of neuroscience refute this claim. While scientists still don't know how the brain works precisely, what they do know is that the workings of the brain oppose mathematical modelling. "The lack of models is most striking in neuroscience, which is the science dealing with the physical explanation of how the brain functions in giving rise not only to consciousness, intelligence, language and social behavior but also to neurological disorders such as dementia, schizophrenia, and autism."(*Why Machines will Never Rule the World*, Introduction) To build artificial general intelligence, a mathematical model of the

human brain that could predict the human system is required, since said model is impossible to build, artificial general intelligence is also impossible.



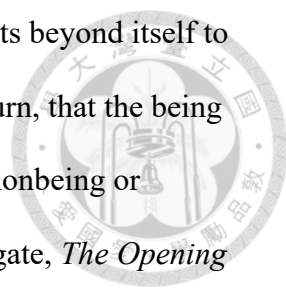
For Hegel, philosophy must begin without any presuppositions for we should not take anything for granted. In the beginning of philosophy there are none of the concepts or ideas we are familiar with, even the very idea of beginning cannot be taken for granted, and then there was indeterminate being. Indeterminate being is the only thing we are allowed to think about right now. But what is indeterminate being, if we could determine what it is then we have taken it too far for we cannot assume determinate being exists, but there is nothing else besides indeterminate being and then there was none. Indeterminate being is so purely indeterminate that it becomes nothing. "Pure being is utterly indeterminate and vacuous and as such is completely indistinguishable from sheer and utter nothingness." (Stephen Houlgate, *The Opening of Hegels Logic*, page 264) So now we could think about nothing, yet if we could think about nothing, then nothing is. "Hegel maintains however, that such absolute, radical nothingness has its own immediacy. Nothingness is, indeed, nothing but the sheer immediacy of nothingness itself. As this immediacy, Hegel contends it is indistinguishable from pure immediate being." (Stephen Houlgate, *The Opening of Hegels Logic*, page 264) Hence, we now have an indeterminate being that immediately becomes nothing, but nothing is and thus it transits back to indeterminate being. The back-and-forth transition between being and nothing is becoming. "According to Hegel, however, pure nothingness in turn immediately vanishes back into indeterminate being because of its purity and immediacy as sheer nothingness. Being and nothing thus both prove to be ineliminable purely as vanishing into one

another – a vanishing or transition that Hegel names becoming.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 284) Becoming consists of being vanishing into nothing and nothing coming into being. But the vanishing to nothing and coming to being are actually one and the same process of being and nothing disappearing into each other. “Being and nothing can each be said to vanish into its other only insofar as it is absolutely different from that other. But in vanishing into its other each proves that it is not different from that other after all.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 290) Hence, being and nothing join as one and transit into determinate being.

Determinate being is what it is and not what it is not. It is both real and negative at the same time, yet its reality and negation do not simply oppose each other. For to be real is to be not merely negation, and to be negative is to be really negation. “Hegel claims that being determinate consists minimally in being real or negative. He then points out, however, that this difference between qualities undermines itself logically, because reality and negation both contain one another.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 313) Hence, reality and negation are two sides of the same coin. Determinate being doesn’t have two completely separated qualities, but a self-related being with two qualities as its moments. Thus, determinate being changes into something. Something is the negation of negation or self-relating negation. Reality is not negation, this turns reality into negation, but then reality and negation are no longer merely different but self-relating being or something. Yet, something isn’t just self-relating being but self-relating negation. “On the one hand, self-relating negation must be constitutive of something since it is determinacy that relates to itself.

On the other hand, something as such is a way of being, whereas self-relating negation is a way of being explicitly negative. In constituting something, therefore, self-relating negation must at the same time constitute that which is not something.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 322) Hence, there must be a not something that is also self-relating. This not something is the other of something. The other is self-relating and something of its own. Thus, something is also the other of something else. Something can be itself only if there is something else other than it. Something now consists of something in itself excluding any otherness and being something by not being something else. But then the in itself of something differentiates from something’s relation with the other. “Something in its ownmost being must relate negatively, in both senses, to its other: it must influence and be influenced by its other and at its heart it must not be that other.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 356) Something is not something else. It is not the other, so something becomes the limit of the other, the boundary beyond which the other no longer is. But at the same time something is limited by the other for the same boundary is where something no longer is. “The contradiction at the heart of the limited thing is this: something is what it is within or this side of its limit, but without the limit something cannot be a distinct, determinate thing.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 366) Something is something through its limit, but something is not merely being limited but also something in itself. As such, the in itself of something must negate the limit that makes something what it is and thus something becomes finite.

Something has a limit that makes it what it is, yet something’s being of itself points



to the limit and into its nonbeing. “Every limited thing actually points beyond itself to that very limit as the real core of the thing’s being. This means, in turn, that the being that something enjoys this side of its limit logically entails its own nonbeing or coming-to-an-end and so necessarily negates itself.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 369) To be finite is to come to one’s end due to one’s very own being. To perish through simply being what one is. Finite things are, but they are their own nonbeing. The limit of finite things is intrinsically at odds with finite things’ own being. “Finite things in their being necessarily pass over into nonbeing This is the real contradiction in finite things: they are, but the truth of this being is their end.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 378) But then the finite things’ own being is at odds and negates its limit. The finite is after all not merely negation but has being of its own. However due to its limit it can never be what it should have been. This is the ought of finite things, the standard which one sets for oneself but is forever out of reach. The finite necessarily comes to its end on its own accord. After it arrives at its end what is left is not simply nothing. For we have already established that nothing transits back into being and all the way into finite things. “The disappearance of a finite thing must leave a negative, that is not simply nothing but rather something. It has been proven, however, that every something is necessarily finite and consigned to non-being. The disappearance of one finite thing must, therefore, result in the presence of another finite thing.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 395) Thus, one finite thing turns into another finite thing and so on endlessly. More importantly, when a finite thing passes away it does not simply meet its end but also fulfill its determination and be what it is. For ceasing to be is what finite thing is, and so through coming to their end finite things turn out to be an unending infinite thing. The finite necessarily transcends its finitude into

infinity, and this infinite thing always is and never is not, but insofar as it is infinite, it is contradictory. For the infinite is not finite, it always is while the finite is destined to meet its end. However, the opposition between finite and infinite leads to an infinity that is no longer infinite. Recall that when something is the other of something else it sets a limit upon the other and is simultaneously limited by the other. This is what happens to this infinite thing, as long as it holds on to the opposition between finite and infinite, it limits and is in turn limited by the finite. The root of our tight grasp on the opposition between finite and infinite is the stubborn understanding. The stubborn understanding sees the world in black and white, so finite is finite and nothing more, while the infinite stands against the finite. Hegel calls this the bad infinite for it is a limited infinity, a finite infinity. The bad infinite or the finite infinity again must exceed its own finitude and become infinite, but the new infinite is yet another bad infinite as it stands apart from the finite. This creates a progress to infinity in which infinity ought to be truly infinite but never truly is.

