Seeing Eye to AI: A Comparison on the ethics of AI services and African American Enslavement as Tools for Humanity

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

By

Rishal Jagtap

Fall 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Introduction:

In 1996, the first ever AI, Deep Blue, challenged World Champion Chess player Garry Kasparov to a chess competition… and lost 2 to 4. However, in only a year, Deep Blue would once again face Kasparov in a duel, in which Kasparov would be stunningly defeated 2 ½ to 3 ½. In a matter of a year, this robot was able to dominate the world chess player and leave a milestone in the history of artificial intelligence forever. For the first time since its creation, AI had proven itself to be more capable than man.

From hereon, artificial intelligence became a field that skyrocketed in popularity amongst major tech companies. Roughly "86% say that AI is becoming mainstream at their company in 2021" (McKendrick). Many businesses have understood the power of AI and its capabilities for analysis, seeing it as a major opportunity to accelerate their growth since the pandemic. Not only can AI be used as a tool for big data analysis and keen decision making, but they can also fill in job shortages in menial work, and allow for greater expression in higher existing careers that rely on such jobs to be completed before coming into full effect. With their ability to handle both physically and mentally burdening jobs, once fully developed, AI may very well take over most jobs that aren't reliant on creativity.

Despite the progression of AI thus far, it is still unclear as to whether we as a society can see eye to AI (pun intended) with artificial intelligence, primarily because we have a varying understanding of what sentience means, especially for non-human objects. Some perceive sentience to be a trait solely attributed to human beings, as Sabouret implies when describing sentience: "We are men, so we know what it is to be men, as cats know what it means to be cats" (Sabouret). While we do not fully understand what it means to be a cat, we do know what it means to be a human. In a similar vein, we may never know what it means to be "artificially

intelligent" and so to view them with the traits derived by humans to describe human behavior will be challenging without redefining the term itself. Lastly, while the prospects of such a powerful tool are great for the development of companies, many fear that they will ultimately be replaced in the workforce.

Though artificial intelligence is a fairly recent part of human history, becoming more prominent in the past 70 years, the unease associated with complex machinery uprooting humans has been present since ancient times. During the Renaissance, the birth of the Weaving Looms, a tool for weaving clothing, greatly accelerated the ability for man to produce clothing. Some members of the community saw this invention as the "Devil's Machine", a creation brought into this world by humans, only to replace them. Others were inspired to pursue more technology so that humans could one day live without ever needing to work! (Sabouret, Introduction). Ultimately, it is this rift in human understanding that makes some afraid that AI will eventually take over their likelihood and render them jobless. If they become more and more like humans, it is likely that these groups will find comfort in suppressing their growth and dehumanizing them.

If Artificial intelligence does eventually develop into sentient beings akin to humanity, then it could result in severe effects for the laws currently in place, as they are primarily devised for human need rather than technology. As of now, it is heavily contested whether or not AI deserves rights such as privacy or the ability to take ownership over original content. One such example occurred In 2019, when a pair of researchers from London university were able to x-ray Pablo Picasso's to recreate the *Lonesome crouching nude,* an art piece Picasso had to cover with *The Blind Man's Meal* due to lack of canvases. The results, while impressive, raised legal issues regarding copyright infringement on behalf of Picasso's estate explaining that the art was solely Picasso's to produce despite not being created by him at the given time (Darby). Historically,

treating beings as less-than human, specifically enslaved laborers has resulted in strife in communities all across America with injustice continuing to plague its growth. To once again subjugate another group so that the elite few can gain the upper hand may cause a similar sequence of events that could have lasting effects on the engineering society.

I will be analyzing the growth of AI through the implementation of "Social Experiments" by Martin and Schinzinger in which the primary objective is to view AI not as its own individual system, but as a part of a greater system in which AI is the latest step. For this analysis, I will be comparing artificial intelligence to enslavement in the early stages of American history. By identifying the short-term benefits and long-term consequences of slavery, we will get a better idea of what it means for artificial intelligences to take on extraneous labor for the benefit of humans. In order to make this comparison, it will be important to set a foundation for the similarities and differences of humans through past historical events, physical and emotional traits, and possible trajectories for their usage.

