OpenWorm - So, can I turn it off?

Seminar - Scientists and Ethics

Emil Suleymanov <emil.suleymanov@tum.de>
Department of Informatics
Technical University of Munich

August 4, 2020

Abstract

With the development of Artificial Intelligence (AI), the day it reaches the level of human intelligence might be achieved if not this decade then perhaps this century [11]. This makes ethical questions regarding said intelligence ever more pressing. In this paper, we try to gain insight into how an AI is perceived and how it should be treated. We start by exploring OpenWorm - a virtual organism on a computer.

1 Introduction

In the past decade, the world has seen great development in the field of machine learning, with systems like AlphaGo [6] surpassing human-level intelligence, given, in a very narrow set of tasks, such as playing Go or various Atari 2600 games at a super-human level, however, with an ever-increasing number and complexity of such tasks.

GPT-3, a recently released language model by OpenAI showed stunning results on many Natural Language Processing (NLP) data-sets, translation, answering questions, using a novel word in a sentence, and more tasks requiring on-the-fly reasoning. This model is a great leap forward towards an all-encompassing AI, covering several "land-masses" as seen on the illustration of the AI's capacity from 2017 (figure 1).

It is estimated that the human brain contains at least 100 trillion or as high as 1000 trillion synapses [3], which are channels connecting individual neurons in the human brain. Parameters or weights in an artificial network could be thought of as an analog of synapses in a biological neural network. GPT-3 contains 175 billion parameters, which took $3.14 \cdot 10^{23}$ flops to train [8]. A reasonable cost estimate for training this network based on the price of Tesla V100 cloud instance is $4.6 million [1]. Assuming that cost of training of a larger network grows linearly with the growth in parameters, we can estimate that a GPT-3 network with 100 trillion synapses could be trained for $2.6 billion dollars

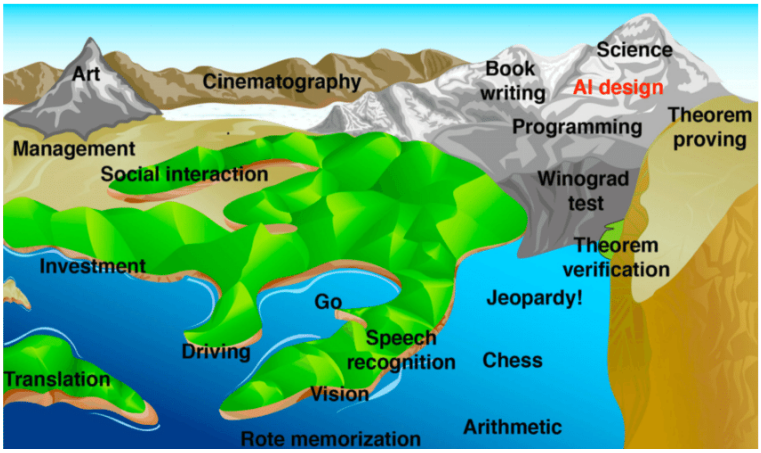

Figure 1: "Hans Moravecs illustration of the rising tide of the AI capacity. Max Tegmark (2017)." Figure from [11].

today. Considering that the efficiency of neural networks has been doubling every 16 months for the past decade and assuming that this trend will continue, we can roughly estimate that the price of training a GPT-3-like network with the same number of parameters as synapses in a human brain (10^{16}) could cost a mere \$80'000 in 20 years [9]. *This calculation is by no means accurate and serves only as a rough estimate.*

With a human-level AI becoming so accessible, several ethical questions can be raised. Is a life of an AI more or less valuable than that of a biological one? Is artificial life - life, and does this question make any sense in the first place? How should an advanced artificial life be treated?

To try answering these questions, I propose considering OpenWorm - a project dedicated to creating a virtual living organism.

2 OpenWorm

As the project is defined on its web-page, it is an open-source project dedicated to creating the first virtual organism on a computer [2]. Current objective of this project is creating a computational model of *Caenorhabditis elegans* (C. elegans) - a microscopic roundworm with 302 neurons, 95 muscle cells, 959 cells in total (see figure 2).

The choice of C. elegans as the target organism is backed by its relative simplicity and a great understanding of this organism. The most important aspect is an almost fully understood neural network of this worm, which enables its simulation. In its current state, the project can only simulate the basic locomotive function of the worm in a virtual environment. This shows that creating a convincing simulation of a living organism or just a nervous system of a living organism is not impossible.

