Are Large Language Models "alive"?

Francesco Maria De Collibus[1]

[1]University of Zurich, francesco.decollibus@business.uzh.ch

Abstract

The appearance of openly accessible Artificial Intelligence Applications such as Large Language Models, nowadays capable of almost human-level performances in complex reasoning tasks had a tremendous impact on public opinion. Are we going to be "replaced" by the machines? Or - even worse - "ruled" by them? The behavior of these systems is so advanced they might almost appear "alive" to end users, and there have been claims about these programs being "sentient". Since many of our relationships of power and social dominance are apparently based on intelligence, where the "smarter" seems to prevail on the "dumber" (as we see for example with livestock farming, or with the distribution of salaries according to formal and intellectual qualifications etc.) the emergence of programs and AI applications of this level of sophistication brings us to the question "what will AI do to us?". Our error is to frame Artificial Intelligence in the same way we frame Natural Intelligence and mistakenly address it as we would do with biological life, even applying to it the same biopolitical categories. We here try to reverse the common arguments in relationship between life and intelligence and show that our progresses in Artificial Intelligence might have no immediate consequence for the topic of life as such, as life and intelligence are well separated issues, and - even when we would like to somehow fuse them together, we are very far away from having an "Artificial Life" to the degree of complexity and effectiveness we have with "Artificial Intelligence".

Artificial Intelligence, AI, Large Language Models, LLM, Life, Existence, Sentience, Biopolitics:

1 Are we alive?

We are alive, and this should never be taken for granted, as it is not so easy to prove that we are alive: as a matter of fact, we can doubt about everything, including our own existence. A classical argument is the one from Descartes: do we exist? We can doubt about everything, but not about the very fact that we are doubting. This is the rock bottom of doubt, and since our doubt is a manifestation of logical thought, Descartes could come with the famous sentence *Cogito ergo sum*, "I think, therefore I am". The topic has been explored and analyzed in any possible aspect by Philosophy, but until very recently the question if machines or computational systems could be "sentient" or "alive" was mostly a fascinating plot element for science fiction, or at best a frontier research topic. Despite enthusiasm and optimism no real system could be really deemed as "sentient" or even close to be capable of human like reasoning, and we know that the topic has been lingering since the very beginning of modern Computer Science, at least since Turing's "imitation game". Interestingly enough, already Turing came to the conclusion that it's almost impossible to come to a clear definition of "thought", valid for humans and machines, so he wrote:

> It was suggested tentatively that the question, 'Can machines think?' should be replaced by 'Are there imaginable digital computers which would do well in the imitation game?' If we wish we can make this superficially more general and ask 'Are there discrete state machines which would do well in the imitation game? [1]

It turns out now we have such machines. The recent - and groundbreaking - progress of Machine Learning (ML) and Artificial Intelligence (AI) driven by the Deep Learning approach [4] has improved the quality of results reached and the capacity of the model to generalize and scale, but, still, for the end users it just looked like as a better solution to an optimization problem. Sure, the progress has been impressive, sure, everybody expected consequences on the job market in the future, but our *natural intelligence* still plays on a league of its own, right? Deep Learning helped evolving Natural Language Processing (NLP) from the early, unsatisfactory approaches into an incredible effective discipline. This happened thanks to seminal papers with new approaches such as [3] and [6]. Recently ([9] such language models were scaled with more and more parameters, well deserving the definition of Large Language Models (LLM), crossing beyond 175 billion parameters (the exact numbers are not always disclosed by the vendors). When certain thresholds are crossed, some emergent properties [5] arise in the Language Models, such as, for example, basic arithmetic, logical reasoning, chain of thought. It depends as well not only from the sheer number of parameters, but from the quality and variety of data used and from the specific optimization techniques, that helped these Large Language Models achieve amazing and previously unseen abilities. This was made evident by ChatGPT, which was the first LLM of this size and power to be publicly available. Many of the end users are impressed: these models seem to reason "as they were really thinking, as they were *alive*". So we should rephrase Descartes into *"Cogitant ergo sunt*, they think, therefore they are? But admitted they think, can "they" be the same way "we" are, can "they" be alive?

Intelligence and life seems to be somehow connected in our perception. It might per part of our *Pareidolia*, we tend to see human-like feature in object, as they were alive, as they were human. This make very good material for entertainment, from Pinocchio to modern science fiction, where puppets or robots seem to desire nothing more than being alive just like us, creatures of flesh and blood.

