Whole Brain Emulation Ethics

Minnal Kunnan

October 10, 2016

Abstract

Advances in neuroscience, computer infrastructure, and neural networks have led to the hypothetical process of Whole Brain Emulation or mind uploading. This process would involve completely uploading one's consciousness onto some hardware running an artificial neural network. Organizations like IBM and the Blue Brain Project have been hard at work trying to make this theoretical process a reality. While technological breakthroughs continue to occur in brain emulation software, many ethical questions have not been addressed. One of these questions is whether or not the resultant emulated consciousness is the same consciousness of the uploader and whether this new consciousness should be offered similar rights. Another ethical question is if the simulated consciousness experiences time in a similar manner to biological humans and if its consciousness can feel pain or "suffer". In this paper, I will examine whether a concrete ethical consensus on Whole Brain Emulation is a necessary precursor to the conducting of brain emulation experiments. This is consistent with Software Engineering Code of Ethics tenet 3.03, which requires software engineers to consider the ethical, political, and economic implications of the technologies they create.

1 Facts

Whole Brain Emulation relies on a fundamental philosophical assumption about the mind and consciousness. It assumes a materialistic view of consciousness in which consciousness arises from various chemical reactions between neurotransmitters. [1] With this underlying assumption, it follows that with enough processing power we could theoretically simulate these interactions with a computer. Advancements in parallel computing, algorithms, quantum computing and Moore's law have futurists optimistic about the timeline for whole brain emulation. Ray Kurzweil, Director of Engineering at Google, believes that whole brain emulation will become a reality by 2029 [1]. There are, however, various degrees of emulation abstraction which begs the question what a true whole brain emulation is. A "functional emulation" considers an emulation successful if its behavior is the same as that which it is attempting to emulate.[2] IBM, The Blue Brain Project, and the Human Brain Project have all had relatively similar levels of success creating functional brain emulations. [5] While these companies have made huge leaps in innovating brain emulation technology they all have little to nothing published on the side of brain emulation ethics. The Human Brain Project does have a public ethics advisory board but information from their meetings is sparse and only tangentially related to brain emulation. [4]

2 Research Question

Are IBM, the Blue Brain Project and the Human Brain Project's attempts at brain emulation ethical in the absence of direct ethical reflection on their projects?

3 Social Implications

The social implications of whole brain emulation are almost limitless. The assumption that consciousness may exist independent of its substrate is itself highly contested. If whole brain emulation manages to prove "successful" it would grant us valuable insight into a meaningful definition of consciousness, perhaps solving the mind-body problem.[4] At the very least it would cause us to reframe the way we understand consciousness and create more meaningful metrics for defining it. Whole brain emulation also opens up the potential for "immortality" as a mind may now "live" under a modular substrate that may be repaired and replaced. [13] This raises the question of whether or not immortality should be something humans strive to create and if governments should consider regulating it. Other various laws would have to be reformed to accommodate brain emulation. For example, should governments extend all general human rights to brain emulations?[6] Should these emulations be considered minors or adults? Should a person be able to upload their consciousness as many times as they desire, or should there be a limit? Other metaphysical questions, separate from consciousness, such as the Qualia of brain emulations will also be brought into question.[4] Will emulations who/which have their processing speed upgraded experience time at a different rate? Will laws regarding retributive punishment such as prison time have to reflect this? While whole brain emulation remains a theoretical concept, experts in technology have no doubts about the drastic changes it will bring to our world. Prominent scientists and entrepreneurs such as Stephen Hawking and Elon Musk have already begun to warn the world about the opportunities and risks that advanced emulation would create. [3]

4 Other Arguments

4.1 Arguments Affirmative

4.1.1 Experimentation is the Solution

Although various ethical implications remain unexplored in regards to whole brain emulation, the argument can be made that we should not foreclose the ability for organizations to continue developing the technology just because of this. The ethicality of brain emulation may then be potentially resolved hand in hand with innovation or after we have a greater understanding of the technology. Experimentation may itself be the only way to resolve these ethical issues.[3]

