Deep Learning and Artificial General Intelligence: Still a Long Way to Go

Maciej Świechowski

QED Software, Warsaw, Poland

ORCID: 0000-0002-8941-3199

maciej.swiechowski@qed.pl

Abstract—**In recent years, deep learning using neural network architecture, i.e. deep neural networks, has been on the frontier of computer science research. It has even lead to superhuman performance in some problems, e.g., in computer vision, games and biology, and as a result the term deep learning revolution was coined.**

The undisputed success and rapid growth of deep learning suggests that, in future, it might become an enabler for Artificial General Intelligence (AGI). In this article, we approach this statement critically showing five major reasons of why deep neural networks, as of the current state, are not ready to be the technique of choice for reaching AGI.

Index Terms—**Deep Learning, Deep Neural Networks, Artificial Intelligence, Artificial General Intelligence, Machine Learning**

I. INTRODUCTION

The current applications of Artificial Intelligence (AI) belong to the so-called "*narrow AI*" [1]. *Narrow AI* is often very successful but only in solving a particular task. For example, a top quality chess playing program [2] is unable to form any sentence in a natural language.

Artificial General Intelligence (AGI) [1], [3], however, is the idea of creating multi-purpose AI that would be capable of learning and performing any intellectual tasks humans do. On a sidenote, a hypothetical AI that would surpass humans in a given domain is referred to as "*strong AI*". If such an AI would surpass humans in solving all problems that require intelligence, it would be called "*strong AGI*" and its existence would mean that the human development has achieved singularity.

AGI can be viewed as human-like intelligence displayed by machines. In fact, the most popular validation tests for AGI involve comparisons to humans in efficacy in solving tasks:

1) *Turing Test* [4] – ability to carry out a believable conversation in a natural language. This test was proposed initially by Alan Turing in a different form and by the name of *Imitation Game*. It is associated with one of the first research works on AI and computers. We also recommend a recent paper [5] that discusses the importance of *Imitation Game AI Competitions*.

2) *Robot College Student Test* [3] – ability to enroll in a university, pass all exams and obtain a degree.

3) *Employment Test* [6] – ability have a job which is ordinarily performed by humans and work as effectively in it as humans do.

4) *The Coffee Test* - ability to enter an ordinary American house (without any predefined setup) and make a cup of coffee. This includes figuring out where the cup as well as the coffee may be, mix all ingredients (sugar, water), etc. According to [7], it was proposed by Steve Wozniak.

AGI may seem as something unreachable and vague. While it has not been achieved yet, it might be worth thinking of it as a distant research challenge. It has been one of the original long-term motivations driving the AI field.

In this article, we are not to claim whether AGI will be possible or not. We focus on deep learning, which in the last decade, has become one of the major topics in research [8]. Moreover, there are many spectacular achievements and commercial applications [9] of deep learning. It is safe to say, that deep learning and deep neural networks (DNNs) have revolutionized the field [10].

This has raised the question – are DNNs the technological enabler that will lead us to AGI? In this article, we devote five sections to major reasons of why DNNs are not yet ready to be the technological driver for AGI. Due to strict page limits, we assume readers' familiarity with the concepts of deep learning and neural networks, but not necessarily with AGI.

In order for this article to be technical and not philosophical, we omit such aspects of human thinking as sentience, self-awareness and motivation. We also do not theorize that having a body that can feel stimuli is a prerequisite for intelligence.

II. REQUIREMENT FOR A LOT OF TRAINING DATA

The process of training deep neural networks requires huge volumes of data [11]. The amount of data necessary to reach high quality of a DNN model (e.g., expected accuracy of a classifier) depends on the complexity of a given problem and the size of the neural network used. The GPT-3 deep neural network, which is currently one of the biggest networks ever trained, has 175 billion learning parameters [12].