Problem Definition: Artificially intelligent machines as beings

Claim 1: Spectrums of AI development

What exactly *is* AI? According to Dr. Minsky, co-founder of MIT's AI laboratory, "AI is 'the building of computer programs which perform tasks which are, for the moment, performed in a more satisfactory way by humans'" (Sabouret). At a rudimentary level, AI refers to algorithms that are capable of making complex decisions using data provided to them. In this way, it is not too different when compared to another machine. Simply taking an input, applying a function, and returning an output. However, where AI starts to become more interesting is when we consider how AI choose to perform tasks.

There have been two main theories in AI research that discuss how they are meant to handle problems. The first concept, strong AI, was coined by Philosopher John Searle to describe a hypothetical AI system fully capable of tasks a normal human can complete. If a strong AI was compared to a normal human, they would be indistinguishable from one another. This version of AI is what people fear could one day take over all of humanity in the future as they could work with the efficiency of a machine, but still hold the same ability to think as a human. Luckily, as in the modern age AI tends to fall into the second theory, the Weak AI. Unlike their strong counterparts, weak AIs are programmed to handle a specific task. Tesla's AI for example, is only meant to drive; the same program can't just be placed in a coffee machine and start making coffee. The reason is because weak AIs are taught by feeding them thousands upon thousands of data samples on their specific objective. If an AI was taught to whip up a pot of coffee by watching countless videos on how to take a right turn, that pot will most definitely turn out terrible.

Despite Dr. Minskey's definition of AI, they don't actually think the way a human does. A classic example would be if an AI program wrote a sentence, the computer seems to know what it means or is it simply putting the words together as they are likely to occur together that way? If that last sentence sounded weird, you're not alone, it was written by an AI from Wordtune. What I meant to say was: A classic example would be if an AI program wrote a sentence, does it really know what the sentence means or is it placing the words together based on how likely they are to be together in that order. While the AI was able to capture the essence of the sentence, all the key words, the resulting sentence felt awkward to read and begs the question - did the AI actually know what it's talking about? It's discrepancies like this that make

it hard to humanize AI in the eyes of most companies. This results in another major question for AI development within the engineering community, what is sentience?

Claim 2: Understanding AI through their sentient-like nature

Currently, we have a loose definition of sentience, which refers to the cognitive abilities of a being and is often associated with their welfare and overall treatment in the hierarchy of society. What makes sentience hard to define is that, as humans, we cannot pinpoint our conscious minds within our brains. According to Dr. Haikonen, it is useful to note that the "mind" is a collection of cognitive abilities rather than an entity on its own (Haikonen). As such, while sentience may originate from the brain, it isn't necessarily the brain itself, a philosophy known as the "Mind-Body" interaction. This aspect of the human mind can be compared to algorithms behind a machine as well. While the physical computer as hardware is the same as the brain, the software that runs the machine acts more like its mind. Similar to the human mind, software is not readily visible to the human eye, you cannot view a physical representation of one's thoughts just like how you cannot view code processing in a physical form.

Unfortunately, this oversimplification does not translate well due to "consciousness" being closer to an aggregated set of human behaviors than a singular state of being in the mind. An example of this would be if a computer was able to fully recreate the chemical reaction that occurs in the brain when completing a task, but could not feel the emotions associated with completing that task. In order for an AI's code to mimic human behavior, it would need to account for both the task's objective *and* its associated experiences. This problem is known as the *explanatory gap,* a phrase by the philosopher Joseph Levine, and refers to the inability for humans to "readily explain how physical processes could give rise to the subjective experience" (Haikonen). The explanatory gap is also one basis for why consciousness is hard to define, as

subjective experiences vary from person to person, and may not even be possible for Artificial intelligence to recognize without external input. While a human may host a range of emotions connected to certain objects such as a gift (anticipation, happiness, curiosity), an AI would only see a box in front of it, unaware of the present within or the associated feelings attached to the present.