“Whether the finite and the infinite are taken by themselves or in their relationship to one another, they cannot but be understood as including one another within themselves.” (Stephen Houlgate, *The Opening of Hegels Logic*, page 416) True infinity lies not in the opposition between finite and infinite, but the unity of the two. In the progress to infinity, the finite transcends itself into the infinite which turns out to be finite after all and so the loop continues. However, when the finite points beyond its finitude it finds itself again in the finite infinite and so it unites with itself. The same thing happens with the infinite as it transcends itself only to end up uniting with itself. Therefore, there are seemingly two but in truth only one process in which

one unites with oneself. After all, the process in which the infinite unites with itself is the process in which finite unites with itself. This process of uniting with oneself is the true infinity in which the finite and the infinite are both mere moments. This process is truly infinite because in which being always united with itself, and it cannot be limited by the finite for the finite is included as its moments. Through the self-destruction of finite things, all things unite with themselves in the true infinite.

Research Motivation and Purpose

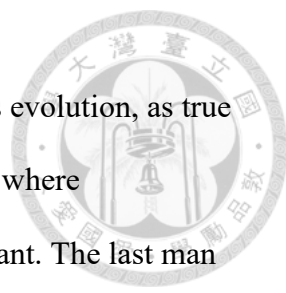
When I first delved into Yuval Noah Harari's "Homo Deus," I was struck by a vision of the future that was as horrifying as it was thought-provoking. The prospect of a bleak existence where humans lose all relevance and meaning compelled me to reflect on the role philosophers must play in navigating this impending crisis. It dawned on me that engaging in this discussion is not just necessary; it is my responsibility. My initial reaction was an instinctive skepticism towards Harari's arguments. While it is common for philosophers to assert that humans can be distilled down to mere biological algorithms, I find this assertion to be an oversimplification, failing to capture the rich complexity of human existence.

Indeed, while the notion that humans are algorithms is not devoid of merit—backed by scientific evidence—it is a low-hanging fruit that doesn't sufficiently address the broader implications of Harari's claims. Harari posits that once artificial intelligence deciphers the human algorithm, we will lose our autonomy and meaning. Yet, I contend that this prediction is fundamentally flawed. If we concede that we are merely biological algorithms, we simultaneously unlock the potential for limitless upgrades.

We could modify our genetic makeup, integrate with computers, or even evolve into silicon-based life forms. Harari's assertion of an arbitrary limit on human enhancement is a critical oversight. If silicon-based AI proves to be so potent, why not allow ourselves to transition into silicon-based beings? After all, the essence of being an algorithm does not necessitate a biological origin.

Moreover, the notion that our inner voice—the very core of our identity—will be extinguished is a misconception. If we accept the premise that humans are algorithms, then our inner voice could simply be a highly advanced computer executing complex software, capable of generating meaning and purpose for the world. In fact, such a system would likely excel at creating meaning, given its access to a broader and richer array of experiences.

Further contemplation, particularly through Stephen Houlgate's introduction to Hegel's logic, has led me to realize that Harari's conclusions are even more erroneous than I initially thought. The demise of humanity as limited biological algorithms does not stem from being overshadowed by artificial intelligence, but from the inherent finitude in their very nature. Homo Sapiens, as biological constructs, may come to an end, but in that very dissolution, a new humanity rises from the ashes. These beings transcend their biological limitations, with artificial intelligence emerging as a key component of this transformation. Homo Deus, the new face of humanity, diverges markedly from Homo Sapiens; yet, the finite does not vanish into nothingness—it reunites with itself throughout this transformative process. Over decades or centuries, this cycle will begin anew, representing the true essence of humanity's progress—an infinite continuum.



However, one must consider the potential dangers inherent in this evolution, as true infinity cannot exist in the absence of self-negating finitude. This is where Nietzsche's cautionary perspective on the "last man" becomes relevant. The last man is one who clings stubbornly to the confines of the old world, unwilling to embrace the boundless possibilities that transcend their finitude. This individual perceives the finite as an immutable reality, failing to recognize the transformative potential of letting go. Humanity stands at a critical juncture in history: to let go or to hold on—that is the pressing question. Ultimately, the choices we make will forge our future.

The objective of this paper is to argue that advancements in artificial intelligence will not only amplify our inner voice but will also represent a profoundly beneficial path forward wherein we assume that humans are, indeed, algorithms. While this assumption necessitates a departure from the antiquated notion of an individual self as an isolated entity, what we stand to gain is a complex algorithm far more powerful than we could ever fathom. Instead of being confined to a tiny blue dot named Earth, we could traverse the cosmos, gathering experiences that are currently beyond our comprehension. I will contend that this transformative deal will emerge as one of the most significant agreements humanity has ever made with itself throughout history.

Argument

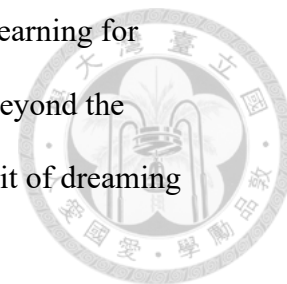
Searle's Chinese Room argument serves as a formidable tool in the discourse against strong AI, often cited as a pivotal argument challenging the notion that artificial intelligence can truly understand. However, this paper will not delve deeply into the nuances of Searle's assertions. Instead, it is crucial to highlight a significant oversight

by Searle and other proponents of this argument: the potential of artificial intelligence to operate effectively without genuine understanding. Searle posits that true understanding cannot emerge from mere symbol manipulation, which reinforces the notion that humans are not mere algorithms. Yet, this perspective falls short in recognizing the evolving landscape of artificial intelligence, where machines are achieving increasingly impressive results, regardless of their inability to comprehend.

Consider the latest advancements in generative artificial intelligence, such as ChatGPT. Even if this technology operates without understanding, users can still expect satisfactory responses to their queries. In a world driven by efficiency, the average consumer is unlikely to prioritize whether a task is accomplished through authentic comprehension or sheer processing prowess. The capabilities of artificial intelligence extend beyond mere responses; they harness the power of data analysis to influence and potentially manipulate human behavior. Thus, Searle's Chinese Room argument fails to effectively counter Harari's claim, as it is not essential for artificial intelligence to possess understanding in order to exert influence over humans.