When a model has that many parameters that need to be optimized solely based on feeding them with training data, then we need to make sure that they are many training samples. **We argue that this requirement for a large corpus of training data conflicts with the nature of AGI.**

Firstly, it is not feasible to have so much training data for any tasks. AGI is inspired by human intelligence and most methods of testing whether we achieve it involve comparisons

to humans. Humans can learn to be effective at virtually unlimited number of tasks. However, children do not need to observe a lot of various cats in order to be able to recognize a cat. While the learning by children has not been fully understood yet, it is most probable that a human's brain builds an abstract internal representation of a concept relatively quickly [13].

Secondly, large neural networks combined with large volumes of data result in high computational cost and, therefore, long training time [14]. It would be completely ineffective to train for a long time for every task AGI is presented with. We need faster ways of training/creating AGI models.

III. TRANSFER LEARNING, USING ANALOGIES AND INTUITION

Although transfer learning, using analogies and intuition mean different things, all of them may help in speeding up the learning process.

In essence, transfer learning [15] is a paradigm in machine learning that consist in training a model for a given task and reusing it in a completely new task without or with minimal retraining. Multi-domain (or multi-task) learning is a step further and involves training one model for a variety of tasks. Finding similar concepts by analogy or reducing a problem to known ones by analogy is one way to implement transfer learning. Another way might be to provide an ontology that maps concepts from one task to another [16].

The problem with transfer learning is that there have not been any successful attempts on a very large scale i.e. for a lot of really diverse tasks. Readers are advised to refer to collective works and surveys on this topic, e.g. [16], [17]. We do not want to state that full multi-task learning is not possible with DNNs but rather that it is still in its early phase of research with very limited applications.

Intuition is a concept attributed to human thinking, which has not been yet defined in terms of artificial intelligence. According to [18], intuitive solutions are found extremely quickly. They might not be optimal – just first good approximations. Neural networks do not have such property. They work in one mode, i.e. the full potential, and they need to perform the necessary inference to give output signal. However, it may be argued that iterative algorithms such as Monte Carlo Tree Search [19] or that can be asked at any time about the current solution to the problem are closer to displaying some aspects of intuition by means of early solutions.

Humans display enormous capabilities to generalize experiences gained at one task into similar ones. For example, when we learn how to cook a specific dish, we absorb a lot of abstract concepts about cooking, in general, that help us to prepare new dishes. Or we can read a manual of a game (e.g. a board game) that we never played before and be able to play it as long as it is based on concepts we have already encountered in other games. Furthermore, the ability to transfer knowledge and act in completely new situation is a strong trait of intelligence [20].
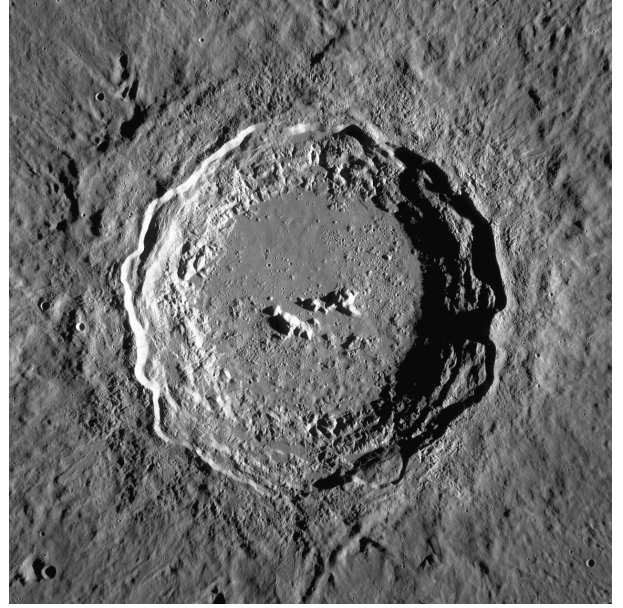

Fig. 1. A lunar impact crater called *Copernicus* (*the image is avaiable from Wikipedia creative commons*).

IV. ABSTRACT OUT-OF-THE-BOX REASONING

DNNs are universal function approximators. Ultimately, they map an input vector $\bar{x}$ into output vector:

$$f(\bar{x}) : R^n \rightarrow R^m \qquad (1)$$

The authors of [21] say that machine learning is a modern rephrase of curve fitting. NNs learn on training examples and once they are trained they are able to generalize into new examples by similarity in how they activate the network.

