Back to Evolutionary Intelligence: Reading Landgrebe and Smith

KIRILL KRINKIN

Abstract: This article is a response to the position of Landgrebe and Smith on the fundamental limitations that prevent the creation of general artificial intelligence (AGI), expressed in their book *Why Machines Will Never Rule the World*. The reasons for failures for attempts to create AGI using formal logic and algorithmic approaches to modeling intelligence are discussed. An attempt is made to define the future direction of intellectual systems development as hybrid evolving systems, as well as a revision of the Turing test statement and language models role.

Despite the fluctuating optimism over many years regarding the creation of artificial intelligence at a human or superhuman level, science and technology still seem to be stagnating and do not appear to have made any significant strides towards this objective.

In their recent book *Why Machines Will Never Rule the World*, Landgrebe and Smith attempt to explain the impossibility of creating what is known as AGI based on current mathematical and computer models and technologies. Their argument relies on the thesis that modem science does not have the capacity to model complex systems, such as humans, their bodies, and their activities. They accurately point out the source of the misconception that has been exploited for many years by scientists and engineers unsuccessfully trying to create the artificial phenomenon of intelligence. The source of this misconception is Descartes' assumption (Descartes 1641) that the mind (in our case, intelligence) can be considered separately from the body, implicitly assuming that cognitive abilities can be reduced to algorithmic manipulation of symbols in a completely abstract symbolic system. Thus, Descartes moved a human being, (more precisely, human's intellectual abilities), from the class of complex systems into a class of systems describable by formal logic and deterministic models.

The simplification made by Descartes has laid the foundation of the scientific method for more than three centuries, and has become naturalized. Thus, at the Dartmouth workshop in 1956 (McCarthy et al. 1955), there was no serious critical analysis of the possibility of creating non-embodied intelligence. In the workshop's justification, we can see:

> The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate

> it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.

Thus, if we draw an analogy with thermodynamics, considering the level of technology available at that time, the goal of the workshop was formulated in the style of: "The study is to proceed on the basis of the conjecture that a Perpetual Motion Machine can exist. An attempt will be made to find a way to eliminate friction and energy losses in a moving mechanical system." It is ironic that Descartes' maxim of 'doubting the obvious' was not used.

In their book, the authors return the focus to the feasibility of AGI by introducing the ***mind-body*** continuum (or even stronger, the mind-body-environment continuum), which is a complex system. It is the central object for consideration, which can be used to discuss the phenomenon of human cognitive functions. Consequently, it can be inferred that living organisms may exhibit stronger adaptive (and cognitive) capabilities than computers, because they are integrated into the environment. This integration into the environment compensates for a lack of computational power through the right context. Essentially, in the ***mind-body-environment*** continuum, the "computational" power of the ***mind*** does not play a decisive role. In other words, "computational" capabilities are necessary, but they are not sufficient and cannot replace the other components of the ***mind-body-environment*** system.

By accepting the thesis that intelligence is only possible within a complex mind-body-environment system, the impossibility of creating AGI from purely physical considerations becomes clear. If we want to create the artificial phenomenon of intelligence, it must be embodied and have comparable abilities to humans in integrating into the environment and ecosystem. However, an intellectual embodied agent cannot be separated from the ecosystem and external environment, as they too provide signals for cognitive functions. Therefore, to operate such an agent, one would have to model or create an environment. But the human habitat cannot be rationally constrained. We reach the following unattainable outcomes: either the artificial intellectual agent must exist in the same environment as a human, i.e., be animate; or it would be necessary to have a model of the existing universe to provide an adequate (identical) environment for an inanimate intellectual agent. The first is impossible, as current technology level does not allow for the creation of life and animate objects. To demonstrate the impossibility of the second, one could rely, for example, on control theory (Conant 1970), which states that a good model should be isomorphic to its object of modeling. In other words, to construct such a model, at best, one would have to duplicate the existing universe, which contradicts the law of energy conservation.

Despite the impossibility of creating AGI as proven by the book's authors, the outlook is not pessimistic. The book rather corrects the future course of creating intelligent systems, pointing out the futility of further attempts to build an isolated abstract intelligence in the format proposed by the founding fathers of AI within the Dartmouth workshop - ***to make machines simulate intelligence.*** The issued to AGI verdict shifts the research focus towards the integration of animate and inanimate (artificial) systems. It becomes relevant to seek an answer to the question - is it possible to form a such integration of machine's "cognitive abilities" with cognitive abilities demonstrated by humans, which would:

- constitute a true mind-body-environment complex system,
- be capable of "cognitive interoperability" of its constituent parts,
- be capable of evolutionary development,
- possess abilities for reflection and self-correction.

The prerequisites for creating such systems are described in (Krinkin et al. 2023), and the intellectual systems of this type themselves are called Co-evolutionary hybrid intelligence (Krinkin et al. 2021).