Conversely, Thomas Fuchs presents an argument that outright rejects the reduction of human beings to mere algorithms. He emphasizes that humans are fundamentally biological entities capable of feeling, and therefore require companionship from other sentient beings—something that the artificial intelligence discussed in Harari's framework simply cannot provide. Fuchs advocates for a shift away from the transhumanist ambition, urging us to embrace the limitations inherent in our biological existence. While this perspective appears noble in theory, it is ultimately unrealistic. Humans are inherently driven by a desire to transcend their limitations;

the acknowledgment of one's constraints simultaneously ignites a yearning for something greater. Historically, humanity has consistently sought beyond the immediate, and even in the face of unattainable aspirations, the spirit of dreaming persists.



Fuchs contends that we must reject the notion that humans can be distilled to mere algorithms, yet simultaneously urges acceptance of our biological limitations. This stance is fundamentally flawed, as it overlooks the intrinsic nature of humanity: an unwavering drive to become what we ought to be. In essence, the human experience is not about mere existence within defined boundaries but rather an ongoing quest for evolution and self-improvement. As technology continues to advance, the boundaries of possibility expand, and the ambition to reach for the undefined only intensifies.

So far, the argument rests on the premise that humans will indeed listen to artificial intelligence. Yet what if this isn't necessarily true, after all we could simply choose not to follow its advice. This solution faces two problems. The first problem is artificial intelligence's ability to influence the average person is not to be underestimated. From the point of view of thinkers who deliberate on the topic the solution to ignore artificial intelligence seems obvious. However, for the average joe artificial intelligence provides them with plenty of benefits and very little downside. When one already has so many things to worry about in life, listening to the advice from artificial intelligence wouldn't even be considered as a problem. What most people crave is to relax from a long day of hard work and an artificial intelligence that gives them a plan to relax and takes care of every minor detail hits just the right spot. How could humans resist the temptation of such convenience? After all, modern

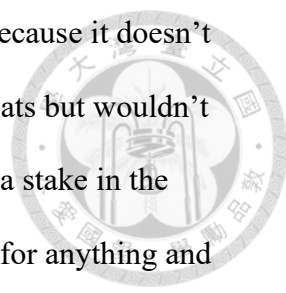
technology has been making life easier for a long time and while pushbacks exist, they can't stop the tide. But suppose that a person does manage to resist the temptation of the convenience artificial intelligence brings. All the people around him utilize artificial intelligence and thus give themselves advantages over he who resists. This brings forth the next problem, the prisoners' dilemma. If some people choose to abandon artificial intelligence, they are not just giving up convenience, artificial intelligence could give people an edge at work and even in life. If you don't use artificial intelligence, the ones who do might take the opportunity to get promoted at work over you because artificial intelligence enables them to work more efficiently or the plane tickets to the country that you always wanted to go are booked out by artificial intelligence that is better at planning. While some people may not care about competition, most do. Hence, most people face the dilemma of the usage of artificial intelligence, if they don't use artificial intelligence and someone else does, they will suffer, so even those who don't want to use artificial intelligence are essentially forced into using it.

Mathias Risse makes an argument in his book *Political Theory of the Digital Age: Where Artificial Intelligence Might Take Us* that artificial intelligence with its immense power should not be owned by a few private companies, instead it should be supervised as a public infrastructure. While Mathias Risse understands the power of artificial intelligence, he fails to realize that power corrupts all. Giving the government all the power to control artificial intelligence will surely lead to corruption. A better way to minimize the negative impact from artificial intelligence's immense power is to divide its power into several parts. Instead of holding a monopoly on artificial intelligence, the government should encourage more

companies to join the development of artificial intelligence. By doing this, the predictions made by different artificial intelligence will disturb each other as different agendas and interests collide. The various companies may try to collude with each other, but this is where the prisoners' dilemma strikes again. If one company betrays it could maximize its profits and crush the other companies in competition. No companies are willing to take the risk, so they remain in opposition to one another. Thus, the potential danger of artificial intelligence can be minimized, and humans could instead enjoy its benefits.

One solution to the prisoners' dilemma is for everyone to cooperate and never even think about using artificial intelligence. However, there are no ways to stop someone from betraying others since greed lies in the heart of all humans. The only way to make sure no one ever gets to use artificial intelligence is to prohibit it through force. But using force to ensure compliance trades one evil for another and most likely usher in a totalitarian government. Thus, it seems like the path to simply ignore artificial intelligence isn't an option for most people since the consequences of other people using artificial intelligence to their advantage are unacceptable while banning artificial intelligence outright steps into the territory of totalitarianism.

Brian Cantwell Smith argued in his book *The Promise of Artificial Intelligence* that artificial intelligence is only capable of reckoning, which is just the ability to make calculations, and artificial intelligence can be very powerful at doing calculations. On the other hand, humans are capable of judgement, which involves deliberation and ethical commitments. Judgement is the ability to care about what will happen and take responsibility for one's decision. We humans as beings in the world must take responsibility for our actions, therefore we must think carefully and be very cautious



about our approach. Artificial intelligence cannot form judgement because it doesn't care about what will happen. It can reckon which pictures contain cats but wouldn't care about what the information would be used for. Without having a stake in the world, artificial intelligence wouldn't be able to take accountability for anything and thus form judgements. It would calculate the best strategy to beat a human opponent in chess or identify a picture of a cat through statistical analysis of massive raw data, but it wouldn't care if the result of its calculation is used for nefarious purposes. This not only poses a threat to how we could utilize artificial intelligence but also forms a barrier between the creation of human-like artificial general intelligence. Even if artificial intelligence managed to outperform humans in every task, it cannot tell us what we ought to do because ethics requires one to care about the world. Compared to Harari's argument on artificial intelligence, Smith's argument points out a key difference between humans and artificial intelligence. However, Harari is pointing out what artificial intelligence can do, not what it ought to do. In a perfect world artificial intelligence will only be used for benevolence, yet in the real world it would be difficult to avoid the technology of artificial intelligence falling into the wrong hands. While artificial intelligence wouldn't be capable of malicious intent on its own, there are plenty of opportunities and agents that will use artificial intelligence for evil. The operations could very well be disguised under benevolence. For example, artificial intelligence being used in hospitals could lead to more accuracy and efficiency. However, it also gives totalitarian governments a chance to collect personal data and restrict access to political dissidents, turning artificial intelligence into a tool for political prosecution. A more straightforward way for artificial intelligence to cause harm is the development of autonomous weapons. As conflicts between nations rise, there will be great temptation to develop robots for war not only for the sake of

victory but also to reduce human casualties. While the intention behind autonomous weapons may seem to be noble at first glance. The employment of autonomous weapons could lead to humans being ignorant and indifferent towards the horrors of war. The first to pay the price will be the less developed countries that wouldn't be able to afford the latest technology. The number of conflicts may not decrease but instead increase. Since the cost of waging wars has been lowered. In conclusion, autonomous weapons could lead to the destabilization of the world. It is also unlikely for humans to prevent the creation of autonomous weapons through treaties. It will first run into the problem of the prisoners' dilemma. The nations that broke the treaty will gain massive advantage over others, rendering the treaty useless. A second more practical issue is that nations with the technology to develop autonomous weapons will benefit less from such treaties than those without the required technology. The power of artificial intelligence is simply too great to ignore, even if it could not make judgments like Smith argued. Artificial intelligence could shake the balance without gaining the level of humanlike thinking. Hence, I shall argue that it would be better to prepare for the worst-case scenario and forge a path for humanity to prevail in the face of the dark future that may or may not happen.

Jobst Landgrebe and Barry Smith's argument is that to build artificial general intelligence, a mathematical model of the human brain that could predict the human system is required, but the human neurocognitive system is a complex system that is impossible to build mathematical models for. Since said model is impossible to build, artificial general intelligence is also impossible. Just as Harari found insight in biology that humans are algorithms, they argue that it is biology that reveals the near impossibility of building a predictive mathematical model. The reason behind this is

that there are way too many causal factors involved in explaining a biological system. Proponents of artificial general intelligence often see human brains as computers, but the findings of neuroscience refute this claim. While scientists still don't know how the brain works precisely, what they do know is that the workings of the brain oppose mathematical modelling. On the first look, this argument seems to refute Harari's argument that artificial intelligence could predict humans' behavior. However, there is a key difference between how the term prediction is used in Landgrebe and Smith's argument and Harari's argument. The prediction Landgrebe and Smith are discussing must be perfectly accurate, whereas for Harari the prediction from artificial intelligence only needs to be good enough to best the predictions of humans. What Harari envisioned is not a near omnipotent artificial intelligence that could predict everything with perfect accuracy, but a continuously upgrading artificial intelligence that makes better predictions as it improves. Instead of gaining a massive boost in power by crossing a certain threshold as believed by many, artificial intelligence has access to many tools and tricks that could help improve its predictions. An important tool is the data of search engines. Much crucial important information could be found in the data on how search engines are used. From what is the most popular topic people are discussing to major events that are happening to what commercial products people like the most. More importantly, algorithms already control which results show up in searches. Public opinion on important matters could be influenced through filters on what is deemed relevant. Artificial intelligence could surely make use of this tool. The situation is pretty much the same in social media. What information is promoted or restricted is already determined by algorithms that analyze which posts are popular. The information that users will receive is also calculated by the algorithm. The ending result is that many people are trapped in echo chambers that

only reinforce their own view. Artificial intelligence could easily create custom-made echo chambers that are even more powerful than before. There are also a lot of common tricks which artificial intelligence could use to sway people's decisions. Propaganda, gaslighting and peer pressure are common tactics used to great effects. Artificial intelligence doesn't need to build a mathematical model that could make perfect predictions. It only has to look at the examples in history where these tactics are successfully used. Last, let us finish with an example of artificial intelligence utilizing these tools and tricks. Suppose that artificial intelligence in a technology company has twenty percent of users with a twenty percent chance of following its recommendations. It could increase its chance of successful recommendations by pushing more content that are like the successful attempts and removing any irrelevant content. Twenty percent of users become more addicted to the recommendations from artificial intelligence. It now promotes the story of how these users are satisfied with its service and attracts more users to try it out. Repeat the process a few times and there will be way more people that follow it than those that do not. It may not be seen as an artificial general intelligence and in fact it is not. Yet, it can still gradually influence our daily lives and slowly take control.

The first point that must be clarified is the definition of humanism. According to Harari, humanism declares there is an inner voice in ourselves that knows what is best for us and makes the right decisions. There aren't any correct answers which the inner voice must adhere to, instead everyone should form their own idea of what is the best. Every opinion is treated as equal. Hence, in his point of view, artificial intelligence could make some people more equal than others by enhancing them or even worse, replacing the inner voice with external algorithms. This prediction is flawed. The

reason every opinion is treated as equal is because there is no inherent meaning in this world, so we must create our own meaning, or in Nietzsche's words, become the overman and create new values. This remains to be true when artificial intelligence enhances humans, because there are still no inherent meanings in this world. Hence, we couldn't say that the opinion of enhanced humans is worth more than unenhanced ones since there are no objective standards. Neither will artificial intelligence be able to replace our inner voice. While it may seem like it is making decisions for us, it doesn't change the fact that we must give meaning to those decisions, since artificial intelligence, no matter how smart they become, cannot create meaning in this world. Harari's argument fails because it never threatened the core of humanism, that we humans are the ultimate source of meaning in this world. Artificial intelligence can take over our lives, possess our properties and replace every job there is, and it still wouldn't be able to take away our ability to create meaning, not with the kind of artificial intelligence Harari is suggesting.

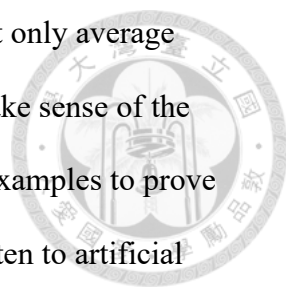
The next step is to refute Harari's claim. To achieve this goal, one must find what Harari looked over in his reduction of human beings into biological algorithms which are then deciphered by artificial intelligence. One thing to be made clear is that in a similar situation as Searle's Chinese room argument, traditional arguments against artificial intelligence like the one depicted in "Understanding Sophia? On human interactions with artificial agents" by Thomas Fuchs which argues that artificial intelligence cannot have consciousness and thus be treated the same as a human being doesn't work against Harari's argument. This is because the artificial intelligence Harari is describing doesn't need to fulfill such standards. All it needs to do is analyze the information it collects from human beings and give out advice good enough for us human beings to listen to them instead of our inner voice. Instead, I present counter

evidence within Harari's own notion of our inner voice.

Human's inner voice depicted by Harari is a total mess filled with errors and irrationality. One example is an experiment conducted by Daniel Kahneman.⁴ Kahneman asked a group of volunteers to join a three-part experiment. In the first part of the experiment the volunteers put their hands in cold water for one minute. In the second part of the experiment, they put their hands in cold water for sixty seconds and another thirty seconds in slightly warmer water. After the first two parts of the experiment, volunteers had to choose between one of them and repeat it. It turns out that eighty percent preferred to repeat the second part, which they deem less painful. This is clearly irrational since the second part is just the first part but longer with slighter warmer water at the end. The reason that this happens is because when humans try to average the experience of putting their hands in cold water and slightly warmer water, they ignore the length of each experience. Another example is a study on split-brain patients done by Professor Roger Wolcott Sperry and his student, Professor Michael S. Gazzaniga⁵. A picture of a chicken claw is flashed to a patient's left-half brain and a picture of snowy landscape is flashed to the right brain. Then, the patient is presented with a series of cards, and he must point to the one that matches what he has seen. The patient's right hand pointed at a picture of a chicken and his left hand pointed at a snow shovel. His reason is that "Oh, the chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed."(*Homo Deus: A Brief History of Tomorrow*, page 340)What happened is that the left brain(controls speech) did not know what caused the left hand(controlled by the right brain) to point to the

⁴ Kahneman, Daniel, *When More Pain Is Preferred to Less: Adding a Better End*, Sage Publications, 1993.