On the other hand there are many examples of beings that are alive and do not stand out as particular intelligent. The mosquito bumping on my window, the house dust mite in the sheets of my bed appear both quite alive, even though they probably did never doubt about their own existence, not the way Descartes did. But we cannot know for sure, they do not bother telling us: again, they are not *intelligent* enough to communicate, for sure not in the way Large Language Models do. The spider on the corner will not talk to me in Japanese or write computer code on the fly, but it can catch the fly, thanks to a very complex web I could never weave myself. So there is indeed some "intelligence" in the arthropods mentioned, they are able to react to external and internal stimuli and to interact in a complex environment: that's at the end most of what life is about. Even the viruses, sitting somewhere at the frontier between life and pure protein-based machine, react to a complex environment. They react like proteins but reproduce and spread like living beings: classifying them is somehow troublesome. Every life form can or cannot react to certain events, some dictated by the environment, other dictated by the surrounding complex social environment, or by its reproductive behavior. We apparently live in the same world with them: the spider spins his web on the corner of my balcony, the dog waits at the door outside my kitchen, but we don't perceive the same world. The world where the spiders are living is most likely made of vibrations, their sight is reported to be quite poor (except for *salticidae*, or jumping spiders, which incidentally are the smartest or most cognitively advanced [17] group of spiders). Dogs - as we all know - can smell way better than we humans do. Maybe on the branch of a tree outside the house we could see, by getting very close, a tick. Perhaps is the same tick described by Jacob von Uexküll. The tick is blind and deaf,

and only react to three stimuli: the heat of the body of a large mammal, the butyric acid spreading out of its skin, and the sense of motion after having landed on that skin. That tick can survive up to 18 years on that tree's branch waiting for a large mammal with the right scent to pass by. Then the tick lets itself fall on it, lays its egg, and dies. As Agamben [18], and Heidegger before him we could ask ourselves if we really share the same world, the same *Umwelt* with that tick, a life form with such an alien way of interacting with the environment and such a impervious understanding of time? Our worlds somehow overlap: the tick might attempt to fall on our skin and bite us as we walk under its tree, or we might squash the tick, if we notice its attempted bite, but not much more than that.

2 Evolution and intelligence

We don't know if we share a world now, but we did share one in the past. We know as a fact that evolution happened, and lifeforms adapted to complex ecosystems, ecosystems varying in time, reacting to mutating geological, climatic and chemical situation in the atmosphere. Our ancestors and the ancestor of the tick, did share *a world*, possibly because it was the same ancestor [2], millions and millions of years ago. Sure it was a very different world, a very different *Umwelt*.

The way life beings interact with this environment might be seen as determining its "intelligence", where intelligence was never per-se a thriving factor, rather adaptation to the ecosystem was. Modern humans have existed for thousand of years before the neolithic revolution (about 12000 b.C.) where agriculture and animal husbandry allowed mankind to thrive. It can be said that intelligence became a thriving factor only when it started to actively shape the environment and the ecosystem around, as in the *Neolithic Revolution* [7].

We know the risks and the limitations of defining intelligence in this way, as we take here for granted the daring "null hypothesis" of Macphail [10], that no real difference in intelligence among vertebrates exist, and extend it even further, stretching it to include all other life forms as well, because our focus lies in the interaction between life and intelligence. So life-beings evolved a complex cognitive adaptation to the ecosystem which we came to define as "intelligence" as it appeared to us Homo Sapiens. But this adaptation does not automatically constitute "life", because the status of being alive predates our adaption in terms of intelligence. The cognitive capability is an expedient life forms evolved to stay alive, not the other way around. Biology showed us empirically that life forms can evolve their adaptive behaviors to a level that can be defined "intelligence", so intelligence - or cognitive adaptation - is a tool for life, the other way around - moving from pure intelligence to life - is a direction that has yet to be proven. It might be possible, but it would be a different journey. We know from Paleontology how life forms adapted in time and from Paleoanthropology - how the complex adaptive behaviour of human beings changed in time. We could even stretch further the thesis that biological life evolved not only natural, but artificial intelligence as well, so even AI can ultimately deemed a "biological" derived construct. Still we have no idea how sheer intelligence or disembodied complex cognitive capabilities could obtain life.