If we accept the discourse of co-evolutionary development, then we need to look at the language system and the Turing test, to which Landgrebe and Smith pay a lot of attention. However, the authors, speaking about the capabilities of artificial intelligence (for example, 8.3, "The training data obtained from recorded conversations are never adequate as material to train a model because the conversations are conducted by

complex systems called people"), implicitly equate the ability to participate in a dialogue with intelligence following Turing. This is a mistake. The current volume of textual data that can potentially be used to train language models far exceeds the volume of texts a person can read in a lifetime. This means that a language model can potentially recognize and synthesize contexts successfully, and steer the dialogue in a certain direction (obviously, the machine cannot have intention, and therefore the ability to choose the direction of the discussion, but this can easily be simulated by a priori random or artificial "belief system" of the language model). In this case, the Turing test is an "imitation game", which the machine will win in the near future as easily as it wins in chess. Obviously, Large Language Models are not an extrapolation of texts (a model of reality is needed for it), but an interpolation. However, due to the fact that the volume of text training samples is orders of magnitude greater than human capabilities, the victory will be achieved.

The main problem is that intelligence, even in a dialogue between two people, is not always manifested. People, especially under typical conditions, tend to act as social automatons. They recognize context, know at which point in the dialogue (to which final state) they need to arrive, and use a rather algorithmic method of "advancement" in dialogue to the final state. Such a dialogue is not generated by a complex system and therefore does not pertain to intelligence. Chat bots, like ChatGPT, approximating the language body, potentially can be successfully used for training or as an interface for navigating through a language snapshot. However, they cannot be used for synthesizing new concepts, as their statistical nature will resist the emergence of objects not present in the training set.

If we consider an evolving hybrid system, the key criterion for the success of the Turing test, in its original formulation, becomes meaningless. The authors of the book in 10.2.4 note: "Language evolves continuously as new lexemes, new phrases, and new styles and registers are invented by speakers. Each undergoes a selection process, and those that survive contribute to a change in the shared body of language." Not the language itself, but precisely the process of its evolution occurs with the participation of intelligence (new abstract concepts are created, endowed with interpretations shared by the participants of the dialogue). Thus, the goal of the updated Turing test should not be to deceive the interlocutor, but to cooperate with the aim of the emergence of new forms, contexts, meanings in the language and the disappearance of old ones (that is, its evolution). This can only be organized in the context of solving a problem that, separately, at the initial moment in time, neither of the interlocutors can solve. Thus, for a hybrid system, instead of an "imitation game," a "cooperation game" is needed that generates a new co-dependent symbolic system. The necessity of dialogue, for the manifestation of intelligence, is conditioned by its intersubjective nature. Beside the communication aspect, symbols (symbolic systems) appeared to compensate for the finite human attention abilities: we need to fold parts of the reality model to word or symbol in order to keep our knowledge solid and connected.

From this perspective Large Language Models (as GPT) can be considered as a tool for navigation through some sort of knowledge graph. It was initially generated by humans and consolidated by machines. Concepts are connected by probabilistic edges based on human 'consensus' implicitly expressed in the native and artificial languages.

Despite the fact that one should not expect a machine to exhibit intelligence in a human sense, a hybrid system can offer new quality. For instance, it can be confidently said that machines are better at detecting patterns and characteristics in large volumes of data than humans. They can't interpret these findings, but a human can (including introducing new concepts into the language for them). Thus, in a hybrid system, humans may be equipped with a new type of perception—a data sense.

On the other hand, being a complex ***mind-body*** system, a human can effectively act in familiar contexts but can't verbalize this experience (that is, convert it into a logical *mind*-description at the level). For example, a professional race car driver or boxer can't create a linguistic description of their actions that could simply be used by another, less experienced athlete. In this case, based on neural networks, a non-verbal probabilistic model of "bodily behavior" can be built, which can be used for further training or teaching others. Narrow AI may open the door to non-verbal behavioral models and non-verbal knowledge.

In their book Landgrebe and Smith clearly demonstrated that the desire to create isolated AGI is akin to attempts to create a perpetual mechanical engine. Indeed, machines will never rule the world in this sense, but we still have a risk if people use the machines without clearly understanding what they can do and what they cannot.

REFERENCES

Conant. R. 1970. Every good regulator of a system must be a model of that system. *Int. J. Systems Sci.* 1(2):89-97.

Descartes. 1641. Rene Descartes. *Meditationes de Prima Philosophia*. In: *Qua Dei existentia et animae immortalitas demonstratur*

Krinkin, K. et al. 2023. Cognitive Architecture for Co-evolutionary Hybrid Intelligence. In: Goertzel, B., Ikle, M., Potapov, A., and Ponomaryov, D. (eds.). 2022. Artificial General Intelligence. AGI. *Lecture Notes in Computer Science* vol. 13539. Cham: Springer.

Krinkin, K.et al. 2021. Co-evolutionary hybrid intelligence. 5th Scientific School Dynamics of Complex Networks and their Applications (DCNA), pp. 112-115. Kaliningrad.

Landgrebe, J. and B. Smith. 2022. *Why Machines Will Never Rule the World. Artificial Intelligence Without Fear.* New York, London: Routledge.

McCarthy, J., M. L. Minsky, N. Rochester, and C. E. Shannon. 1955. A proposal for the Dartmouth summer research project on artificial intelligence. http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf