

# A Way toward Human Level Artificial Intelligence

W. L. Dunin-Barkowski<sup>a, b, \*</sup> and V. V. Shakirov<sup>a, b</sup>

<sup>a</sup>Scientific Research Institute for System Analysis, Russian Academy of Sciences, Moscow, 117218 Russia

<sup>b</sup>Moscow Institute of Physics and Technology, Dolgoprudny, 141701 Russia

\*e-mail: wldbar@gmail.com

Received December 11, 2018; revised December 11, 2018; accepted December 11, 2018

**Abstract**—The dispersion of estimates of the time to achieve human level AI is discussed at length. Some of the reasons behind this diversity are exposed and thoroughly analyzed. A special role of human language in providing both natural human intelligence and AI is extensively discussed. The more straightforward and expectedly much faster than currently pursued way of proceeding to the goal is dotted and discussed in crude detail.

**Keywords:** human level artificial intelligence, Natural Language Processing, Real world learning, Semi-Supervised Learning, Pavlov Principle

**DOI:** 10.3103/S1060992X19010041

## 1. INTRODUCTION

We present hereby an outline of a method to proceed from the current state of affairs in the artificial intelligence (AI) to the “full scale AI”, “strong AI”, or “Artificial General Intelligence” (AGI), all three terms we consider as synonyms. The scheme has emerged as a result of our earnest attempt to solve this problem with minimal resources in the last two years. In that work, we did get ideas and principles, which seem to be promising for the fast movement to the goal. Also, we came to crude evaluation of the minimal resources which are needed to solve this task. In strict terms, we present here just the ideas which emerge as a result of our goal-directed work on the problem of making the AGI as fast as possible with minimal resources. Also, we put additional arguments on the role of the natural human language in procuring human intelligence. We believe that even ideas in this field worth publication as the main problem in question—the real, full scale artificial intelligence seems to be really important for too many people.

The structure of the paper is following. First, we shortly review the currently adopted approaches to problem solution and pinpoint their probable weakness. Then we are characterizing some ways of solution of problems, comparable in society impact scale to the AGI problem. And finally we outline contours of the approach, which we suggest, and shortly characterize its expected features. The paper is targeted to narrow professional audience with omission of references, probably needed for relatively distant from AGI problems researchers. Also, we have recently published “Review of State-of-the-Art in Deep Learning Artificial Intelligence” [1] and suppose that readers can have something like this paper under hand for references, when they are needed.

## 2. THE PROBLEMS

One of the problems with AGI is that there is no rigorous definition of the term. We have claimed earlier that attempts to give the definition are actually not much productive [1]. Just by often cited analogy of constructing AGI with creating airplanes one can see that definitions might not be necessary. Say, in spite of the fact that airplane flights officially started on December 17, 1903 (Wrights brothers’ first flight) there were flights before that date, while the real air-planes did not appeared before the WW1. One of the first formal AI definitions, the Turing test, has been passed by multiple systems and currently has no practical value [2]. In spite of this critics of formal definitions, a practically important (although very simple in formulation) definition has been recently proposed: “AGI is an intelligence, which is not weaker than any human intelligence” [3]. This definition might be practically useful (a jury of human experts can compare “strength” of different intelligences), but it is nonconstructive as it yields no concrete features of intelligence, which might be declared as concrete goals of AI construction.

There are many judgments of professionals and non-professionals on the question, when AGI is coming. The diversity of answers (actually, from zero (“AGI already exists in some secret place”) to infinity (“AGI is impossible”), which can be discovered by simple quest to the Internet, reflects uncertainty of the current situation as well as the importance of the event.

It is appropriate here to make reference to a historical precedent. In the beginning of the year 1957 there were many “optimistic” prognoses that artificial satellites of the Earth will be launched in the next 50 years to 1957. In fact, the Soviet first Sputnik has been launched on October 4, 1957.

The pessimistic prognoses for AGI among other reasons [2] often reiterate kind of the mantra: “They have promised it to come yesterday, but it didn’t”. Indeed, this critique has real grounds.