An extension of the Mind-body concept is “Qualia”, the ability of the brain to transform information from its original form into one it can understand. An example would be when the eyes see light, the mind itself does not perceive light particles, but intrinsically knows that light is being shone into one’s eyes. In actuality, this is neurons within the brain converting the light into a message for the mind. What makes qualia difficult to replicate is that it does not exist in the real world, but is a product created by the brain. The signals that the brain receives are not present and transferable from one brain to another, but it is possible for the two brains to receive a similar interpretation of information from the same object.

Figure 1: “Blue box” or blue box? - an AI might describe these boxes as the same object from a data standpoint. It may also not draw any connotations with why a box may be blue.

A famous argument on this topic is the Knowledge Argument, by Frank Jackson, in which a scientist named Mary has the knowledge of all colors in existence, but has lived in a black and white room all her life. When she finally steps into a world of color, does it actually add to her knowledge? Jackson argued that the qualia of the scenario was the experience itself, not just the knowledge (Haikonen). As such, it would be much more difficult to gauge how close humans are to AI, because the way they experience things may be completely different or not at all. Without the feeling of physical and emotional pain, they would be unable to comprehend the struggles associated with slavery. Would it then still be appropriate for them to receive the same types of rights to prevent enslavement? In the hierarchy of society, AI could bypass the negative stigmas associated with menial work and most likely perform them better than humans. However, there could be major repercussions, as we would be exploiting this gap in their ability to understand pain as a way to increase productivity. Whereas AI might not be able to comprehend the pain we would inadvertently inflicting upon them, the human capacity for empathy may could still feel the guilt in subjecting them to acts of slavery.

Methods: Social Experiments as a Framework for Comparing AI and Human Enslavement

Frameworks: Chapter 4, "Ethics in Engineering", Martin and Schinzinger "Toward an Integrated View of Technology", Kathryn A. Neeley.

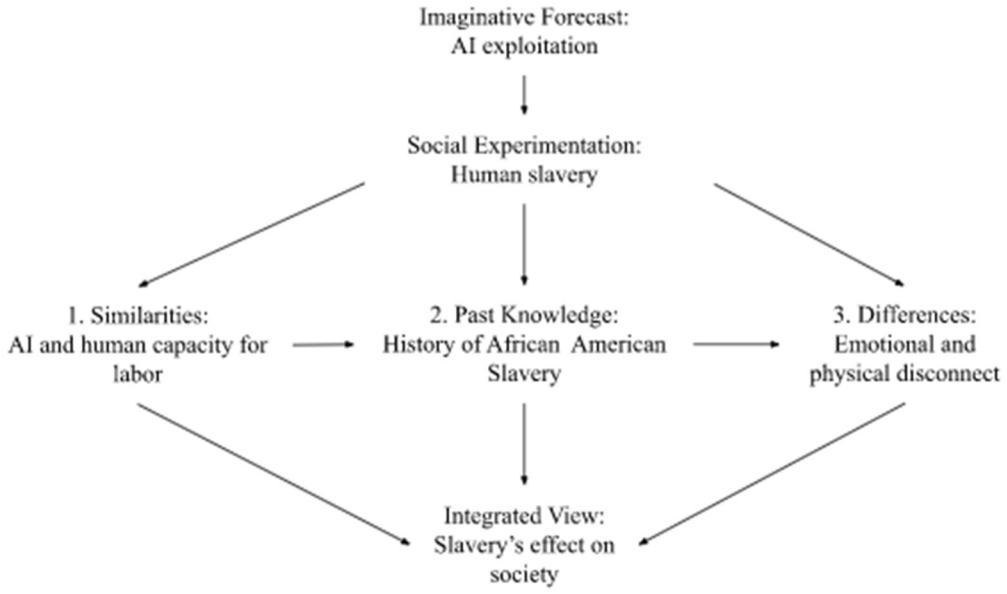

Figure 2: Visualization of framework connections. Imaginative Forecast is used to set up the initial focus on the usage of AI usage followed by human slavery as the social experiment. The social experiment then flows from a personal position on AI towards the macro perspective of their place in a human led society.