4.1.2 Ethical Reflection Unnecessary

Current moral frameworks may already deal with the issue of whole brain emulation. If so, then an explicit ethical reflection of whole brain emulation may be unnecessary. The issue of whole brain emulation might be a philosophical debate so convoluted that there might never be a more concrete answer than what is currently known. It might be a question on the scale of other philosophical conundrums such as "What is the meaning of life?" or "Where did we come from?" If this is the case, the only reason to not pursue the study of brain emulation is if it possesses an existential level risk to humanity.[14]

4.2 Arguments Negative

4.2.1 Emulations are Moral Agents

One primary objection to Whole Brain Emulation research is the belief that emulations should be treated as conscious and therefore may have the capacity to experience suffering. If this is the case then even primitive emulations may have some degree of sentience. Then we must weigh different ethical systems to determine if this suffering is justified or not.[6]

4.2.2 Technology with Unknown Consequences

Some singularity doomsday scenarios may be put into motion by successful Whole Brain Emulation. This

may be brought on by a brain emulation that recognizes a way to enhance its intelligence and does so over and over again or simply copies itself millions of times.[15]

5 How the SE Code Applies

5.1 Definition of Software Engineers

The Software Engineering Code of Ethics defines a software engineer as a person "who contributes by direct participation...to the development of software...in accordance with their commitment to the health, safety, and welfare of the public". Are whole brain emulation programmers acting in accordance with their commitment to the safety of the public by exploring brain emulation projects? [16]

5.2 SE Code Tenet

We can narrow our question further by using SE code tenet 3.03 which states that software engineers should "Identify, define and address ethical, economic, cultural, legal and environmental issues related to work projects." [16]

5.3 Domain Specific Rule

Our question now specifically is the following: have brain emulation projects identified, defined and addressed ethical, economic, cultural, legal and environmental issues related to their work?

6 Analysis

In order to answer our research question, we must first conduct a thorough investigation of the issues related to whole brain emulation.

6.1 Moral Status of Simple Emulations

The first question to consider is whether ethical considerations should apply to current emulation attempts. However, judging conscious states and the sentience of a being is not a simple task.

6.1.1 Uncertainty of Machine Consciousness

Various academics have posited extensive theories on what would constitute a machine conscious. An early model was introduced in the 1970s by Robert Freitas Jr. who created the "Sentience Quotient". This simple formulation of sentience simply takes the information processing rate of the being and compares it to the size and number of processing units. Using this scale, Humans and IBM's Watson have roughly the same degree of sentience. David Gamez suggests a probability scale for the likelihood of a machine to be attributed mental states. [6] Gamez goes on to write that it may be impossible to determine whether a machine is conscious or not.

Current emulations which are fairly primitive under the above theories have also come under scrutiny. Imagine a scenario where a researcher asks "We want to genetically engineer mentally retarded infants. We need infants with cognitive deficiencies to study their behavior." This would be a fairly appalling idea under most ethical frameworks, yet Thomas Metzinger warns this is exactly what might be happening when machines satisfy a sufficient number of qualities for conscious experience. [8]

The question of whether or not a machine is truly conscious may also be reduced to the 'hard problem of consciousness': "The hard problem of consciousness is the problem of explaining why any physical state is conscious rather than non-conscious." [9] This problem, defined by David Chalmers, states that this issue may persist even after emulated behavior begins to match that of its origination. Whether or not the hard problem of consciousness will continue to persist after the functional behavior is emulated is controversial. This alone demonstrates the uncertainty behind what a true machine consciousness may look like.

6.1.2 Principle of Assuming the Most

Because of much of the uncertainty surrounding software emulations, it would be prudent to use a very cautious approach. The safest assumption would be to assume that any emulation could have the same mental capacity as the system it is trying to emulate. Using this view it would be necessary to treat the emulations with the same ethical consideration as their originals. [10]

To meet this requirement the emulations would have to be consistent with general animal testing standards, requiring minimization of suffering and improving the quality of life for the emulations. Reducing suffering may be easier in this instance as we might be able to virtually block nociceptors that induce pain and create things like the "perfect painkiller". Improving the quality of life of animals would entail maintaining a habitable virtual environment for the emulation which is not difficult to do. [10]

6.2 Existential Risks

Sandberg and Bostrom's roadmap to whole brain emulation suggests that it may be feasible by mid-century.[13] Of the various singularity oriented risks, whole brain emulation poses some extremely unique ones. Bostrom defines an existential risk as "one where an adverse outcome would either annihilate Earth-originating life or permanently and drastically curtail its potential". [5]

6.2.1 Arms Race

As whole brain emulation becomes less and less of a pipe dream many nations may invest significant amounts of resources to be the first to reach this new technology. The first nations to create a complete whole brain emulation would have a distinct advantage over other

nations similar to how industrialized nations managed to colonize less "advanced" nations. This is due to the fact that brain emulations may be so beneficial to economies in which they are created that other nations would need to create their own just to keep up and not become dependent financially. The nature of brain emulations is similar to resource arms races that already exist, however, the exponential nature of these emulations would create conflicts on a much larger magnitude and without warning. [5]

6.2.2 Superorganisms

The creation of a sufficiently advanced brain emulation may quickly lead to the creation of Superorganisms. This is partially due to an emulation's substrate circuitry hardware, which is replaceable, repairable, and even improvable. This is equivalent to increasing a computer's clock speed and creating clones of this computer. This duplication technique may even be applied to individual tasks. Emulations may keep a track of their own productivity and find optimal moments during their emulation where they complete a respective task incredibly quickly. These states may then be saved and reused when that task may need to be completed again. [11] This, combined with the ability for emulations to create clones of themselves allows for unlimited potential and the ability for recursive self-improvement. [7]

6.3 Ethical Framework

To conduct a thorough analysis of the above ethical issues in regards to brain emulation we must also establish a moral framework to judge them on. We can do this, not by creating an entirely new framework but by elaborating on previously defined frameworks. While explicit ethical reflection on software projects seems like a prudent step, it is at times unnecessary. For example, it would be unfair to require programmers who are building a simple calculator to create a system of ethics justifying why their project is in accordance with public safety and health. In cases where a normative moral framework may already apply we can apply Armin Grunwald's argument that "technology development free from ethical reflection could use a provisionary moral framework to continue the development process without explicit ethical reflect" [13]

We can apply Grunwald's framework with the knowledge that the creation of brain emulations would likely result in an intelligence explosion [14]. While the results of such an explosion are still up for debate we know that the world will be forever changed after the fact. Because of this, a techno-pessimist approach to the issue would be prudent as we may then be able to more accurately predict whether the creation of whole brain emulations will be a benefit or detriment to society. [15]

6.4 Application of SE Code to Moral Status of Current Emulations

The moral status of current emulations is a more pressing issue in regards to brain emulation since it is already occurring. Since there is currently no standard for emulation ethics and no laws in place to enforce it, it is the responsibility of private companies themselves to maintain an ethical approach to brain emulations.

6.4.1 Identify

Brain emulation projects themselves have not attempted to identify issues related to whole brain emulation, however, much progress on this issue has been made in the field of academia.[6] The case can be made that since these issues have been identified and information about them has been released into the public sphere, the companies themselves need not reinvent the wheel. We can then apply Grunwald's argument that a sufficient provisionary moral framework exists for the development of brain emulation technology without the need for identifying ethical issues surrounding it.

6.4.2 Define

The definition of ethical issues differs from their identification depending on the context in which these ethical issues are being explored. While, in an abstract sense, ethical issues about the moral status of emulations have been identified, they have not been discretely defined under specific contexts. For example, IBM has not attempted to apply the Principle of Assuming the Most to their emulation of a cat's brain or if they have that information is not publicly available. Not much information is available on the ethical considerations of companies attempting brain emulation. Only the Human Brain Project has a listed ethics advisory board and their standard operating procedure does not go into detail about the ethical concerns they consider. Their compliance management standard operating procedure lists two ethical requirements that are tangentially related to whole brain emulation, research on animals and "other ethical issues".[3]

Without the evidence of any consideration involving brain emulation ethics, we cannot assume that these companies have done their due diligence in providing a sufficient quality of life and minimization of suffering for their emulations. Considering that emulations continue to increase in complexity at an incredibly rapid rate, this lack of ethical inspection is troubling.