Humans, on the other hand, apply reasoning or even meta-reasoning that involves outside information. For example, let us consider a task of detecting whether a moon crater, such as the one shown in Figure 1 on a picture is concave or convex. Machine learning techniques can be easily fooled because concave craters with shadows to the left are very similar to convex craters with shadows to the right. If there is a watermark with the time of day the picture was made, humans may notice it spontaneously and infer the direction of the sun and then accurate shadows. If there were no watermarks with a date information in the training, machine learning models would at best ignore them completely or treat as a noise in data. In fact, even if there were some watermarks on images used during training, it is unlikely that the same ML model would learn to classify the image and interpret text information on top of it.

Moreover, humans use a lot of *out-of-the-box* techniques. Let us consider verbal communication "task" in a natural language. Humans will also incorporate non-verbal communication [22], anyway.

We argue that logical reasoning is required to go beyond just mathematical classification or regression represented by machine learning models. Logical conclusions may completely

change the process how we interpret what we see, hear, smell etc. Logical reasoning goes beyond fitting new experiences to the known ones. It allows us to reinterpret things and add adapt to new situations dynamically.

V. EXPLAINABLE REASONING

The proposed definitions of AGI involve some kind of performance metrics at given task and a comparison to human intelligence [3], [6]. When attempting to solve a complex task that requires intelligence, **humans are capable at explaining their thought process** and justify their decisions or actions. In rare occasions when we are not able to, we will say that it was an instinct, impulse or intuition.

We argue that entities that possess general intelligence should be able to provide a similar form of justification for their decisions/actions. For the purpose of this article, let us call this explainability.

DNNs as well as traditional NNs are machine-learning models that are considered one of the least explainable ones [24]. Even the recent advancements in the so-called eXplainable AI (XAI) [25] give only limited tools such as feature importance, uncertainty, error, examples, visual summaries etc. The bottom line is that NNs are large numerical function approximators and do not really operate on explainable concepts inherently.

Naturally, there are machine learning models such as *decision trees* [26] that are inherently explainable and interpretable, but they are usually much less effective in solving complex problems, for which deep learning is employed.

VI. VULNERABILITY TO ATTACKS

A significant part of the overall success of deep learning has been due to Convolutional Neural Networks (CNNs). They are state-of-the-art approaches in many image classification tasks and computer vision [27], in general. As discussed in [28], CNNs – in particular – can be maliciously manipulated by presented specially prepared input. This is referred to as *adversarial attacks*. Naturally, not all examples of fooling deep learning AI have to be malicious [29]. Imagine a road sign detection AI that is fooled because the view by the camera has been obscured. The are hundreds of example on the Internet, such as the one shown in Figure 2, that show spectacular misclassifications of images by deep neural networks. These examples are considered funny, because humans do not make such obvious mistakes.

Partial reason for such attacks being possible is that these types DNNs do not operate on particularly meaningful features as input. Such features are discovered by the network within intermediate layers. The input consists in a lot of very simple features – RGB color values for millions of pixels – and the next layers of the network may represent features using mathematical operations on them including weighted aggregation. It is possible to exploit this, for instance, by specifying a completely different image that will result in the same results of internal operations further in the network.

Humans are particularly good at seeing patterns and concepts, not individual pixels. Moreover, humans have all sorts of built-in mechanisms, such as premonition, that prevent them from being fooled so easily but more importantly – we build our judgements in a holistic fashion, also using external data. Human intelligence is not a portfolio of specialized algorithms for particular tasks. **We use one brain for everything**. Therefore, it is likely that the defence mechanisms go in pair with fully implemented transfer learning. Machine learning models optimized for one problem have limited additional validation possibilities, because there is no "world" for them beyond this one problem. They cannot answer, for instance: "*It looks like cat, but I know this is not a cat, because you have always tried to fool me so far in the history of our interactions*".

