What is a machine? Exploring the meaning of 'artificial' in 'artificial intelligence'

STEFAN SCHULZ[1]
AND
JANNA HASTINGS[2]

[1] Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz
[2] Institute for Implementation Science in Health Care, Faculty of Medicine, University of Zurich/School of Medicine, University of St. Gallen/Swiss Institute of Bioinformatics/Centre for Behaviour Change, University College London

Abstract: Landgrebe and Smith provide an argument for the impossibility of artificial general intelligence based on the limits of simulating complex systems. However, their argument presupposes a very contemporary vision of artificial intelligence as a model trained on data to produce an algorithm executable in a modern digital computing system. The present contribution explores what it means to be artificial. Current artificial intelligence approaches on modern computing systems are not the only conceivable way in which artificial intelligence technology might be created. If there are conceivable routes by which an artificial intelligence might be developed that are not constrained in the same way as the current generation of artificial neural networks, then these might be plausible routes towards the engineered generation of artificial general intelligence that are not precluded by the Landgrebe and Smith argument against the possibility of artificial general intelligence.

INTRODUCTION

Landgrebe and Smith, in their recent book *Why Machines Will Never Rule The World: Artificial Intelligence Without Fear* (Landgrebe and Smith, 2022) provide an argument for the impossibility of artificial general intelligence (AGI) based on the limits of mathematical simulation. They present both an argument for why current digital systems will not serve as a basis for the development of AGI, and an argument for why it will *never* be possible to develop AGI.

The L&S argument against the emergence of AGI within the current computing paradigm (including with approaches based on neural networks) can be summarised broadly as follows (our formulation):

I. Everything that runs on a digital computer must be represented as a mathematical function, whether explicitly programmed or learned from data.
II. However, intelligence, as an emergent feature of a complex system, cannot be approximated by a mathematical function.
III. Therefore, intelligence cannot be simulated on a digital computer.

This argument takes aim at those who claim that modern technologies such as language models will soon have, or already have, the foundations of AGI. For example, a team at

Microsoft Research recently published an article claiming that the large multimodal model GPT-4 shows ‘sparks’ of AGI (Bubeck et al. 2023).

We agree with L&S that the only kinds of things that can run on digital computers as we know them are mathematical functions, whether designed or learned from data, and that this is not sufficient to replicate all aspects of human intelligence, as it is not exactly describable or predictable by a mathematical function. Mathematical functions can only approximate, but not truly reproduce the behaviour of complex systems. Thus, while digital computers are able to mimic or approximate many of the behaviours of intelligent systems (“weak AI”), they are not able to emulate human intelligence (“strong AI”, often equated with “AGI”).

However, we will leave this argument aside in the remainder of this contribution.

The L&S argument for the *impossibility* of developing AGI *in any form* in the future is the one to which we will speak in this contribution. This can be summarised broadly as follows (again, our formulation):

I. Complex systems cannot be engineered (that is, they cannot be recreated as machines)
II. Importantly, here, engineered, i.e. being a ‘machine’, appears to mean for L&S, systems for which:
   1. the outputs are known and intended by the engineers,
   2. some mathematical model of the behaviour exists to predict and verify the behaviour in any given situation (they often refer to this need for prediction and verification for safe use of a technology)
III. Intelligence is an emergent feature of a complex system that cannot be represented in an engineered, mathematically predictable machine
IV. Therefore machines will never possess intelligence, ergo, there will be no AGI.

In broad outline, this argument seems plausible. As the many examples in the book illustrate, complex systems indeed have many features that preclude their formulation as mathematical models, and physicists have long been familiar with the inherent challenges with prediction of the behaviour of even very (relatively) simple complex systems.

However, is there perhaps a danger that L&S are drawing too narrow a picture of what form AGI may take with their clause (II), and therefore arguing with a straw man? In what follows, we aim to raise this possibility.

What does it mean to be ‘artificial’?

Let us for a moment imagine that in a previous cosmological epoch, somewhere in our galaxy, a civilization existed that had developed on a purely inorganic basis (according to (Cooper et al. 2011)properties, and composition (see picture and (Barge et al. 2015), inorganic life is not inconceivable). Over a long period of time, following an evolutionary process, this civilization had achieved a high level of cultural and scientific complexity similar to ours. Just as human civilization learned to use metals and ceramics for creating artefacts and gradually developed new materials through chemical synthesis, that remote civilization discovered the potential of organic chemistry and eventually created numerous new artefacts based on carbon compounds. Their engineers developed molecules with interesting properties, including the potential of information encoding by polynucleotides, the biosynthesis of proteins, and the creation of lipid-enclosed microcompartments in which such reactions could autonomously evolve. Thus, that remote civilization had come to the technological excellence of engineering self-replicating microsystems in the form of simple prokaryotes.

In short, a remote civilization produced artefacts, i.e. microscopical machines capable of self-replication and self-modification, as the ultimate achievement of their scientific and technological progress.