The topic of disembodiment has been intensively debated in Philosophy and Psychology. The thought experiment known as "brain in a vat" could be described as a human brain being artificially suspended in a jar, where it is provided with all kinds of neuronal stimuli he would have experienced were it still in the skull of a human body. So the brain could think "I am eating ice cream" and be convinced of having ice cream, feeling its flavour and its cold while still being inside the tank, as it feels exactly what it would have felt in the "real" world. It's not easy to prove we don't live in a simulation: the argument of Putnam [11] is that the simulated world and the real world are not really the same, even if they appear so at first glance. We would

be living in two different worlds. This is as well one of the starting points of David Chalmers [12] article "Could a Large Language Model be Conscious?". The author starts by quoting Thomas Nagel: "We have no idea how it feels for the bat being a bat". So we have not really an idea on how LLMs perceive the world, even admitted they can "perceive" something. Its - purely hypothetical -feeling of the world, would be radically different from our own. The argument of Putnam still considered the brain in a vat to be functioning exactly like ours own, except for the sensorial part, while the LLMs have a very different inner working and therefore would perceive a different world in a rather different way, more similar perhaps to the one of the tick than to the one of already troublesome human brain in a vat. Despite everything, could they obtain life? Ruth Millikan [15] developed the concept of *Biosemantics*, to investigate how living beings interact with intentions and representations, but still she is assuming biological existence, which right now is not at all provided in such AI applications. Are we speaking then about sort of "brains in a vat", do we have at least minds here? Philosophy of Mind is a fundamental topic in contemporary philosophy (we could cite here just as a starting point the work of Jaegwon Kim ([16]). The problem is that we now have some advanced cognitive capability, but we can't yet really talk about mind the way we humans do. So LLMs have no biological existence and apparently no mind either: what they exactly have?

3 Large Language Models

Researchers in the Natural Language Processing (NLP) field, with the evolution of Deep Learning technology, and the increase of parameters of model, have been able to produce systems that seems to talk and reason almost like human beings. Still it made a sensation when Blake Lemoine, an engineer at Google, announced that the Large Language Models he was working on, LAMDA, was "sentient". Probably that specific LLM was not: the engineer was later fired from Google, and there are conflicting points of view in terms of research (see again Chalmers [12] for a complete overview). Emily Bender and Timnit Gebru ([13]) state that these Large Language Models behave as *stochastic parrots*, simply echoing what we expect to hear from them, even making up an apparently random answer when there is not enough knowledge to provide a meaningful one (hence the fascinating "hallucination" problem, that seems to originate from a problem with the methodology for training the summarization capability as explained in [19]). We cannot exclude that - by raising further the number of parameters up or by some yet unknown emergent phenomenon or by finding a new optimization method we'll get to even better results, mitigating this "parrot" issue. Already GPT-4 does way far less mistakes than the 3.5 based ChatGPT and seems to get way better results with conceptualization in reasoning tasks, achieving what it appears as a "mental" model of the discussed topic. Some researchers claim GPT-4 has already reached at least some of the skills associated with Artificial General Intelligence. [20]. The road ahead - as usual for science and research - is very difficult to foresee, especially for empirical, statistic-based disciplines such as Machine Learning (ML). With the current approach we will eventually hit a limit of some sort, as we already hit limits in the past: we don't know yet where the limits are, and if or how these hypothetical problems will be solved. Let's leave aside the "sentience" topic, let's oversimplify by assuming it is a further evolution of intelligence (which we defined as a complex adaptive behaviour) when the complex adaptive behaviour is capable of conceptualizing itself. Since we have a working biological brain, which is made of matter, we have to assume that the phenomenon of our human intelligence can be somehow imitated. The only way we can exclude this possibility is by assuming some metaphysical reason preventing it, which we cannot neither exclude nor confirm. Surprisingly even Alan Turing spent some interesting pages discussing the metaphysical and religious aspects of thinking machines [1], coming to the interesting conclusion that we are we not in the position

of limiting the power of the Almighty to put thoughts and maybe even souls inside machines, if that pleases him.

4 What life means for machines?

We came to the conclusion that life, intelligence and their relationship require different analyses. If computing systems can achieve some of the hallmarks of intelligence, what about life? Computer Science and Biology have experienced a very favorable cross-pollination. There are many field like robotics that study the the interaction of machines in complex environments. John Holland introduced the Agent Based Modelling in his book "Complex Adaptive Behaviours" [8]. Genetic Algorithms achieved remarkable results. Conway's game of life inspired generations of computer scientists since its firs appearance in 1970. But the life we can simulate in this way is very basic. There are discussions about our entire world being a simulation, a thesis that is present in many different traditions of philosophy around the world and was recently made once again popular by movies such as *The Matrix* or by the three theses of Nick Bostrom. Could we live entirely in a simulation? We've already discussed the question regarding the "Brain in a vat" topic from [11]. Some results from Quantum Physics seem to deny this ([14], but for our scale of observations, it does not really matter if we are in a simulation. We care to know if - in the current reality or simulation - we can simulate (or sub-simulate) something that reaches the status of life. Therefore we postulate that the layer of simulation does not play a role, only the final results do, i.e. quality and level of complexity achieved. With the increase of number of parameters in Large Language Models we started at least to get to an order of magnitude of what human brain is capable of, we can model - for example - non linearity and threshold activation. Sure what is running on computing machines remains a numerical simulation of reasoning, which is happening on our brain on the fairly different basis of electrochemical signals, so we cannot say the two phenomena are really equivalent, but we assume a good approximation could be reached.