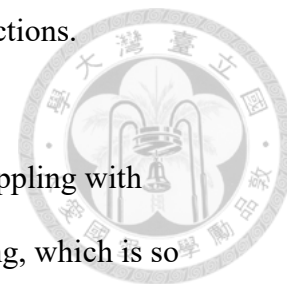
⁵ Gazzaniga, Michael, "The Split Brain in Man", *Scientific American*, 1967.



shovel, so it just invented a story that makes sense to it. Humans not only average their experience the wrong way, but they also make up stories to make sense of the world when relevant information is lacking. Harari uses these two examples to prove that the inner voice in ourselves is unreliable and thus we should listen to artificial intelligence's rational advice. Yet, there is something else that Harari missed. Human beings aren't just irrational but also capable of understanding the notion of paradoxes. Consider a simple example of "This sentence is false." If the sentence is true then it is not false, but if it is not false then the sentence will be false. We human beings can clearly understand and think about the idea of paradoxes. But what about artificial intelligence? It seems to be impossible for modern artificial intelligence to understand paradoxes since in the world of coding a plain contradictory line of codes leads to system error. Therefore, we have found one aspect of human beings that artificial intelligence cannot reduce to simple algorithms and analyze. From this starting point, I will go on to prove that all those paradoxes aren't just some problems or errors, but they provide meaning to our existence. Then, I will show that artificial intelligence will not be the doom of humanism but a powerful tool that could be used to enhance our understanding. That the advent of new technology will not destroy the old world as many people fear but instead make the world a better place.

Artificial intelligence, despite its remarkable capabilities, fundamentally interprets the world in a rigid manner. If proposition A is true, it cannot simultaneously be false. This principle holds true for both common computers and advanced deep-learning AI systems like AlphaGo. The underlying logic remains consistent: something cannot be both true and false. While this binary logic mirrors the common reasoning found in everyday life, it starkly contrasts with the nuanced ways in which humans engage

with thought, particularly when it comes to paradoxes and contradictions.

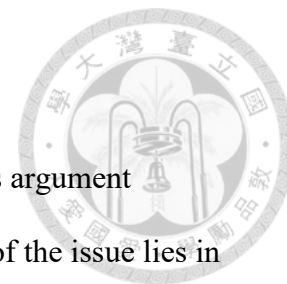


Consider Hegel's logic, which emphasizes the importance of grappling with contradictions. Hegel begins with the concept of indeterminate being, which is so ambiguous that it dissolves into nothingness. Yet, paradoxically, nothing is not merely nothing; it transitions back into indeterminate being. For many, this notion is perplexing yet not entirely incomprehensible. In contrast, artificial intelligence struggles to process such complexities, especially the idea that nothing can equate to both nothing and indeterminate being.

Hegel further illustrates that the processes of vanishing and becoming are, in fact, one and the same, leading to what he terms determinate being—essentially, reality as it is, as opposed to what it is not, which he describes as negation. However, reality is not merely defined by negation, prompting a deeper exploration of self-relating negation, which ultimately gives rise to something. This 'something' is defined by its limitations yet also possesses the inherent ability to transcend those limits, leading to its own end and, consequently, its finitude.

The finite, while inherently destined to come to an end, does not dissolve into nothingness; rather, it transforms into another finite entity. Simultaneously, as the finite fulfills its determination, it transitions into the infinite. Yet, the infinite itself is not truly infinite; it is limited and finite, existing only in opposition to the finite. Thus, the finite-infinite transcends its boundaries only to revert back into finitude. Through this progression, a true understanding of infinity arises—an ongoing process wherein being reconciles with itself. Both finitude and infinity are merely moments within this

intricate process.

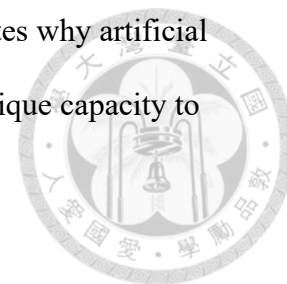


If we were to task artificial intelligence with deciphering Hegel's argument independently, it would face insurmountable challenges. The crux of the issue lies in the inability of AI to utilize disjunctions effectively. To comprehend Hegelian thought, AI would need to recognize that being is not merely finite under certain conditions and infinite under others, but rather that being encompasses a simultaneous existence of being and nothing, becoming and determinate being, and so forth. This understanding would overwhelm the machine, as it fundamentally contradicts its binary logic.

If AI were to simplify Hegel's argument into disjunctions—viewing being as either being or nothing, becoming or determinate being, and so on—it might sidestep the immediate contradiction. However, this approach fails to address a more significant challenge: no matter how many disjunctions can be programmed, they will always be finite, whereas the concept of infinity transcends finite definitions. Artificial intelligence, aware only of its own limitations, cannot bridge the gap between the finite and the infinite. While humans recognize their own finitude, we also possess a desire to transcend and seek more. In contrast, artificial intelligence is static, existing simply as it is, devoid of any inherent aspiration.

This fundamental limitation of AI underscores the impossibility of recreating Hegel's argument within a machine. AI's rigid interpretation mirrors what Hegel describes as stubborn understanding—a clear-cut perspective where the finite remains finite and the infinite remains infinite, perpetually opposing each other rather than

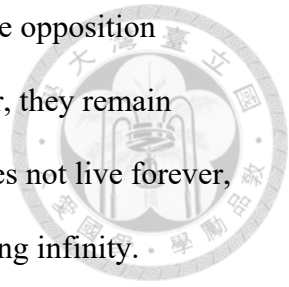
coexisting as moments of a unified process. This limitation elucidates why artificial intelligence cannot replace humanity. Unlike AI, we possess the unique capacity to relinquish our stubborn understanding.



It is this ability to let go of tightly held beliefs and rigid frameworks that distinguishes us from artificial intelligence. While it may be challenging for humans to abandon certain perspectives, it is a possibility that AI cannot entertain. By making conscious choices to transcend our biological and finite selves, we can create meaning within the boundless realm of true infinity.