The further fate of that remote civilization is of minor interest. The decisive factor for our short narrative is that the prokaryotes manufactured by them might have persisted through space and time, embedded in frozen matter, accidentally reaching our planet Earth. Here, they might have become the foundation of biological and intelligent life as we know it today.

So far, this has been a simple science fiction story, and we want to make it clear that we consider this scenario highly improbable, but nevertheless feasible. We are not interested in proposing any new narrative about the origins of life, but just want to take it as a thought experiment encompassing the following components:

- A1: A complex civilization (C1), exhibiting the characteristics of intelligent beings
- A2: That civilization was successful in engineering artefacts ("machines")
- A3: Certain artefacts were capable of self-replication
- A4: They persisted through time and space, and were able to self-replicate, leading to evolution
- A5: They turned into seeds of a second civilization (C2)
- A6: This new civilization developed entities with intelligent characteristics ("mind-body continua" according to L&S), *viz.* Homo Sapiens.

In this counterfactual scenario, one could claim that "machines rule the world" under the following assumptions:

- It cannot be excluded that there are other intelligent, conscious entities like homo sapiens somewhere in the universe (A1).
- These intelligent agents have the capability to engineer "machines", i.e. complex material or immaterial artefacts for certain purposes under their control (A2).
- This also includes "machines" with evolutionary potentials (A3, A4)
- "Ruling" could be seen as referring to some significant control, influence, or power (A5)
- "World" (A6) means a certain portion of reality

Then the hypothesis that "machines will never rule the world" could already be refuted by this hypothetical counterexample under the premise that machines or subunits thereof can undergo evolution. It is likely that L&S would disagree with this possibility, however, to better understand the point of contention requires us to elaborate on the ontological nature of machines.

That "machine" is not a well-defined term becomes obvious when comparing existing definitions. Depending on which sources are referred to, machines must have moving parts, do a particular type of work, use a powered engine, and / or transform energy. Others require that machines be manufactured, others, again include "molecular machines" that are the products of human engineering design but may be manufactured by living processes such as genetically modified bacteria. In computer science, "Turing machine" and "machine learning" refer to a particular meaning of "machine", which contradicts some of the mechanistic definitions, as digital processors do not have moving parts. L&S obviously refer to "machines" in that latter sense when stating that "machines with a general cognitive performance even at the level of vertebrates such as crows" (Foreword x) are infeasible. And they refer frequently to "engineering", which we assume to be synonymous to "creating an artefact" (A2), set forth by intelligent beings (A1). Unfortunately, the key terms "machine", "artificial" and "engineering" are missing in their glossary.

Engineering pursues a goal. An artefact can have different purposes, one of which might be the capability of self-replication (A3, A4). Self-replicating machines have mostly been discussed in space engineering contexts, harnessing 3-D printing as a key technology (Abdel-Rahman et al. 2022; Ellery 2016). Clearly, even artefacts such as these that have the capability of self-replication would still be artefacts. The question is now how to classify the next generations of these entities, those that are themselves created by machines. Are they still artefacts, insofar as they resemble those of the first generation? If we accept this, then there is no reason to consider them and subsequent descendants any less as engineered artefacts than the products of a fully automated assembly line. But what about a scenario in which further generations gradually undergo a Darwinian process of mutation and selection? Would they still count as machines even after diversification into entities of different kinds, analogous to biological speciation? This is by no means inconceivable: in the abovementioned discourse for applications to space engineering, potential evolutionary divergence is seen as a risk to be prevented in evolutionary machines (Ellery 2022). And it is noteworthy that the boundary between engineering and natural science is continually being eroded, such as is evidenced for example by the development of ‘dish brain’, a system of around 800,000 human neuronal cells in a dish that were induced via artificial feedback to learn to play the computer game ‘pong’ (Kagan et al. 2022; Milford et al. 2023).

Our elaborations show that at the very least, both “artefact” and “machine” require better definitions than those provided in L&S. Without an ontological clarification, any discourse about evolutionary and potentially intelligent machines would remain vague. Therefore, we here propose definitions of artefact and machine. Regarding the definition of intelligence, we subscribe to section 3.2 in L&S.

“Artefact”:

> Definition: An artefact is a material or informational entity that results from an engineering process. The agent of this production process is an intelligent being that executes a plan specification pursuing a specific goal.
>
> Elucidation: the engineering process may be simple but also supported by other material entities (tools) or informational entities. The outcome can be highly complex, such as a spacecraft, a symphony, or a CAD software application, or it can also be simple like a dugout canoe, a lullaby, a copper alloy or “Hello World” in Java.

“Machine”:

> Definition: A machine is a material entity made up of a multitude of specialised parts, each with a specific function. Machines can run or stand still. When they run, they consume energy. A machine serves a specific purpose. A machine is an artefact or has evolved from an artefact.
>
> Elucidation: Machines may incorporate informational entities such as algorithms.