Life on the other aspects is a very complex phenomenon, perhaps even more complex than intelligence itself. On one hand the relationship life has with entropy, order and chaos and on the other hand what has been defined by Arthur Schopenhauer as *Lebenswille* as a basic property of living form, the struggle for life, to persist against all odds in the current, transient, state of life. The complex organization of multi cellular organisms, where life as a whole coexist with the life of the single cells composing the whole is of extraordinary level of complexity. Can we simulate all this? Projects like Open Worm [21] are trying such a simulation and still struggling to model a nematode worm, *Caenorhabditis elegans* in all its complexity.

So if we made good progress in simulating intelligence - or complex cognitive behavior - we are still lagging behind in simulating life. Sure, we can define "game of life" or "agent based models" with some parameters modelling life and ecosystems, we can set a boolean variable "alive" to True or False, but we are far from the level of complexity found in real life forms.

This again does not mean this simulation is impossible. But it makes no sense to ask ourselves if any Artificial Intelligence project is alive if we have not defined in a similar level of complexity what Artificial Life can really be. The same applies to robots, which even with their remarkable abilities and stunning development are still orders of magnitude away from the complexity of biological life. The level of sophistication of current actuators and sensors is simply not there yet. Sure, we can simulate some parameters, but we are not yet to the point where it makes sense to ask if those systems are alive. They are not, and will not be in the foreseeable future.

Is at least the current software architecture of Large Language Models somehow peculiar? What we know of current LLM model architecture points to a default Machine Learning project, powerful and expensive indeed, since the incredible user load

OpenAI servers had to endure, but still based on the usual paradigm of a client-server application or a normal stand-alone application when locally hosted, as some of the available models allow. The Large Language Models run in a machine, can be turned off, restarted, disabled, without being "alive" or "dead" at all. Synthetic biology and artificial life as a topic are making progress, so it could be that a Large Language Models, way far more advanced than the current ones, residing in this kind of artificial life environments could reach the hallmarks of life. When this happens, it might very well be that these new artificial life forms will behave with their own goals, goals that might be more comparable to the one of the Elephant than to the evil AI trying to rule the world and/or disrupt our society. An Elephant with its mass could easily crush a human being, and sometime - by accident or when provoked it does crush people, but most of the time the elephant just minds his own business trying to get water, food and the other usual resources for life, being it natural or artificial. That's still the most likely scenario for LLMs (or any other advanced enough ML or AI applications) that receive an advanced enough level of artificial life. Such artificial life which might very well be of a level of complexity pushing forward the limit of what we define as *life*. Until the field of Artificial Life has made such progress, all these applications need to be considered simply as a tool, and nothing more than that.

5 Conclusion

While addressing the question about the status of life of the Artificial Intelligence applications known as Large Language Models we proposed the following theses:

- ◇ Life and intelligence are two separated topics
- ◇ Biological life evolved intelligence as a complex adaptive behavioral response to the evolving challenge of environment, which is the decisive challenge for life beings.
- ◇ Up to now there are no evidence of intelligence being able to create life at will: even assuming biotechnological manipulations, this possibility has yet to be fully proven.
- ◇ Artificial (or simulated) Life is possible, but the currently level obtained are very, very far from the level of complexity reached by biological organisms.

So our conclusion is: no Large Language Model or any other Artificial Intelligence application is "alive" or can be deemed "alive" by our current standards and understanding of it. Assuming so through oversimplifying conditions for what life is would be incorrect and unjustified. To even ask ourselves if AI applications can be alive we should deeply advance the field of Artificial Life. Only when Artificial Life as a sector reaches a level of complexity comparable to what our understanding of life is we can ask ourselves if those AI applications could somehow be "alive". This come not by limit of computer science or robotics, but from our own gnoseological limits of human beings. We are alive, we know the world with our *kantian* structures. Getting back to the question of Agamben and Thomas Nagel, we can only know what humans know, and humans will always perceive the world as humans, even when trying to "impersonate" a different entity, be it a spider, a bat, a tick, or a Large Language Model. For us being alive will always mean something comparable to our own experience, and that's the only way we can frame this question.