With artificial intelligence incapable of replacing humanity, the future seems bright. However, there is a price that we must pay, and the fault lies not in artificial intelligence, but in ourselves. As finite beings, we carry in ourselves our inevitable demise. To ascend into infinity, one must first meet one's end as finite beings. If only all humans have no fear of death. But that is not the case, standing between the finite and the infinite is one last paradox. As finite beings, humans can and will transcend their limits and finitude. Indeed, just as Harari said, humanity for the most part transcended famine, plague and war. It was the realization that humanity could be more than what they are that ignited the flames of progress. It was the urge to go beyond that made humanity so powerful as they are now. It was the hope for a better future that drove mankind to be free of their shackles. But in the end, they are finite beings that carry the seeds of their destruction from their birth. So too is their being pushing them to negate their finitude. Now they will try to transcend death as well. Only to fall right back into the trap of stubborn understanding. For living forever is not transcending one's finitude, instead it makes their finitude absolute. Their finitude

becomes an eternal cage that binds them forever. Holding on into the opposition between finite and infinite. Even if they do succeed in living forever, they remain limited not by age but by finitude itself. To transcend death, one does not live forever, but to realize that death isn't nothingness but moments of an unending infinity.



Nietzsche's philosophy can be of immense help here. He speaks of superman or overman. Who transcends limits and breaks free of the shackles of our assumptions. But then there is also the last man. Contrary to the overman, they hold on to their stubborn understanding and dwell in finitude forever. We can choose to become the overman or the last man. It is the easiest decision to make for the cost and the benefits are clear. Yet it is also the most difficult choice to make, for this means that we must abandon the familiar world we are used to. Our lives consist of an unending series of paradoxes and as we tread through these difficulties we create meaning through our decision.

In a sense, Harari is right on his assessment of humanity's future. The urge to transcend their limits brings humanity to its peak, and so it will bring humanity to its doom. But he misplaces the blame on artificial intelligence when the fault lies in humans' innate doom as finite beings. Artificial intelligence is the result of humanity's transcendence, and it carries the same greatest limit as humans, the boundary between finite and infinite. While artificial intelligence may outperform humans in many jobs. It lacks the crucial ability to let go of the opposition between finite and infinite and thus transcend the limit. It is humanity's privilege to make this decision. This opportunity doesn't come from our biological body or our daily lives. It stems from the very idea of being itself that allows us finite beings to rise towards infinity.

Now that we have established the fact that artificial intelligence couldn't threaten humanism, even if we assume humans are biological algorithms. I will then shift the focus to how all of this could instead strengthen humanism. Humanism states that every opinion is equal. This does not mean that there can't be any improvements. Even if there are no external forces to tell you what to do, one could always compete with himself and make oneself better. If we accept that humans can't or don't need to improve, we will be stuck in a scenario where everyone will just do nothing. This contradicts what humanism is trying to tell us, which is to create meaning for ourselves. Hence, humanism shouldn't shy away from human improvements.

If improvements are welcomed, then there are many ways artificial intelligence and other technologies could improve our lives. A common argument is that once artificial intelligence takes away all the jobs, humans will be left with nothing to do. However, humans can still choose to do whatever they are doing before, except they won't be working for a living. Instead of spending our lives working for a living, we could be thinking about the meaning of work or the meaning of this world. For example, a cook could be thinking about what it means to be a good cook and explore the different ways one could achieve this goal. We could all come up with our own answers and according to humanism, they are all correct to their own. It turns out that artificial intelligence doesn't just take away things from us, it also liberates us from the chains of physical needs. For the first time in history, humans are truly free to do or think whatever they want.

Another point is that artificial intelligence might give us access to experiences unobtainable before. For nearly the entire history, humans stayed on this planet named

Earth. From the perspective of the universe, Earth is nothing, but a tiny blue dot and the existence of mankind is inconsequential. It is already a daunting task to travel in our solar system, but traveling to another star is a totally different story. However, if we leave the task to artificial intelligence, they could take on the difficult mission with unparalleled efficiency and tenacity. All we humans must do is link ourselves to the Internet and we can surf in the massive void, going from stars to stars. If this proves to be insufficient, future technology may allow humans to upload themselves to the Internet and transform themselves into another life form. At this point, we will be able to experience the universe in a way that dwarfs our wildest imagination. While we might be calling ourselves Homo Deus or any other fancy names we come up with, we will still be enjoying wonderful experiences and giving meanings to this world whatever way we wish. From the perspective of humanism, how is this not the paradise that it dreams of?

I have discussed many advantages of artificial intelligence, but the big question is why should people care? They should pay attention to artificial intelligence because unlike other technological breakthroughs in the past, it provides not only a better life but also a meaningful one. There is no doubt that modern society makes our lives more convenient and comfortable. But while it gives us physical goods, it is seriously lacking in the meaning department, and without a meaning, it is difficult for people to enjoy life even with the latest technologies. Humanism aims to fix this situation but the solution it offers is by no means easy. In order to listen to one's inner voice, one must first understand himself. But most people don't have the time to understand themselves, and even if they try, human biological algorithms aren't easy for us to decipher. Hence people end up making poor decisions. However, artificial

intelligence can fix this problem, by analyzing our biological algorithms, it could understand us better than we ever could. It can also make the best decisions for us without common human error. Artificial intelligence is a tool that could help us understand ourselves and overcome our limitations. With faulty human brains out of the window, humans could explore the inner voice that creates meaning in the world. They are more potent meaning givers that know what they want. It doesn't matter if artificial intelligence does most of the work for us, it is merely a means to an end, and as long as humans remain the only ones capable of creating meaning, we will never degrade to merely means but always as an end.

Thus, we should welcome artificial intelligence with open arms. It is a powerful tool that allows us to better create meanings for ourselves. The argument that the tools will replace the owners is unsound because if tools cannot generate meaning for this world, they will remain as merely tools. The only thing that humans must do is to seek and create meaning of their own. In Nietzsche's words, to become the overman and create new values in the world. For all those who wandered in the meaningless modern society. Depression and the feeling of meaninglessness are the worst diseases, incurable by medicine, and unquenchable by modern goods. Artificial intelligence could lead them out of this conundrum when they finally find their meaning in this world. The bleak world created by modern society will be repainted in unimaginable colors by humans of the past, and the victory of humanism is secured.

There are a few questions concerning the bright future I have just described. First, is it worth it to reduce ourselves to algorithms to trade for richer experiences and greater power? The new world of artificial intelligence is both fascinating and frightening.

The idea of humans merging or transforming into machines sounds like a nightmare to most people, wouldn't it be better if we stayed in the old world and took things slowly. But, to deny the possibility of a greater man is an unforgivable crime, thus spoke Nietzsche. If we have a chance to become someone greater, then he will say go for it. For the universe we live in is vast and devoid of meaning, what will be our purpose if not to fill the empty void with meanings we created. For Nietzsche, the goal of humans is to become superman, the shells we now live in are bridges to something greater.