“Evolutionary Machine”:

> Definition: A machine that is the offspring of a lineage of machines that resulted from an evolutionary process, i.e. involving replication, mutation and survival in a given environment.
>
> Elucidation: Whether a given entity is an evolutionary machine or not depends on the first step, i.e. whether is an artefact. Under the hypothesis of the above thought experiment (i.e. that intelligent beings can be traced back to an act of engineering by another intelligent being), but equally under a doctrine that denies evolution (e.g. creationism), humans would be machines, according to this definition.

What does this mean for human civilization? If it were possible to create artefacts, based on the current technology but with the ability to develop, replicate autonomously and evolve due to evolutionary pressure, which would still count as machines by the above definition, then we could not rule out the possibility that machines would ever rule the world.

It may be difficult to imagine that the production of integrated circuits and the production of materials that are necessary to manufacture them can be completely taken over by machines. It equally requires some imaginative power to consider how such machines might reach autonomy in their environments and develop the potential for self-replication, which would then lead to Darwinian mutation and selection cycles in which the fittest machines survive and eventually, as a consequence of evolutionary pressure, may even acquire qualities that can be compared to human intelligence.

The low probability that this will occur, and the long time it would take to reach a level of complexity that enables intelligent behaviour to evolve, does not concern us at present: insofar as it is possible at all, then L&S have not succeeded in showing that AGI is impossible.

CONCLUSIONS

We are broadly in agreement with L&S that an emulation of a complex system such as the human brain, created artificially and with intent as an entire engineered system, is not feasible. However, in our opinion this is not the only way intelligence can emerge in the future. The fact that our own brains are the result of an evolutionary process shows that intelligence can evolve in response to selection without any explicit design. Yet, what we humans create has the potential to join the same natural world, and insofar as we are able to create artefacts with evolutionary potentials, we lay the foundations for evolutionary processes to operate under similar selection pressures, in ways that can be anticipated or even planned.

For the future, we should therefore be wary of drawing any too simplistic boundaries between what is natural and what is a machine. For an artificial system to be intelligent means almost by definition that it will not be predictable or mathematically describable. Yet that does not mean that creating such a system is impossible. It only means that such a system, if we do manage to create it, will not fit the L&S definition of 'machine'.

REFERENCES

Abdel-Rahman, A., Cameron, C., Jenett, B., Smith, M., and Gershenfeld, N. 2022. Self-replicating hierarchical modular robotic swarms. *Communications Engineering*, 1(1): Article 1. https://doi.org/10.1038/s44172-022-00034-3

Barge, L. M., Cardoso, S. S. S., Cartwright, J. H. E., Cooper, G. J. T., Cronin, L., De Wit, A., Doloboff, I. J., Escribano, B., Goldstein, R. E., Haudin, F., Jones, D. E. H., Mackay, A. L., Maselko, J., Pagano, J. J., Pantaleone, J., Russell, M. J., Sainz-Díaz, C. I., Steinbock, O., Stone, D. A., … Thomas, N. L. 2015. From Chemical Gardens to Chemobrionics. *Chemical Reviews*, 115(16):8652-8703. https://doi.org/10.1021/acs.chemrev.5b00014

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., and Zhang, Y. 2023. *Sparks of Artificial General Intelligence: Early experiments with GPT-4* (arXiv:2303.12712). arXiv. https://doi.org/10.48550/arXiv.2303.12712

Cooper, G. J. T., Kitson, P. J., Winter, R., Zagnoni, M., Long, D.-L., and Cronin, L. 2011. Modular Redox-Active Inorganic Chemical Cells: iCHELLs. *Angewandte Chemie International Edition*, 50(44):10373-10376. https://doi.org/10.1002/anie.201105068

Ellery, A. 2016. Are Self-Replicating Machines Feasible? *Journal of Spacecraft and Rockets*, *53*(2):317-327. https://doi.org/10.2514/1.A33409

________. Curbing the fruitfulness of self-replicating machines. *International Journal of Astrobiology*, 21(4):243-259. https://doi.org/10.1017/S1473550422000246

Kagan, B. J., Kitchen, A. C., Tran, N. T., Habibollahi, F., Khajehnejad, M., Parker, B. J., Bhat, A., Rollo, B., Razi, A., and Friston, K. J. 2022. In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron*, 110(23):3952-3969.e8. https://doi.org/10.1016/j.neuron.2022.09.001

Landgrebe, J., and Smith, B. 2022. *Why Machines Will Never Rule the World: AI without Fear*. New York and Abingdon: Routledge.

Milford, S. R., Shaw, D., and Starke, G. 2023. Playing Brains: The Ethical Challenges Posed by Silicon Sentience and Hybrid Intelligence in DishBrain. *Science and Engineering Ethics*, 29(6):38. https://doi.org/10.1007/s11948-023-00457-x