It's very possible that for alien beings the current machines are already "living" in some way. But in our experience as human beings - the only one that currently matters - they are not and they will not be for the foreseeable future.

Acknowledgement

The author would like to thank Massimiliano Lorenzo Cappuccio, Raffaele Mauro and Amirehsan Sarabadani Tafreshi for their help in reviewing the draft of this paper and contributing to its improvement with their valuable input.

References

[1] Turing, A.M., Computing Machinery and Intelligence, Mind, Volume LIX, Issue 236, October 1950, Pages 433–460, https://doi.org/10.1093/mind/LIX.236.433

[2] Evans, S., Hughes, I., Gehling, J. & Mary L. Droser Discovery of the oldest bilaterian from the Ediacaran of South Australia. *Proceedings Of The National Academy Of Sciences*. 117, 7845-7850 (2020), https://www.pnas.org/doi/abs/10.1073/pnas.2001045117

[3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L. & Polosukhin, I. Attention Is All You Need. *CoRR*. abs/1706.03762 (2017), http://arxiv.org/abs/1706.03762

[4] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, 521, 436-444. http://dx.doi.org/10.1038/nature14539

[5] Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E., Hashimoto, T., Vinyals, O., Liang, P., Dean, J. & Fedus, W. Emergent Abilities of Large Language Models. (2022) https://doi.org/10.48550/arXiv.2206.07682

[6] Radford, A., Narasimhan, K., Salimans, T. & Sutskever, I. Improving language understanding by generative pre-training. (2018 https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

[7] Bocquet-Appel,J.P. ,When the World's Population Took Off: The Springboard of the Neolithic Demographic Transition.Science 333,560-561(2011). https://doi.org/10.1126/science.1208880

[8] Holland, J. Adaptation in Natural and Artificial Systems. (University of Michigan Press,1975), second edition, 1992

[9] Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Ngo, H., Niebles, J., Sellitto, M., Sakhaee, E., Shoham, Y., Clark, J. & Perrault, R. The AI Index 2022 Annual Report. (2022) https://arxiv.org/abs/2205.03468

[10] Macphail, E. M., Barlow, H. B., Vertebrate Intelligence: The Null Hypothesis [and Discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences, 308(1135), 37–51 (1985). https://doi.org/10.1098/rstb.1985.0008

[11] Putnam, H. (1992), "Brains in a Vat", in K. DeRose and T.A. Warfield (eds.), Skepticism: a Contemporary Reader (Oxford: Oxford University Press). https://philosophy.as.uky.edu/sites/default/files/Brains%20in%20a%20Vat%20-%20Hilary%20Putnam.pdf

[12] Chalmers, D. Could a Large Language Model be Conscious?. (2023), https://doi.org/10.48550/arXiv.2303.07103

[13] Bender, E., Gebru, T., McMillan-Major, A. & Shmitchell, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . *Proceedings Of The 2021 ACM Conference On Fairness, Accountability, And Transparency*. pp. 610-623 (2021), https://doi.org/10.1145/3442188.3445922

[14] Campbell, T., Owhadi, H., Sauvageau, J. & Watkinson, D. On testing the simulation theory. (2017) https://doi.org/10.48550/arXiv.1703.00058

[15] Millikan, R. Biosemantics. *Journal Of Philosophy*. **86**, 281-297 (1989) https://doi.org/10.2307/2027123

[16] Kim, J. (2011). Philosophy of Mind (3rd ed.). Routledge. ttps://doi.org/10.4324/9780429494857

[17] Cross, F., Carvell, G., Jackson, R. & Grace, R. Arthropod Intelligence? The Case for Portia. *Frontiers In Psychology*. **11** (2020), https://www.frontiersin.org/articles/10.3389/fpsyg.2020.568049

[18] Agamben, G., Attel, K. (Translator), The Open: Man and Animal *Meridian, Crossing Estethics*. (2003)

[19] Maynez, J., Narayan, S., Bohnet, B. & McDonald, R. On Faithfulness and Factuality in Abstractive Summarization. (2020) https://arxiv.org/abs/2005.00661

[20] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. & Zhang, Y. Sparks of Artificial General Intelligence: Early experiments with GPT-4. (2023) https://arxiv.org/abs/2303.12712

[21] The Open Worm Project https://openworm.org/science.html