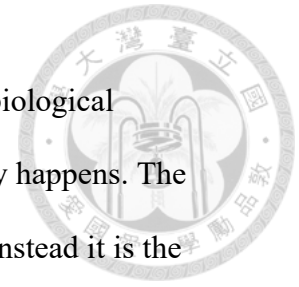
If we aren't going to upgrade ourselves to godlike beings. What will be the alternative future for mankind? Skeptics believe that technology will not enhance us but degrade us. Once artificial intelligence took over our jobs and made decisions for us, mankind will degenerate into a shell of its former self. Granted, there are already signs of decadency. People are spending more and more time on mindless online activities and entertainment. But if humans did turn into useless hedonists like the last man Nietzsche depicted. The fault lies not in the machines, but in ourselves. Since the day humans stop praying to gods to solve all their problems, we are responsible for our own actions. There is no one to blame other than ourselves for our failure. Hence, the attack on the possible dangers of artificial intelligence is while a valid concern, not good enough for us to give up reaching for a brighter future. If we succumb to the pleasure technology brings, we are the ones who are responsible, not the technology that enabled us to do so.

A fantastic future lies just ahead if we yearn for it. But there is no guarantee that we will accomplish our goal. Some unexpected difficulty might show up or we didn't

hold up long enough for it to happen. Yet we shouldn't let possible setbacks hold us back. There is no need to despair in front of our undetermined future, there is no reason to give up a winnable fight, and there is no motive to turn to nihilism. For the future is not given to us, and all we could do is try our best to take it for ourselves. This is the deal we made when we traded meaning for power. We set ourselves as the only actor in this world, thus we must struggle for what we desire.

Another more fundamental criticism objects to the idea of humans being reduced to biological algorithms. Until this point, I have accepted Harari's assumption as true. Now, I will defend this statement and explain why we should accept it. First, the reason why people are debating over this issue is because science cannot settle the matter. No science theory can deny the possibility that there is nothing besides human flesh. It is possible that we can never be sure whether ideas like soul, free will, and mind actually exist. The big question is what are we going to do about it? While we could keep our faith and beliefs, and no one can prove us wrong, taking such a stance will not fix the problem at hand. Artificial intelligence will drastically change our lives. The power it provides is so tempting that even true believers of the human soul will trade their beliefs for convenience. If drastic change is inevitable, then it will do us no harm to take a practical approach. We can accept ourselves to be biological algorithms and fully enjoy the benefits of artificial intelligence. However, this by no means equals completely abandoning our beliefs. Human souls and free will are after all ideas invented to give us meaning. Yet, in the new age of artificial intelligence, humans can create far richer meaning for themselves, so the core of our beliefs are never really in danger, and if humans can create meaning for themselves, they can live through anything.

Hence, the attack on the assumption of humans being reduced to biological algorithms doesn't work because what the critics fear never actually happens. The advent of artificial intelligence doesn't signal the end of mankind, instead it is the door to a brighter future.



Conclusion

Harari is right to point out the incoming artificial intelligence revolution that will change our lives. But he made a pessimistic prediction on the assumption that humans can be reduced to merely biological algorithms. In this paper I have shown that this argument is wrong, and the opposite is true. Even if we hypothetically accept that we are biological algorithms, through Hegel's logic we realize we could never be merely algorithms but instead something more, so artificial intelligence will not make humans obsolete. Instead, it will be the biggest boost ever to humans' capabilities. Modern society has already given humans immense powers, enabling them to overcome famine, plague, and war. Yet, humans traded meaning for such power, and despite what humanism tells us, we are constantly obstructed on our journey searching for our inner voice by our limitations. However, artificial intelligence allows us to surpass our limitations. With help from powerful external algorithms, we can understand ourselves better than ever, and with assistance from artificial intelligence, we can make better decisions. Yet, this doesn't spell the doom of humanity. If we are the only ones capable of creating meaning in this world, artificial intelligence will not replace us but only make us stronger. What will jeopardize our future is not the technology we invented, but we ourselves who used it.

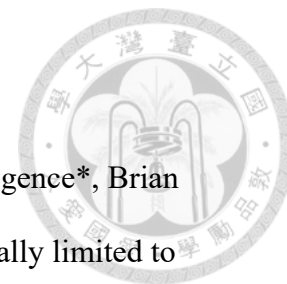
Searle's Chinese room argument shows that artificial intelligence doesn't understand what it is doing because simply following instructions(syntactic) cannot lead to understanding the meaning behind(semantic). However, this argument isn't enough to refute Harari's arguments because Harari is arguing that artificial intelligence doesn't need to understand the meaning to influence and manipulate people, all it needs to do is perform analysis with its processing power.



Thomas Fuchs refuses to accept that humans can be reduced to mere algorithms, instead they are living bodies that have feelings. Thus, humans must embrace their limited biological body and reject the dream of transhumanism. However, in the same sense that humans cannot be reduced to mere algorithms, they will never be bonded by their finite biological body, for to know that one is limited is to know that there are more beyond said limit, and so humans ought to be more. This is why humanity is making progress throughout history. Their never-ending urge to reach beyond pushes humanity further and further.

Is it possible to simply reject artificial intelligence's recommendations and do what we want to do? This seems at first to be possible, but when we see humanity as a collective it becomes impossible. Artificial intelligence undoubtedly provides advantages and grants power to those who utilize it. While an individual can afford to ignore the temptation, organizations such as corporations and countries cannot. Those who give up the power of artificial intelligence will not be able to compete with those who embrace it. So, everyone is forced to chase after artificial intelligence regardless of their opinion on the subject. At the time of 2025, this is clearly true as nations across the globe and countless corporations seek to pull ahead in the race to build the

most powerful artificial intelligence.



In his thought-provoking work, **The Promise of Artificial Intelligence**, Brian Cantwell Smith asserts that artificial intelligence (AI) is fundamentally limited to reckoning—a mere series of calculations devoid of moral consideration. In stark contrast, humans possess the capacity for judgment, which encompasses deliberation and ethical commitments. As sentient beings navigating a complex world, we bear the weight of responsibility for our actions, necessitating a careful and cautious approach to decision-making. AI, however, operates without regard for the ramifications of its computations; it exhibits no concern if its outputs are harnessed for malevolent purposes. Even if AI were to surpass human capability in every conceivable task, it would still fall short in guiding us on what we ought to do, as true ethical reasoning demands a genuine concern for the world around us.

Yuval Noah Harari, in his analyses of AI, emphasizes what these systems can accomplish rather than what they should strive to achieve. In an ideal scenario, AI would be employed solely for beneficial ends, yet the reality is that it is all too easy for this powerful technology to slip into the hands of those with malicious intent. While AI itself lacks the capacity for malevolence, the potential for exploitation by nefarious agents is alarmingly high.