On the side of optimistic points of view, we can cite I. Pavlov and B. Russel [4] (the latter being really charmed by Pavlov’s experimental results and the perspectives, which were exposed by them [5]). According to Pavlov, numerous interacting conditioned reflexes can explain human intelligence. The Deep Learning revolution (started with amazing achievement of AlexNet in 2012 [6]) actually proved that Pavlov’s ingenious guess was correct. Since 2012, little by little it becomes clear that simple adaptive change of many connections in brain-like constructs (in the “deep” neural networks) leads to emergence of intelligent properties of the constructs. Also important is that until November 2014 there was practically one effective method of tuning (“training”, “teaching”) of neural networks – error Back Propagation (BP). And there were multiple evidences that nothing like BP can be implemented in physiological neural networks. The work of Timothy Lillicrap et al. [7] has demonstrated that training of deep neural networks is possible without BP. Moreover, the later works [8–11] have shown that several well-known subsystems of real neural systems (say, subsystems of such neural modulators as dopamine, norepinephrine, 5-HT, etc.) fit very well to implement in brain the schemes of training the artificial neural networks. Thus, although not firmly proven by targeted experiments or reliable theory, the idea that intelligence of artificial and natural neural systems have the same general mechanism, looks now fairly plausible.

Thus, with a high degree of confidence we can state that Deep Learning is the mechanism, which enables smartness of neuron systems (the term “neuron” has been proposed as a general term for both, artificial and natural neural systems [12]). And DL can make the neuron systems intelligent. However, the mere presence of DL in artificial systems is not enough to get the human level AI as the same principles work in neural systems of all animals. Surprisingly, I.P. Pavlov also gave the proper and highly relevant in present days identification of the neural sub-system, operated only by humans, and necessary for the human level intelligence. Pavlov has coined the term “the Second Signal System” (SSS), denoting by it the language [5]. We argue below that the term is precise, meaningful and at least for two millenniums (since the famous first words of Gospel of John became known: “In the beginning was the word”; end of citation) was considered (consciously or unconsciously) the main essence of being human. There is no generally accepted concept about the origin of the language. One of the simplest idea here is that the language has been invented by first humans. We intentionally use the expression “first humans” as before appearance of language there was no human society, just herds (or other ensembles of individuals) of human-looking animals. We reiterate here that the language origin has not been definitely determined yet. Nevertheless, the simplest hypothesis, based on historical data on human culture is that language has been invented by humans. It has been proven that next stages of language development: writings and sign languages were indeed invented. The most important feature of language, compared to the other systems of communication between individuals is that this communication channel has a high informational capacity. Numerically, human language by capacity exceeds communication channels between other animals in several orders of values. It is really striking that while I.P. Pavlov did not know precise ways of quantitative characterization of communication channels (he passed away in 1936, many years before Shannon published initial basics of the Information Theory), he has realized that language is extremely powerful and qualitatively distinct (from other interpersonal signaling systems) way of communication, the Second Signal System. Saying by computer analogies, language provides wide bandpass communicational pipeline between “biological computers” of individual human beings. So, the term Second Signal System surprisingly precisely denotes the phenomenon, which makes humans to be human.

We will add few more words about the significance and the role of language in human life from the informational technologies point of view. First, in this respect, the uniqueness of a man, as a biological species, is in that humans is the only biological species, which *can* fluently use language and *actually uses* it. The attempts to prove abilities of some other animal species to use human language obviously failed (although, some dogs, cats, parrots, apes, etc., can recognize some words and even phrases). The evolutionary role of the language in terms of species computational (and, consequently, adaptational) abilities looks the following way. Indeed in the evolutionary branch of primates the size of the human brain substantially exceeds (more than twice) brains of the other primates. Less sure, but one can argue that human

neural scheme construction might be better than animal's neural design. The advantage in size, considering the computational power, is approximately proportional to the size (that is true for computers of the same hardware types). That makes humans possess 2–3 times more computational power than any other species in a comparable ecological niche (to exclude from comparison elephants and dolphins). However, by such a measure of functionality as the ability to use language, the human brain surpasses all other brains. The language as wide bandpass communicational pipeline between human “biological computers” makes possible formation of clusters of these “computers”. Due to that fact, computational power of human brains clusters might surpass the power of competing species in many orders of value. Also, language provides humans with exocortex (first, books and, since the XX century, computer type memories) of unlimited informational capacity. And we will reiterate that most probably, the language has been invented by humans. The main arguments for this option: it is within human capabilities and is simple. Occam razor criterion in fact cuts off other possibilities.

