Machines that dream: A brief introduction into developing artificial general intelligence through AI-Kindergarten

Danko Nikolić

- Department of Neurophysiology, Max Planck Institute for Brain Research, Deutschordenstraße 46, D-60528 Frankfurt/M, Germany
- Frankfurt Institute for Advanced Studies (FIAS), Ruth-Moufang-Straße 1, D-60438 Frankfurt/M, Germany
- Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, Deutschordenstraße 46, D-60528 Frankfurt/M, Germany
- Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia

Correspondence:
Danko Nikolić
Max-Planck Institute for Brain Research
Deutschordenstr. 46
60528 Frankfurt am Main
email: danko.nikolic@gmail.com

Abstract:

Development of artificial general intelligence (AGI) may not be possible exclusively through human-created algorithms. Many aspects of human brain are not understandable to human scientists and engineers. Instead, AGI may require machines to create their own algorithms i.e., machines that learn to learn. It has been proposed that this can be achieved through AI-Kindergarten. In AI-Kindergarten machines are not left alone to figure out on their own the necessary algorithms, but they are heavily guided through human feedback. The feedback comes in a form of everyday interactions but also in a form of scientific knowledge about development of species and individuals. Information obtained from humans is integrated through a computational process that corresponds to the biological function of sleep and dreams. Importantly, AGI created that way has no danger of going rogue. It is completely safe while maximally benefiting humanity.

There is a long-lasting dream in creating artificial general intelligence (AGI). Today's artificial intelligence (AI) is not yet there. The approach of today is to implement algorithms based on insights of human programmers and engineers. Hence, much effort is being invested into engineering new learning algorithms and information processing systems. The hope is that a right set of algorithms will be eventually created — making up a machine that will be able to learn on its own to the extent of becoming an AGI.

A possible problem is that this effort based on human-developed algorithms may not be sufficient to bring about AGI. The reason is simple: It is likely that a human engineer cannot understand the complex processes of the brain and mind sufficiently well in order to create a computer program e.g., in C++, that would then result in AI that is generally intelligent.

There is rich evidence suggesting our limited capability of understanding engineering details of the brain. For example, a biologist cannot infer what changes on behavior of an animal will be caused by a change of a nucleotide in DNA. The interactions between genes themselves and between genes and their environment are just too complex to ever be understood by a human mind with such a precision. Similar evidence comes from mathematical theory of dynamical systems. From chaos theory we know that there are mathematical systems consisting of only a few equations that are too complex for a human to understand the behavior of the equations. The only way to know how equations will behave is to run them in a computer simulation. Often these incomprehensible mathematical systems consist of only a few equations — a minimum of three for continuous systems, but already a single discrete equation can be chaotic (e.g., a logistic map).

What are then our chances to understand the brain, which probably has the number of interacting equations in the order of thousands if not millions or billions? How can we possibly understand the brain to a sufficient engineering detail? AGI created through human insights into workings of the brain may not be very likely simply on the ground of the underlying complexity.

Today's efforts in AI present no significant alternative to human-created learning algorithms. The only known alternative would be to use raw computing power to try out

randomly created learning equations, and select them on the basis of a fitness function — much like the natural evolution did. This approach is computationally unfeasible. Therefore, there seem to be no other option but employ human engineers to think up novel algorithms.

The result is a large number of solutions, but for very specific problems. New general algorithms, once that could bring us closer to AGI, do not seem to come out easily from such efforts. Some of the best general algorithms used today (e.g., deep learning) stem largely from 1980's.

So, what can we do? Is there any alternative or are we simply stuck with specialized AI?

The answer is: Yes, there is an alternative in the form of AI-Kindergarten. AI-Kindergarten is a method for development of AGI that uses a novel approach to the problem (Nikolić 2015a).

First, AI-Kindergarten is not much about developing new algorithms by human engineers. In fact, only a few relatively simple algorithms are needed to operate AI-Kindergarten.

AI-Kindergarten is more about offering the intelligent agents different levels of organization at which they can learn and thus become able to create their own algorithms. The algorithms created by humans operate at much lower levels of organization that in the traditional AI. These simple algorithms lay behind the ability of the agents to create (or "learn") more complex learning algorithms that human themselves could not possibly understand. And then these more complex algorithms operate in order to create behavior of human-level intelligence.