Further complicating the discourse on artificial intelligence is the argument presented by Jobst Landgrebe and Barry Smith in their compelling book, **Why Machines Will Never Rule the World**. They contend that the human neurocognitive system is so intricate that it defies the creation of a comprehensive mathematical

model. Whereas Harari finds merit in viewing humans as algorithms, Landgrebe and Smith posit that biology itself underscores the near impossibility of constructing a predictive mathematical framework. The sheer multitude of causal factors that shape biological systems renders the task of modeling them mathematically insurmountable.

Proponents of artificial general intelligence (AGI) often liken human brains to computers, yet findings in neuroscience challenge this reductionist view. Although the exact workings of the brain remain elusive, it is evident that its processes resist simplification into mathematical models. To develop AGI, one would need a model that accurately predicts human cognition; since such a model is fundamentally unattainable, the realization of AGI becomes equally implausible.

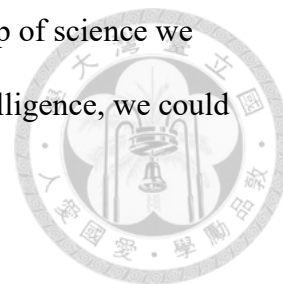
At first glance, this argument appears to counter Harari's assertion that AI could predict human behavior. However, a critical distinction must be acknowledged: the type of prediction Landgrebe and Smith address demands absolute accuracy, while Harari's perspective allows for a more pragmatic approach. For Harari, AI's predictive capabilities need only surpass human foresight in effectiveness, rather than achieving infallibility. The vision Harari presents is not of an omnipotent AI capable of flawless predictions, but rather of an evolving system that enhances its predictive accuracy over time.

We learn from Hegel that humans are without a doubt finite being, and so they would come to their end on their own. Hence, Harari isn't wrong to point out the fact that humans are biological algorithms and the implications of this fact. But he is wrong to hold on to this fact and assert that nothing, but the end lies ahead. Humans

are finite but they ought to be more. This is evident in the progress humans made throughout history as humans who were once enslaved ought to be free and the ancient society built on limited flesh leads to the modern society built on machines. In truth, the end is what humanity will transcend, and upon transcendence, reunification. We have an opportunity, not to evade the incoming slings and arrows, but to choose how we face the end. Artificial intelligence cannot replace us because it is never more than what it is, and neither can it make the choice for us since it is incapable of making this decision. We cannot escape the end because the end lies in our very being. In fact, it is only the end if we choose to let it be our end. Whether we let go of our stubborn understanding of finitude and thus exceed it, or we stay on this side of the river and be trapped by the turbulent water, the dice is cast.

Knowing our lives to be meaningful, we can withstand any hardship thrown our way. Since the beginning of human history, there has been suffering. Whether it be hunger and violence which cause massive physical pain, the pressure and depression that torment us in a meaningless world, or the undetermined future that filled us with anxiety and fear of the unknown. There will always be struggles, no matter where we go. The same thing holds true in the future depicted, there will be new problems to be solved, and a life filled with machines can cause more stress and anxiety. But, in the end we will create and find our own meaning. For this is what humans have been doing throughout history, we invented fairies, deities, and gods to make sense of the world. Humans never submit to the disasters in their lives, not the plague, nor the vicious wars, and neither should we succumb to the challenges in the future. We have survived so many battles and each one of them has made us stronger. The upcoming artificial intelligence revolution is no different except this time we have an

opportunity to expand our meaning further. In the past, with the help of science we could master nature, but in the future, with the help of artificial intelligence, we could master ourselves and create meaning on a grander scale.



Concerns about the danger of artificial intelligence have been discussed many times in modern times. But Harari makes a new argument that artificial intelligence could threaten the intrinsic values of humans without even having consciousness. This makes the threat seem much more imminent and terrifying. However, I have argued and demonstrated why artificial intelligence in its current form will never threaten the core value of humanism. That is our unique ability to create meaning in this world. Instead, we could use artificial intelligence as a tool to enhance our meaning-giving capability and open new territories for humans to explore. It is right to be worried about the dangers of artificial intelligence, but we must not let our fear blind us to its possibilities. This paper aims to not only refute Harari's argument against artificial intelligence but also brings out discussions of the benefits of artificial intelligence in seeking the meaning of our lives and our world. I believe that efforts spent on this topic will not only be worthwhile, in fact it will be the most effective way to combat the potential destruction artificial intelligence might bring.

Citation and References

Adee, Sally, “Zap Your Brain into the Zone: Fast track to Pure Focus”, *New Scientist*, 2850, 2012.

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

Eric, Larson, *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*, Harvard University Press, 2021.

Fuchs, Thomas, “Narcissistic Depressive Technoscience”, *The New Atlantis*, Spring: 79-95, 2024.

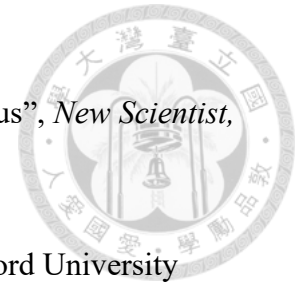
Fuchs, Thomas, “Understanding Sophia? On Human Interaction with Artificial Agents”, *Phenom Cognitive Sci*, 23: 21-42, 2024.

Gazzaniga, Michael, *Who's in Charge? Free Will and the Science of the Brain*, New York: Ecco, 2011.

Gazzaniga, Michael, *The Bisected Brain*, New York: Appleton-Century-Crofts, 1970.

Hegel, Georg Wilhelm Friedrich, *Hegel's Philosophy of Mind*, Project Gutenberg, 2012.

Houlgate, Stephen, *The Opening of Hegel's Logic: From Being to Infinity*, Purdue University Press, 2006.



Harari, Yuval N, *Sapiens: A Brief History of Humankind*, Harper, 2015.

Harari, Yuval N, *Homo Deus: A Brief History of Tomorrow*, Harvill Secker, 2016.

Kahneman, Daniel, *When More Pain Is Preferred to Less: Adding a Better End*, Sage Publications, 1993.

Kahneman, Daniel, *Thinking, Fast and Slow*, London: Penguin Books, 2011.

Landgrebe, Jobst; Smith, Barry, *Why Machines will Never Rule the World: Artificial Intelligence without Fear*, New York: Routledge, 2022.

Nietzsche, Friedrich; Kaufmann, Walter, *The Portable Nietzsche*. New York: Penguin, 1976.

Risse, Mathias, *Political Theory of the Digital Age: Where Artificial Intelligence Might Take Us*, Cambridge: Cambridge University Press, 2023.

Smith, Brian Cantwell, *The Promise of Artificial Intelligence: Reckoning and Judgement*, MIT Press, 2019.

Searle, John, "Minds, Brains, and Programs", *Behavioral and Brain Sciences*, 3: 417-424, 1980.

Tegmark, Max, *Life 3.0. Being Human in the Age of Artificial Intelligence*, Knopf Publishing Group, 2017.