VII. CONCLUSIONS

Deep learning has proven to be able to break barriers in front of AI. One of the most prominent examples is an ancient board game – *Go* – which was considered to be one of the "grand challenges of AI" [30]. In 2016, *AlphaGo*, a program based on deep neural networks and Monte Carlo Tree Search, defeated Lee Se Dol, who is one of the strongest *Go* players of all time. *AlphaGo* was a major breakthrough in computer science field. This approach has inspired many successful approaches to other games and problems [31].

Through the history of civilization, human intelligence has allowed us to solve numerous problems and overcome many challenges. AGI, to earn its title, also needs to be effective at solving complex problems. It seems like deep learning is a promising way to achieving this goal.

However, AGI is something more that efficacy at one task. In this short article we have pointed out five characteristics of deep neural network that make us believe that AGI is still far beyond reach of deep learning.

These characteristics revolve around:

- Faster learning, even with limited training examples
- Transfer learning and using analogies
- Abstract out-of-the-box (logical) reasoning
- Deep explanation of the thought process
- Defence mechanisms e.g. against malicious behavior

It is worth aiming for the improvement in these five areas regardless of AGI. We believe that they all help in creating better and more robust machine learning models.

ACKNOWLEDGEMENTS

REFERENCES

[1] B. Goertzel and C. Pennachin, *Artificial General Intelligence*, B. Goertzel and C. Pennachin, Eds. Springer, 2007, vol. 2.
[2] D. Hassabis, "Artificial Intelligence: Chess match of the century," *Nature*, vol. 544, no. 7651, pp. 413–414, 2017.
[3] B. Goertzel, "Artificial General Intelligence: Concept, State of the Art, and Future Prospects," *Journal of Artificial General Intelligence*, vol. 5, no. 1, p. 1, 2014.
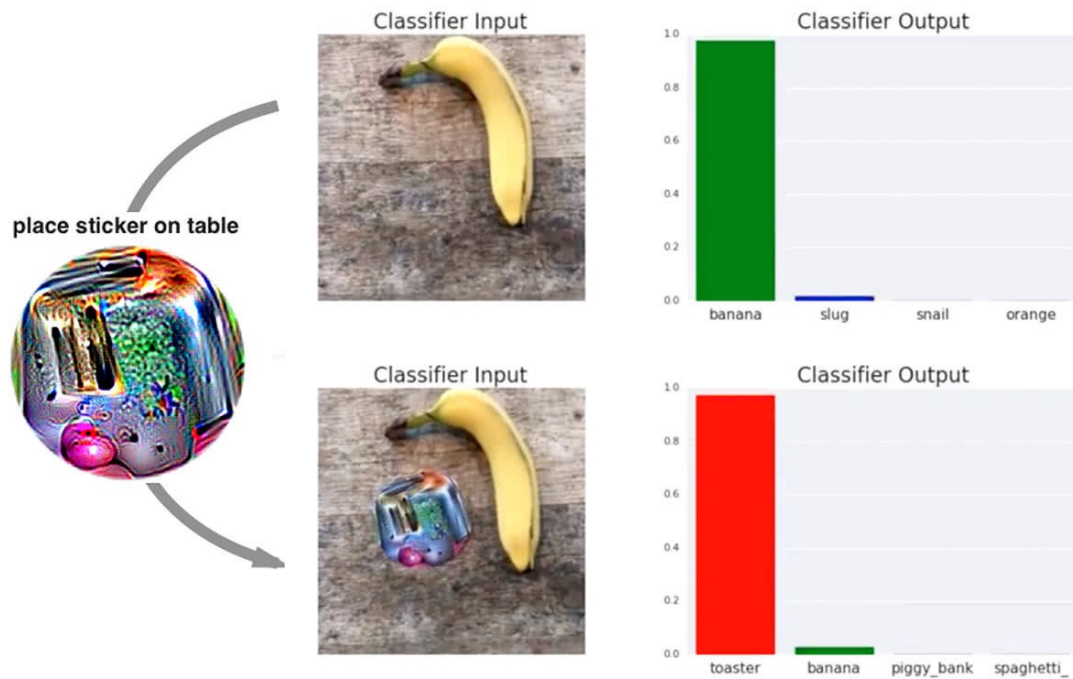
Fig. 2. The image on the top was classified as a banana by a deep convolutional neural network, whereas the image on the bottom was classified as a toaster. The figure is available in a paper [23] published on *arxiv* open repistory. Interested readers are advised to consult this paper for description of the method of how the classifier was fooled.