As follows from the arguments, given above, we understand, first, how it happens that neuronics systems are smart. The answer is—DL systems can be smart. Second, what is important in human level AI? The answer is—the second signal system of humans, i.e. the language. In next section we will address this problem.

### 3. THE LANGUAGE

Technically, there are no obstacles in wide bandpass communication between an autonomous intelligent agent and humans or other intelligent agents. So, to see how fast communication leads to human intelligence we ought to trace how an intelligent animal, acquiring the language, becomes human. Happily, there is no need to trace the historical processes of how the humankind has got its intelligence. That is because each human individual gets her/his intelligence in ontogenesis from birth to adulthood. The wide bandpass communication option in particular means that one agent can in fact effectively use sense organs of other agents remotely in space and/or time. Language provides availability of such functions. We will point also to two other functions, available to agents with help of the language. First is decomposition of the mental task between members of a social group. Second, is the same for all other social tasks. With handling natural languages one should understand, first of all, that language is a tool (or toolkit, if we mean a set of words, and other language instruments) which serves to human social life. Other tools, say, music or other non-verbal arts are also involved in that function. Besides, by simple self-introspection anybody of us, humans, is aware of the fact that language, words are intensively used in our individual mental processes. Of course, streams of words are not always streams of thoughts, but words very often are included in latter streams. So, language can serve as thought enhancer.

Two more considerations should be added in description of language properties. First of them is connected with the fact that for the last hundred thousand years humans did not significantly change biologically. That means that the neural representation of language (and, of particular words) should not substantially differ from representation of other objects in brain. In that sense, words have animal origin. Connected with that, but rather opposite are following considerations. Let us reiterate that the words are tools. We know that all tools are man-made. Besides, all toolkits, which appeared early in human history, are permanently evolving due to human efforts. A simple example of that in relation to language is well-known invention of words by concrete people (“neuron”, “synapse”, “quantum”, etc.). From this fact, i.e. a permanent enhancement of language usability qualities in historic process, one can suppose that current natural languages might accumulate through the many millenniums long history a lot of special qualities, well-suited for effective employment, which are hard to replace by fast engineering. These qualities might refer to any capacity of using of natural languages, either for communication or for thoughts enhancement. This argument gives a technical reason why we should enable AI with fluent use of the natural languages.

### 4. ON ROADMAPS AND CORNERSTONES

The Zeno's paradox of the Tortoise and Achilles is well-known and is just a sophism. However, a kind of a likely paradigm, with substantial practical effects (contrary to the original one), currently might be suspected to interfere with the “international race for AGI”. Here is what and why.

Mikolov et al. [13] writes: “since solving AI” seems too complex a task to be pursued all at once <...> the computational community has preferred to focus on solving relatively narrow empirical problems that are important for specific applications”. With this argument, these authors propose to move on the way to AI “a concrete roadmap <...> in realistic, small steps, that are <...> incrementally structured in such a way that, jointly, they should lead <...> close to the ultimate goal of implementing a powerful AI.” Unfortunately, the proposed list of concrete actions (cornerstones?) is, first, very long and, second, it is not clear

now, how the execution of those steps will move us on the road to AI. In fact, there is no constructive definition of the “so much needed AI” and a priori nobody can tell how the proposed list can lead us close to the (undefined) goal. It is worth not to look farther than to any handbook of philosophy, philology or psychology to see lists of human abilities, and new abilities are being permanently discovered in new researches in those fields. All of them might be proclaimed important for the artificial intelligence. In rather skeptical attitude of a physicist to the listed above “soft sciences” one can tell that currently considered “lists of human intellectual abilities” are overwhelmingly large and redundant. The tasks in that list, while seemingly being somewhat different from each other, are fairly probably based on a limited number of base skills (alas, the list of those skills is not yet known). That’s why it’s much more important to learn how to learn to behave “like humans” than to waste time on solving in fact exponential number of different (but might be essentially same) tasks, with a finite amount of efforts and time, needed to resolve each of them. That makes the solution time for the whole problem exponential in reality, not just in imagination as in Zeno’s paradox. The exponent rate of this process might be of the order of number of words in vocabulary. That means that this road to AGI might happen to be infinite. In next section, we propose the other strategy of engineering the human level AI.

## 5. THE STRATEGY: EYES ARE SCARED WHILE HANDS ARE DOING

The title presents a Russian proverb, which means that the planning stage (sometimes) might be skipped even in serious enterprises.