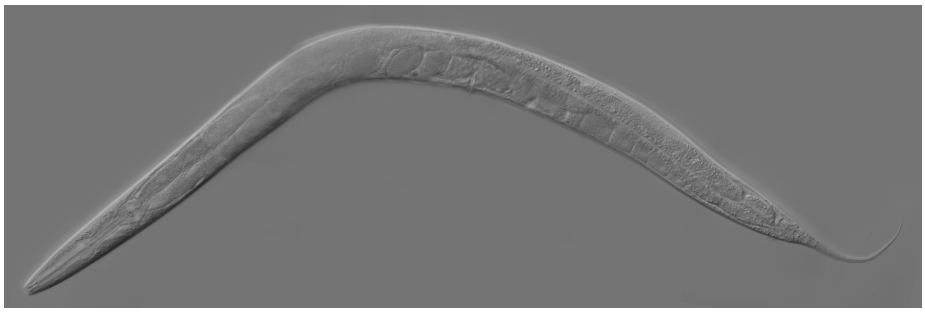

Figure 2: Adult Caenorhabditis elegans. Figure from [7].

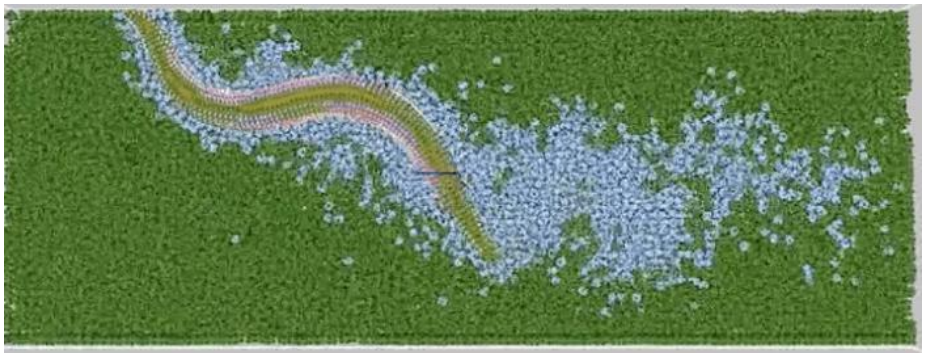

Figure 3: A snapshot of a simulation of C.elegans. Figure from [2].

OpenWorm consists of several separate parts (see figure 4). Let us go through some of them:

Sibernetic is a fluid mechanics simulator using Smoothed Particle Hydrodynamics (SPH) algorithm. This part of the project deals with simulating the body and the environment of the virtual worm.

Geppetto Simulation Engine is an open-source modular platform that enables an interactive simulation of biological systems.

Optimization engine is used to fill in the gaps in the knowledge of the electrophysiology of neurons and muscle cells of C. elegans, by using optimization techniques like genetic algorithms.

Movement validation project is a way of validating the model by comparing it to the real worm data.

Muscle-Neuron Integration project aims to create biologically realistic models for neurons and muscles by creating realistic models for cellular ion-channels.

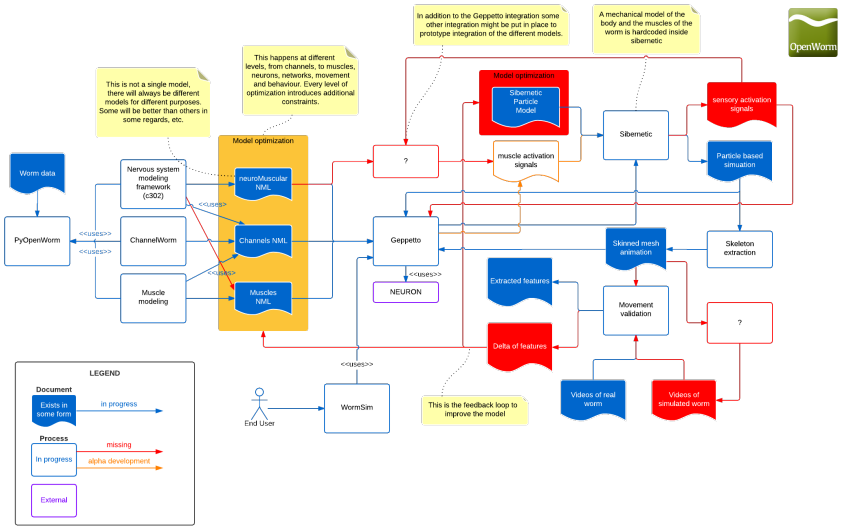

Figure 4: Structure of OpenWorm. Figure from [2].

3 Is it alive?

Is this simulated worm alive? To answer this question, first, we need to define what alive means. In figure 5 you can see a chart with key requirements for a biological definition of alive.

Assuming that this virtual C. elegans behaves in-line with the biological one, we could expect these requirements to be met as the project progresses. With these assumptions, if the virtual worm shows these behaviors, should it be considered alive? This question can be approached from two points of view:

Outer-world perspective. From the outside world, this is just a simulation that mimics life. The virtual worm is not storing any real energy, but merely a programmed representation of it.

Inner-world perspective. From the inside of the simulation, this organism obeys the laws of physics in its virtual world and therefore meets the requirements for being alive. The programmed representation of energy is as real to it as energy is to us in our reality.

With that in mind, one could ask if it is "morally wrong" to kill or turn off this worm and its simulated universe? Or, is it more "wrong" to kill the virtual worm, than the biological one?

Given that most people would have a hard time sympathizing with a worm, and that it could be argued that simulation of the environment and the full body of a virtual life-form is not at the core of the issue, let us consider a simulation of a human nervous system on par with OpenWorm, upscaled from 302 to 86

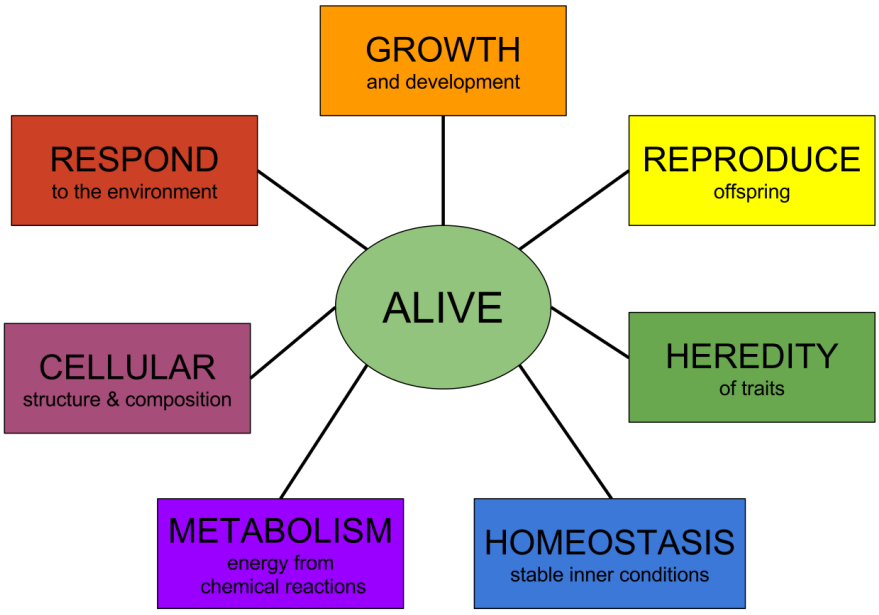

Figure 5: Characteristics of life. Figure from [10].

billion neurons of a human brain [3]. Should it be treated as a human? Should it be treated with the same ethical concerns as a human? The answer to these questions depends on your life stance.

We consider the following ethics based on historical relevance and progression between them:

- Abrahamic religions
- Humanism
- Sentientism

3.1 Abrahamic religions

Abrahamic religions include Christianity, Judaism, Islam, and many more religions, and have the largest cumulative following among religions with 55.5 percent of the global population [4]. As summarized on Wikipedia, the Abrahamic religions believe in judging, paternal, fully external god to which the individual and nature are subordinate [4]. This puts any artificial life-form or intelligence into the category of man-made things, which diminishes their right for ethical consideration.

Given that the source of value in Abrahamic religions can be considered God, or rather God's creations, let us explore if the following creatures deserve ethical considerations within this ethical system:

- A biological worm?
  - Not so much.
- An artificial worm?
  - Probably no, since it is man-made.
- A biological human?
  - Yes, if certain parameters are matched.
- An artificial human?
  - Probably no, for the same reason as an artificial worm - it is man-made.