For this, a theory of organization of biological systems was needed that was more general than any theory so far such that the theory would be equally applicable to different levels of organization within living systems (cell, organ, organism) and non-living adaptive systems (AI). This theory is called practopoiesis (Nikolić 2015b) and is fundamentally describing the workings of a hierarchy of cybernetic controllers as it is founded in two fundamental theorems of cybernetics: requisite variety (Ashby 1947)

and good regulator theorem (Conant and Ashby, 1970). But this was not enough. Practopoiesis only provided the basic structure of adaptive systems. It was necessary also to specify how many levels of organization were needed and what the function of each level of origination was. It turned out that, to create AGI, we need more levels of organization than what has been imagined by the current brain theories or AI theories. Namely, adaptive agents that mimic biological intelligence need to operate at three levels of organization (see tri-traversal theory of the mind in Nikolić 2015b). This implies that for an AGI, it is not sufficient to have an advanced learning algorithm or multiple such algorithms. An AGI needs to rely on a set of algorithms that enable it to 'learn' new learning algorithms. And this has to be done on the fly. In effect, this requires conceiving an agent capable of AGI as having one more level of adaptive organization than what we thought so far.

The real implementation problem arises from the fact that these algorithms on how to learn-to-learn are also incomprehensible to human engineers or scientists. These algorithms correspond to the plethora of plasticity mechanisms that are encoded in our genes and are driving the development of our brains and for all our instincts. It is practically impossible to even enumerate those rules, not to mention understanding of the principles of their functioning. To solve that problem, AI-Kindergarten is invented (Nikolić 2015a) as a method understandable to a human mind for providing the most fundamental learning-to-learn algorithms for AGI.

Second, AI-Kindergarten is not about autonomously self-developing AI. A popular science fiction meme is that it is sufficient to give a smart AI an access to the Internet. Then the AI can download all the necessary information on its own and learn and develop autonomously. One just needs to wait until the machine spits out a super-smart agent. To the contrary, in AI-Kindergarten much of human input and supervision is needed all the way along the process of developing AGI.

However, this input is not in a form of direct engineering. It is a different type of human input, related to showing our intuition and demonstrating our own skills of dealing with world, and also related to our scientific knowledge of biology and psychology.

AI-Kindergarten takes advantage of the fact that biological evolution has already performed many, many experiments until it came up with rules for building our own brains and guiding our behavior. AI-Kindergarten is about extracting this existing knowledge from biological systems and implementing it into machines.

To do that AI-Kindergarten uses inputs from human trainers. If human engineering fails to specify the learning rules for the machine, human intuition can specify for the machine what kind of behavior the machine should produce, and then the machine is left to find out the proper rules on learning how to learn. We need to tell machines which behavior is in which situations desirable. This is provided during interactions with AI in a way similar to that in a real kindergarten — where teachers interact with our own children.

But AI-Kindergarten requires something else in addition. While our kids learn only at the level of developing their brains, AGI needs to learn at one organizational level lower, that is, at the level of "machine genes". To achieve that, AI-Kindergarten must combine ontogeny with phylogeny (i.e., development of an individual with development of the species). For that, data from biology and psychology are needed to structure the stages of AI development. That way, existing scientific knowledge on brain and behavior plays a much more important role in AI-Kindergarten than in classical AI in which the engineers are supposed to assimilate that knowledge and apply it in algorithms invented by themselves.

It would be incorrect to think that AI-Kindergarten does not require a high intensity of computation. On the contrary, we cannot run away from intensive computations to develop AGI. These heavy computations are primarily needed for integrating knowledge acquired from humans. The process of integrating knowledge within AI-Kindergarten corresponds to what biology has already invented when endowing us with the capability to sleep and dream. Much like our dreams are needed to internally integrate knowledge that we have acquired throughout the day, the AI developed in AI-Kindergarten needs to integrate knowledge acquired through interaction with humans. Consequently, without intensive dreaming, AGI cannot be developed.

Finally, due to the continuous interaction with humans and feedback, which occurs along all stages of AI development, the resulting AI remains safe in terms of developing exactly the type of behavior and instants that the creators require. The motives, instincts and interests of the resulting AI are carefully crafted and shaped through this process such that they match the needs of humans. There is a concern that AI could surprise us with some unintended type of behavior becoming rouge or rebellious (Bostrom 2014). AGI developed in AI-Kindergarten cannot do that. Much like selective breading of dogs makes them reliably gentle and human-friendly, a super-human intelligence produced in AI-Kindergarten has even more thoroughly imprinted into its “machine genes” the basic instincts of not harming humans. AI-Kindergarten, by its very nature, produces safe AI.

References:

Ashby, W. R. (1947) Principles of the self-organizing dynamic system. Journal of General Psychology 37: 125–128.

Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.

Conant, R. C. & W. Ashby, R. (1970) Every good regulator of a system must be a model of that system. International Journal of Systems Science 1(2):89-97.

Nikolić, D. (2015a) AI-Kindergarten: A method for developing biological-like artificial intelligence. (patent pending)

Nikolić, D. (2015b) Practopoiesis: Or how life fosters a mind. Journal of Theoretical Biology 373: 40–61.