6.4.3 Address

There are no complete solutions to the ethical dilemmas surrounding brain emulation. In fact, the controversy over testing brain emulations is a rather new topic and currently proposed solutions are wholly theoretical and not considered public practice anywhere. [11] Considering that emulations of animal brains have already been "successful" we are approaching, if not already in, an ethical gray area. [6] The worst case

scenario is that we are enacting Thomas Metzinger's warning that implementing partially successful brain emulations is similar to genetically engineering mentally retarded infants. The best case scenario is that the ethical issues of brain emulation turn out to be unfounded and the testing of brain emulations can be justified ex-post facto. Both options are unfavorable and present a clear violation of the ethical framework laid out above.

6.5 Application of SE Code to Existential Risks of Whole Brain Emulation

The urgency of companies to consider the existential risks of whole brain emulation is noticeably less than that of the current concerns over the moral status of emulations. That does not, however, deem it a topic not worth exploring. Ethical concerns in technology have been notably justified ex post facto over the course of history. [15] This has been demonstrably the case with nuclear and chemical weapons. Unprecedented technology is usually invented before restrictions for them are applied.

6.5.1 Identify

While the academic world has a lot to say on existential scenarios for humanity, this does not excuse the organizations themselves for not developing their own consistent system of ethics for their projects. This is especially important as the organizations themselves will be on the frontier of developing this technology since the main limitation is massive processing power. Unfortunately, these issues seem to have been deemed irrelevant, obscure or a problem for future generations to deal with. [15]

The intelligence explosion that may follow the creation of whole brain emulation make this an object of dire importance as it will be near impossible to retroactively apply restrictions to such a rapidly growing invention. Organizations developing whole brain emulation should at bare minimum have a set of procedures or plans of action in case of these risks years before the potential chance for these risks. Currently, there exists no documentation by any of the projects listed above on these potential scenarios. [7]

6.5.2 Define

Defining the existential risks of progressing towards whole brain emulation without ethical consideration is also not a simple task although there are some metrics in place. There is a concept of a hard versus soft takeoff in terms of the intelligence explosion where in the soft takeoff scenario a generalized AI or brain emulation self-improves at a "human-like" pace. This takeoff would be preferred to the hard takeoff where a superintelligent machine improves "too quickly for significant human-initiated error correction or for gradual tuning of the machine's goals". [5] Defining exactly how each scenario might occur under one project may provide valuable information into avoiding unwanted scenarios. Currently, there is no documentation of these projects conducting this ethical investigation.

6.5.3 Address

Since the advent of a superintelligent machine is not likely for the next few decades there are no ways to address concerns related to them as of yet. [7]

7 Conclusion

Based on the arguments above it is not ethical for brain emulation projects to continue their experiments until they identify, define and address the ethical concerns listed above. While in some instances the projects may not be in direct contradiction to the Software Engineering Code of Ethics, they are treading a fine line with the ethicality of their projects. As emulations become increasingly more complex the unattended ethical consideration of these emulations as "conscious" or "sentient" is a form of negligence on the part of the organizations conducting the research. While we are still in the process of researching brain emulation, the experimentation alone is enough to violate ethical issues recognized by experts in the field. The high magnitude impact of potential future scenarios in regards to brain emulation may also be sufficient to force organizations to consider these scenarios seriously. These concerns may be mitigated due to the fact that they lie in the distant future; however, the tendency to value progress over consequences, both intended and unintended, particularly in the field of technology, may be a dangerous path to take.

References

1. "KurzweilAI — Accelerating Intelligence." KurzweilAI Well Be Uploading Our Entire Minds to Computers by 2045 and Our Bodies Will Be Replaced by Machines within 90 Years Google Expert Claims Comments. Web. 22 Apr. 2016.

   Kurzweil estimates the time it will take for brain emulation to become a reality. Includes a broad survey of whole brain emulation, its philosophical assumptions and an exploration of consciousness.

2. Sandberg, Anders. "Feasibility of Whole Brain Emulation." Studies in Applied Philosophy, Epistemology and Rational Ethics Philosophy and Theory of Artificial Intelligence (2013): 251-64. Web.

   Discusses roadblocks, concerns, and breakthroughs in regards to whole brain emulation. Provides a broad view of the issue including ethical dilemmas that develop from it.

3. Ethics Advisory Board. Retrieved May 27, 2016, from https://www.humanbrainproject.eu/ethics-advisory-board

   Human Brain Project Ethics Advisory Board. Has information on the team's standard operating procedure and frequently asked questions.

4. Robinson, H. (2003). Dualism. Retrieved May 27, 2016, from http://plato.stanford.edu/entries/dualism/

   Stanford University Encyclopedia of Philosophy entry on the mind body problem. Discusses the potential for computer simulated consciousness to shed light on the issue.

5. Ward, M. (2015). Hawking, Musk warn of 'inevitable' AI arms race. Retrieved May 27, 2016, from http://www.cnbc.com/2015/07/27/hawking-musk-warn-of-inevitable-ai-arms-race.html

   Discussion from academics on the risks of an intelligence explosion and the AI arms race that leads up to it.

6. "Are Brain Emulations People Too?" Inverse. N.p., 21 Apr. 2016. Web. 22 Apr. 2016.

   Analysis of the differences between brain emulations and humans as moral agents.

7. Bostrom, Nick. Superintelligence: Paths, Dangers, Strategies. Print.

   Defines existential risks to humanity in the context of technology. Exploration of why brain emulation may push us closer to these risks.

8. David Gamez, Progress in machine consciousness, Consciousness and Cognition, Volume 17, Issue 3, September 2008, Pages 887-910, ISSN 1053-8100, http://dx.doi.org/10.1016/j.concog.2007.04.005. Keywords: Machine consciousness; Artificial consciousness; Artificial intelligence; Models of consciousness; Synthetic phenomenology

   Older paper on the discussion of various metrics in reference to artificial intelligence and consciousness.

9. Metzinger, Thomas (2003). Being No One. Cambridge Massachusetts: The MIT Press Philosophical argumentation on whether or not experimentation may constitute an ethical violation.

10. David Chalmers. The Conscious Mind: In Search of a Fundamental Theory (1996). Oxford University Press

    The conception of the Hard Problem of Consciousness. Details why consciousness may be illusory or inexplicable and why it matters in regards to emulated consciousness.

11. Sandberg, Anders. "Ethics of Brain Emulations." Risks of Artificial Intelligence (2015): 213-42. Web.

    Ethical overview of ethics in brain emulation and artificial intelligence. Explains how superorganisms may come into existence and what that would mean for humanity.

12. Shulman, Carl. "Whole Brain Emulation and the Evolution of ... - MIRI." Machine Intelligence Research Institute. Web. 4 May 2016.

    Discussion of current attempts in brain emulation. Applies the principle of assuming the most to consider the potential "sentience" of a brain emulation

13. Sandberg, Anders, and Nick Bostrom. "Whole Brain Emulation: A Roadmap - University of Oxford." Future of Humanity Institute, 2008. Web. 4 May 2016.

    Detailed roadmap to the creation of brain emulation. Includes parts list, algorithms, hardware dependencies, biotechnology integrations and ethical analysis.

14. Grunwald, Armin. "Against Over-estimating the Role of Ethics in Technology Development." Science and Engineering Ethics SCI ENG ETHICS 6.2 (2000): 181-96. Web.

    Lays the groundwork for using a provisionary moral framework to continue software development in cases where explicit ethical reflection is unnecessary.

15. Agar, Why Techno-Optimism Is Dangerous. Retrieved May 25, 2016, from http://www.huffingtonpost.com/ni agar/techno-optimism-dangerous_b_7746682.html

    Analysis of techno optimism in the context of artificial intelligence.

16. IEEE, Association for Computing Machinery. Retrieved May 27, 2016, from http://www.acm.org/about/se-code

    IEEE ACM Software Engineering code of ethics. Creates the framework for the analysis in this paper