Any list of contemporary achievements of Deep learning (e.g. [1]) demonstrates how many human abilities AI systems currently have. And our experience of permanent tracing of arXiv publications on AI tells us that these days each week at least ten sound publications appear. One can even tell that almost any particular trait of human intelligence has a solid counterpart in already implemented AI system. What is really lacking – it the “real intelligence” of AI systems. And availability of intelligence currently can be judged only by human experts. No other exact or deep formulation of the Artificial General Intelligence exists. Can we move further to AGI just having what is stated above? We argue that we can. After years of week to week following recent successes in Neuroscience and AI [14] we coined a current opinion on the way to move to creating the AGI.

Our general idea in fact coincides with the original moving force of the DL in 2012 and further. We understand that nobody understands what exactly is AGI. However, observing concrete behavior of AI agent people can tell if it is closer to AGI or farther from it. So partly supervised learning, where teacher guides the learning process in direction of making the AI agent be more like AGI has chances really achieve that goal.

We propose the direction of immediate works toward human level AI as defined by the following wordy formula: Adaptive Integrator of Artificial Intelligence Competences (abbreviation “AI2C”, pronounced as: “AI double C”). In this formula “Adaptive” means adaptive to demands of the AI agent human operators, who urge the agent behave like humans. “Integrator” stays for the computer, endorsed with visual, acoustic and tactile sensors, graphical and acoustical outputs at which installed are several trained modules, which can intellectually process sensory input information. Those modules can be taken from the current pool of “AI competences”, i.e. from the rich depository of the intellectual programs available at GitHub [14]. Among the modules necessarily should be present “The Adaptator”, which works on principles of semi-supervised learning and learns to make computer output signals complying with correcting actions and demands of human computer operators. The life long record of computer inputs and outputs should be also available to the system. We suppose that the goal-directed training of an AI2C agent to teach it human intellectual behavior in human environment is the shortest way to AGI. The goal is human level AI. The training process should follow the standard way of teaching intellectuals from less intelligent children. And we reiterate that the proposed approach is the precedent-based way of teaching computer things, which we do not clearly understand. In fact, the same strategy has provided the DL “revolutionary” improvement in 2012 [6].

To be clear, we give here the explicit comparison of the proposed way to approach AGI with the well-known way of successful solution of the ImageNet task. The classical approach to the latter problem was to obtain lists of features of all visual objects, to understand, how these features can be detected on the images and to give definition of the objects to be recognized in terms of the features. The DL approach did nothing of that. It just gives the correct answers to the visual questions. The answers to the questions have been made by humans. Later, the computations have revealed that response patterns of lower levels of artificial neural networks practically coincide with analogous “layers” in the vertebrate visual system [15]. We can conclude that the most important for object recognition is not the precise description of

object, but self-organization of object recognition networks under the demand to operate “like humans”. And we propose the same methodology to get the AGI agents.

Besides the above given general framework for composing of new AI agents we have additional concrete ideas for their practical implementation. They are in line with the recently proposed platforms for robotic skill Learning [17, 18].

Following that works we propose sharing experience replay memory, for RL agents to be able to use the accumulated knowledge of other previously existed robots who lived within the analogous environment. If agents use the same pre-trained network for primary processing of sensory data, then sharing experience replay is trivial, because any agent with any working algorithm would understand shared sensor embedding. In fact, they “speak” the same sensor language. It has a clear parallel with human cultural evolution, which outpaced biological evolution tremendously. We argue for the beginning of robotic cultural evolution. Collections of experience replay data are essentially equivalent to textbooks and instructional videos – they store and transmit knowledge. In modern schools and universities, humans quickly accumulate the fruits of billions of human-years of exploration. Without that accumulation, it’s very hard for a human to invent even simplest Oldowan tools [19], and it’s almost impossible to invent a bow, within human lifetime. So even if RL agents learn slowly, it seems like their full potential isn’t revealed and realized yet to the full extent. This is especially important when we speak about really hard and complex environments like real world. It might be too much of wasting time and computational resources to learn real world, starting each time from zero. Using the proposed approach, it would be much more easy and fast to create different agents with different algorithms.

We would also propose an annual Intelligence Competition, say, at NeurIPS. At this panel the team of human judges will rank the systems presented to their attention, based on the demonstrated intelligence.