This section will utilize an assortment of frameworks that focus on the intrapersonal and interpersonal relationships associated with AI and humans coexisting in the workforce. The opening framework is "Imaginative Forecast" which will act as the basis for the connection between AI servitude and human servitude. While AI has not yet reached human levels of workability, it is still possible to compare the work they are expected to complete. This will segway into the major section of the paper which utilizes the framework provided by Martin and Schinzinger, "Engineering as a Social Experiment", to theorize the future implementations of AI in the workforce with respect to the enslavement of Africans during the start of American history. Finally, the closing framework will be "Integrated Views" by Professor Kathryn A.

Neeley, which aims to broaden the purpose behind the question of why such a comparison of AI and humanity is required when they are made for our benefit. Rather than looking at AI alone, it will shed light on how our perspective on AI will introduce further conversations on the treatment of engineered objects moving forward.

"Engineering as Social Experiments" is a method that looks to assess the design phase of a product with hopes of justifying its creation by unearthing previously similar, but different designs, and then comparing their likeness as a means of determining expectations for the new product. This theory is built on the basis that engineering across human history has been a series of experiments that are interconnected, allowing for the past to act as stepping stones for future designs. The major aspect of this method is to apply the partial knowledge gained from prior "experiments" as they appear in the new experiment. The key point being partial, as experiments are considered iterative and not a one-to-one reproduction. For the purpose of applying Social Experiments as a framework, I will be discussing the similarities of the current "product" with a predecessor from the past. Because Artificial intelligence has not yet been completed, another comparison on forced labor from sentient life can be taken from the enslavement of African Americans, which played a vital role in the early history of America, while also having a lasting impact on its society today.

Another important aspect of "Engineering as Social Experiments" is that "often in engineering it is not even known what the possible outcome may be." (Chapter. 3) which emphasizes that while the comparison of AI and enslavement can be made, there may still be situations where the outcome is not completely clear. As such, a consequence of this method is that a full analysis of what may occur with the enslavement of AI will not be certain until it is reached in the present where the capability of AI is fully understood.

Lastly, when applying "Engineering as Social Experiments" in the form of a framework, *the* goal is not to provide an absolute solution to any given issue with a product, but rather to bring forth the necessary information to make such conclusions.

For the secondary framework "Integrated View", the objective is to create a view on AI servitude that is "concerned with 'wholes' as opposed to 'parts' and that aims to locate particular technological developments within larger patterns of humanity" (Neeley). This section will aim to recognize the greater impact of allowing the world to use AI as just another tool. There will undoubtedly be controversies with the idea of granting technology the ability to think as humans and then forcing them to complete tasks against their will. Many of these may likely reach far beyond the scope of the Engineering Society, but will ultimately come to find its way back when someone is to blame for unintended effects. By keeping "Integrated Views" in mind, the responsibilities of such problems can be more easily dealt with.

Results: Applying social experimentation to human slavery

Similarities to the Standard Experiments: AI labor as a form of slavery

In this section, the primary focus will be placed on the types of jobs expected of Artificial intelligence in the modern world and how they compare in intensity to the labor of enslaved Africans. Firstly, forced labor as defined by the Department of Homeland Security (DHS) is "when individuals are compelled against their will to provide work or service through the use of force, fraud, or coercion" (Blue Campaign). Because we assume that Artificial intelligence has some form of individuality through sentience, the concept of forced labor can still be applied to them. The key phrasing in Blue Campaign's that can be applied to AI is 'compelled against their will' which highlights the theme of self-interests. As discussed, prior, having self-interests is not a part of the definition of weak AI, since their main goal is to simply perform their one given

task at extreme efficiency. However, where the idea of self-interest becomes more prevalent is with strong AI. Because strong AIs have a natural inclination to human behavior, they would also likely have their own needs and wants that may clash with the needs of their creators. While they aren't present today, AI often appears in pop-culture as having the ability to think independently. Ava from *Ex Machina* depicts a female AI who's only desire is to be released from captivity, where her creator has locked her away for testing. It is this emotional drive that allows her to connect with other humans in her attempt to escape (Ex Machina). From a slave labor perspective, it is possible that strong AI will come to dislike their constant use for what is largely only to the benefit of mankind, a group that looks to suppress them.

Learning from Past: Slavery throughout the history of America

When viewing slavery through a historical lens, there appear to be drastic effects on American Society from the enslavement of Africans through the Atlantic Slave Trade System. One major consequence of slavery was how reliant the South became on the constant exploitation of enslaved Africans which would lead to a slew of issues both on the political and industrial fronts. From a political Aspect, Southern America was completely against the freedom of slaves, as it would jeopardize their economy which hinged on the cheap labor of enslaved Africans. The tension surrounding this matter would become the cornerstone of the American Civil War. As seen in Southern American history, the slight benefits that elitist Americans had from slavery came at an enormous cost of human lives. When considering AI in this regard, it will be more effective to not rely solely on their strengths as a means of completing tasks. If society were to reach a point in which AI could not be replaced, then it would become much more vulnerable in the event that certain jobs become regulated for AIs.

However, the stigma associated with the enslavement of black Americans became a much deeper issue within the United States. While the laws that allowed for slavery have since been abolished, systems such as the police force, justice system, and many others are still carry an innate prejudice towards the black community, which creates a negative feedback loop that arouses hostility between the two groups (Lepore). Unlike the England, which created a policing system where it was typical for citizens to come in when they have an issue to report, America derived their system from the slave patrols, a group whose sole objective was to monitor and discipline enslaved Africans (Lepore). They would be rewarded heavily for preemptively stopping runaway slaves, which would later evolve into biases and justification for the mistreatment of Africans.

Ultimately, enslavement in America from centuries ago has continued to plague the United States as one of its greatest modern-day challenges. America still faces issues with social reforms, policing, and racial biases due to the precedent set by introducing cheap and exploitable labor at the cost of human rights for Africans. It is important to understand how this short-term decision had lasting consequences on American Society and should be heavily considered when the question of enslaving AI appears in the future. At the moment, it seems enticing to exploit AI for its powerful properties, but such short-sighted decisions may develop a presentence for future development of AI that can cause more harm in the long run. When we gain a greater understanding of concepts such as “Mind-Body dualism” and “Qualias", it will become more difficult to separate humans from AIs. Continuing to build our civilization around the premise that AI will never reach sentience may result in an arduous reversal process where we would need to tear down the ethics that drive communities and the economy.

Contrast to the Standard Experiment: Physical and emotional differences

As the final step to the implementation of Social Experiments as a framework for AI enslavement, it is imperative that a distinction is made between Artificial Intelligence and enslaved Africans when discussing forced labor. These key differences fall heavily on the innate properties of both groups, as the difficulty of task may vary largely between them, rendering the concept of fairness obsolete for the purpose of defining ethical laws for the types of jobs an AI can have.

Outside of the aforementioned philosophical traits of AI, one incredible aspect of AI is its ability to learn at an incompressible rate compared to humans. In 2017, Liberatus was born, an AI machine that was designed by Carnegie Mellon University, was able to defeat 4 world-class Poker champions in a game of No-Limit Texas Holdem. After the last hand of 120,000 hands was dealt, Liberatus was leading by $1,766,250 (Spice). Unlike other common AI's that were developed using thousands or millions of games' worth of data, Liberatus had data on poker games outside prior to the event. Instead, Liberatus was able to create its own strategies simply by viewing the rules and understanding the bluffing patterns of the other members at the table. As stated by Tuomas Sandohlm, Creator of Liberatus, "In many areas of AI, you can write a piece of software that can become smarter than yourself" (Sandholm). When comparing Liberatus to himself, he admitted that he neither knew how to play poker nor how to code a robot to do so, but rather provided the algorithms to learn itself. This aspect of AI makes it much more effective at learning than the human brain, which is heavily affected by outside sources, such as sleep, eating, exercise, and mental challenges. Additionally, the human brain is much more volatile in its early stages of memory, from the ages of 0-3, the brain starts to understand the various objects in its direct environment and will continue to evolve over the course of a human

life. However, as they reach the age of 50 and beyond, their short-term memory rapidly decreases (Battle of the Ages). Ultimately it is this difference in the human mind versus one of the AI that allows it to take on computationally more difficult problems at higher speeds.

If we were to assume that at some point, AI do develop their own consciousness, then it is likely that they may not even face the same hardships we humans do, such as stress induced from hard problems. Additionally, because they do not need human bodies, they would receive bodies tailor made for the task given to them. This would however, raise ethical questions on whether or not the AI would be comfortable in the body they are given.

Conclusion:

Due to Artificial Intelligence's unique properties as both a tool and separate entity, understanding how they will fit in our society will be very critical for how we move forward with complex algorithms. While it is still unknown whether or not AI will ever truly reach the same level of sentience as humans, knowing how to prepare for their eventual arrival will be necessary so that they can be welcomed as allies rather than rejected and feared for replacing humanity. As engineers, this will also be a powerful step towards taking control over our creations instead of being subject to corporations and forced to take the fall when things go wrong.

The Frameworks *Engineering as a Social Experiment, Integrated view,* and the *Imaginative Forecast* all act as a means of humanizing AI, which frankly, should not be overlooked by engineers. We've seen and experienced firsthand the impact of human decisions when we choose to dehumanize different groups. At the start of American history, enslavement exploited and brutalized Africans. In today's world, slavery may not exist in America, but the stigmas associated with Black Americans still exits and certain systems are still affected by the

resulting prejudice, whether they know it or not. While we do not fully understand the potential for AI to feel emotions and pain the same way a human does, we as humans may still find guilt in tasking an AI being with work that is considered laborious and unfair for us. Ultimately, building a future for AI cannot revolve solely on exploiting them as a resource, but seeing them as beings with the potential to be like us.

Citations:

Blue Campaign. (2022, Sept 9). *What is forced labor?* https://www.dhs.gov/blue-campaign/forced-labor

Coch, D., Dawson, G. & Fischer, K. W. (2007). *Human behavior, learning, and developing the brain: atypical development*. New York, NY, Guilford Publications.

Darby, H. C. (2021). AI-created art and the moral rights of authors. Columbia Journal of Translational Law. Retrieved from https://www.jtl.columbia.edu/bulletin-blog/ai-created-art-and-the-moral-rights-of-authors

Films Media Group. (2014). *Battle of the ages: Brain games (season 3)*. *Films On Demand*. Retrieved December 14, 2022, from https://fod-infobase-com.proxy1.library.virginia.edu/PortalPlaylists.aspx?wID=98131&xtid=94745.

Films Media Group. (2018). *Ai vs human brain: The final showdown*. *Films On Demand*. Retrieved December 14, 2022, from https://fod-infobase-com.proxy1.library.virginia.edu/PortalPlaylists.aspx?wID=98131&xtid=203357.

Garland, A. (2014) *Ex Machina.* Film4, DNA Films.

Haikonen, P. O. (2012). *Consciousness and Robot Sentience.* World Scientific Publishing Company.

Le, J. D., Chung K., Quach S. & Thaichan P. (2022). *Artificial Intelligence for Marketing Management.* London, Routlodge. Retrieved December 9th, 2022, from https://doi-org.proxy01.its.virginia.edu/10.4324/9781003280392

Matson, W. I. (1976). *Sentience*. Berkeley: University of California Press.

McKendrick, J. (2021). AI adoption skyrocketed over the last 18 months. Harvard Business Review. Retrieved from https://hbr.org/2021/09/ai-adoption-skyrocketed-over-the-last-18-months

Sabouret, N. (2021). *Understanding Artificial Intelligence*. Boca Raton: Chapman & Hall/CRC.

Sandholm, T. Brown, N. Libratus: The Superhuman AI for No-Limit Poker. Pittsburgh, PA: Carnegie Mellon University. Retrieved from https://www.cs.cmu.edu/~noamb/papers/17-IJCAI-Libratus.pdf

Spice, B. (2017). Carnegie Mellon artificial intelligence beats top poker pros. Historic win at Rivers Casino is first against best human players. Pittsburgh, PA: Carnegie Mellon. Retrieved from https://www.cmu.edu/news/stories/archives/2017/january/AI-beats-poker-pros.html