A curious argument against ethical considerations for human-level AI from a religious point is that only God can create "souls" which are supposed to be essential to being a human. Alan Turing gives an interesting remark on the issue, stating that whatever the mechanism for creating "souls" in biological humans is, it can work as well in artificial systems [12].

3.2 Humanism

Humanism, as defined by the International Humanist and Ethical Union, advocates that human beings have the right and responsibility to give meaning and shape to their own lives. One could argue that humanism is at the core of modern societies.

Humanism provides two major points in defense of an artificial life form. It values human life, which may encompass artificial humans as well, depending on the definition of a human being. And is discourages mysticism, which leaves us with just our physical world in which our biological brains reside, functioning within the laws of physics, which theoretically enables the existence of a human being in a simulation.

Considering that the source of value is human beings having freedom over their lives, let us explore if the following creatures deserve ethical considerations within this ethical system:

- A biological worm?
  - No, since it is not a human being.
- An artificial worm?
  - No, since it is not a human being.
- A biological human?
  - Yes.
- An artificial human?
  - Probably no, unless considered human.

3.3 Sentientism

Sentientism, as defined on Wikipedia, advocates that all sentient beings deserve moral consideration [5]. Sentience, a core concept to sentientism, is defined on Webster as the capacity to feel, perceive, or experience subjectively. One could think of sentientism as an extension of humanism. Even though this ethics originally was tailored towards animal rights, it is the best hope for an artificial life-form or an AI.

Considering that the source of value is sentience, let us explore if the following creatures deserve ethical considerations within this ethical system:

- A biological worm?
  - Yes to an extent.
- An artificial worm?
  - Yes to an extent, if the sentience can be verified.
- A biological human?
  - Yes.
- An artificial human?
  - Yes, if the sentience can be verified.

An issue with the sentientism is that, for example, people in coma do not fall under the category of protected beings, since they are unconscious, and do not experience anything subjectively. This is why it should be used as an extension of humanism.

3.4 Ethics summary

These ethics are just one of many, and even though they are not perfect for our case, they provide us with valuable insight into the issues of ethics surrounding artificial lifeforms and AI.

Humanism is a good starting point but it has too great of an emphasis on biological humans. Sentientism appears to be the best hope for artificial life, but there is an issue with the verification of sentience. And Abrahamic religions are outdated for answering these questions.

4 Future/Conclusion

Simulation of a human nervous system could be equated to the artificial general intelligence, which, as shown at the beginning of this paper, could be achieved rather sooner than later, considering the rapid development in the field of machine learning. This makes ethical concerns ever more pressing. We cannot

simply ignore these questions because they are hard to deal with, and an objection that this is far enough in the future not to think about it is unrealistic.

In the end, the question we should be asking ourselves is if we want slaves, partners, or masters.

References

[1] Openai's gpt-3 language model: A technical overview. https://lambdalabs.com/blog/demystifying-gpt-3/.

[2] Openworm: Building the first digital life form. open source. http://openworm.org/.

[3] Scale of the human brain. https://aiimpacts.org/scale-of-the-human-brain.

[4] Wikipedia: Abrahamic religions. https://en.wikipedia.org/wiki/Abrahamic_religions.

[5] Wikipedia: Sentientism. https://en.wikipedia.org/wiki/Sentientism.

[6] Alphago is the first computer program to defeat a professional human go player, the first to defeat a go world champion, and is arguably the strongest go player in history. https://deepmind.com/research/case-studies/alphago-the-story-so-far, 2017.

[7] Zeynep F. Altun. Adult caenorhabditis elegans. https://en.wikipedia.org/wiki/Caenorhabditis_elegans#/media/File:Adult_Caenorhabditis_elegans.jpg.

[8] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.

[9] Danny Hernandez and Tom B. Brown. Measuring the algorithmic efficiency of neural networks, 2020.

[10] Chris Packard / CC BY-SA (https://creativecommons.org/licenses/by sa/4.0). Characteristics of life. https://commons.wikimedia.org/wiki/File:Characteristics_of_life.svg.

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

[12] Alan Turing. Computing machinery and intelligence, 1950.