## 6. CONCLUSIONS

We have discussed several issues of the problem of achieving human level AI. First, we discussed the absence of definition of the goal and demonstrated that the precise definition might be not that helpful. Then we discussed deep learning and natural language as the most important components of the human level intelligence. Then we put forward arguments why some roadmaps and cornerstones might be excessive if not misleading on the way to AGI. Finally, we have proposed a strategy, which probably would yield a reasonably short way to the goal.

## ACKNOWLEDGMENTS

The work was supported by Scientific Research Institute for System Analysis, by iPavlov project on artificial neural intelligence at Moscow Institute of Physics and Technology and by Russian Foundation of Basic Research grant no. 16-07-01059.

## REFERENCES

1. Shakirov, V.V., Solovyeva, K.P., and Dunin-Barkowski, W.L., Review of state-of-the-art in deep learning artificial intelligence, *Opt. Mem. Neural Networks*, 2018, vol. 27, no. 2, pp. 65–80.
2. Russel, S.J. and Norvig, P., *Artificial Intelligence. A Modern Approach*, Edinburgh Gate: Pearson, 2016, 3rd ed.
3. Dunina-Barkovskaya, N.V., A public statement at the 148-th Working session “Calderon of Ideas of Neuronet,” 2018. <https://leader-id.ru/event/11435/>.
4. Russel, B., *The Scientific Outlook*, London: George Allen and Unwin, 1931.
5. Pavlov, I., Lectures on Functioning of Brain Hemispheres, 1927. <https://archive.org/details/b29929222>.
6. Krizhevsky, A., Sutskever, I., and Hinton, G.E., ImageNet Classification with Deep Convolutional Neural Networks, 2012. <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>.
7. Lillicrap, T.P., Cownden, D., Tweed, D.B., and Akerman, C.J., Random feedback weights support learning in deep neural networks, arXiv:1411.0247v1, 2014.
8. Solovyeva, K.P., Shchukin, T.N., Ivashchenko, A.A., and Dunin-Barkowski, W.L., Basic principles of neural processing, *III All-Russian Conference with International Participation “Hippocampus and Memory: Norm and Pathology,”* Pushchino, 2015, pp. 33–34.
9. Lillicrap, T.P., Cownden, D., Tweed, D.B., and Akerman, C.J., Random synaptic feedback weights support error backpropagation for deep learning, *Nat. Commun.*, 2016, vol. 7, no. 13276, p. 10.
10. Nokland, A., Direct Feedback Alignment Provides Learning in Deep Neural Networks, 2016. <https://arxiv.org/pdf/1609.01596.pdf>.

11. Dunin-Barkowski, W. and Solovyeva, K., Pavlov principle and brain reverse engineering, *2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, St. Louis, MO, 2018, pp. 1–5. doi 10.1109/CIBCB.2018.8404975
12. Dunin-Barkowski, W.L. and Solovyeva, K.P., *Pavlov Principle in Neuronic Systems, Updated, BICA 2017, Invited Talk*, 2017. doi 10.13140/RG.2.2.34785.20322
13. Mikolov, T., Joulin, A., and Baroni, M., A Roadmap towards Machine Intelligence, arXiv:1511.08130v2, 2016.
14. <http://rebrain.2045.com/>.
15. <https://github.com/>.
16. Yamins, D.L.K. and DiCarlo, J.J., Using goal-driven deep learning models to understand sensory cortex, *Nat. Neurosci.*, 2016, vol. 19, pp. 356–365.
17. Christiano, P., Leike, J., Brown, T.B., Martic, M., Legg, S., and Amodei, D., Deep reinforcement learning from human preferences, 2017. <https://arxiv.org/abs/1706.03741>.
18. Mandlekar, A., Zhu, Y., Garg, A., Booher, J., Spero, M., Tung, A., Gao, J., Emmons, J., Gupta, A., Orbay, E., Savarese, S., and Fei-Fei, L., RoboTurk: A Crowdsourcing Platform for Robotic Skill Learning through Imitation, 2018. <https://arxiv.org/abs/1811.02790>.
19. Morgan, T.J.H., Uomini, N.T., Rendell, L.E., Chouinard-Thuly, L., Street, S.E., Lewis H.M., Cross, C.P., Evans, C., Kearney, R., de la Torre, I., Whiten, A., and Laland, K.N., Experimental evidence for the co-evolution of hominin tool-making teaching and language, *Nat. Commun.*, 2015, no. 7029.