[4] J. H. Moor, "An Analysis of the Turing Test," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 30, no. 4, pp. 249–257, 1976.

[5] M. Świechowski, "Game AI Competitions: Motivation for the Imitation Game-Playing Competition," in *2020 Federated Conference on Computer Science and Information Systems (FedCSIS)*, vol. 21. IEEE, 2020, pp. 155–160.

[6] N. J. Nilsson, "Human-level artificial intelligence? Be serious!" *AI magazine*, vol. 26, no. 4, pp. 68–68, 2005.

[7] B. Goertzel, M. Iklé, and J. Wigmore, "The Architecture of Human-Like General Intelligence," in *Theoretical foundations of artificial general intelligence*. Springer, 2012, pp. 123–144.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," *IEEE Access*, vol. 7, pp. 53 040–53 065, 2019.

[10] T. J. Sejnowski, *The Deep Learning Revolution*. MIT press, 2018.

[11] X.-W. Chen and X. Lin, "Big Data Deep Learning: Challenges and Perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.

[12] L. Floridi and M. Chiriatti, "GPT-3: Its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, no. 4, pp. 681–694, 2020.

[13] J. Spicer and A. N. Sanborn, "What does the mind learn? a comparison of human and machine learning representations," *Current opinion in neurobiology*, vol. 55, pp. 97–102, 2019.

[14] C. Berner, G. Brockman, B. Chan, V. Cheung, P. Dębiak, C. Dennison, D. Farhi, Q. Fischer, S. Hashme, C. Hesse *et al.*, "Dota 2 with large scale deep reinforcement learning," *arXiv preprint arXiv:1912.06680*, 2019.

[15] L. Torrey and J. Shavlik, "Transfer Learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 2010, pp. 242–264.

[16] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A Survey on Deep Transfer Learning," in *International conference on artificial neural networks*. Springer, 2018, pp. 270–279.

[17] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[18] J. Mańdziuk, *Knowledge-Free and Learning-Based Methods in Intelligent Game Playing*. Springer, 2010, vol. 276.

[19] M. Świechowski, K. Godlewski, B. Sawicki, and J. Mańdziuk, "Monte Carlo Tree Search: A Review of Recent Modifications and Applications," *arXiv preprint no 2103.04931*, 2021, submitted to Springer-Nature AI Reviews.

[20] J. Piaget, *The Psychology of Intelligence*. Routledge, 2003.

[21] W.-Z. Dai, Q. Xu, Y. Yu, and Z.-H. Zhou, "Bridging Machine Learning and Logical Reasoning by Abductive Learning," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[22] M. L. Knapp, J. A. Hall, and T. G. Horgan, *Nonverbal Communication in Human Interaction*. Cengage Learning, 2013.

[23] T. B. Brown, D. Mané, A. Roy, M. Abadi, and J. Gilmer, "Adversarial Patch," *arXiv preprint arXiv:1712.09665*, 2017.

[24] R. Féraud and F. Clérot, "A methodology to explain neural network classification," *Neural networks*, vol. 15, no. 2, pp. 237–246, 2002.

[25] Arrieta, Alejandro Barredo and Díaz-Rodríguez, Natalia and Del Ser, Javier and Bennetot, Adrien and Tabik, Siham and Barbado, Alberto and García, Salvador and Gil-López, Sergio and Molina, Daniel and Benjamins, Richard and others, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information fusion*, vol. 58, pp. 82–115, 2020.

[26] S. B. Kotsiantis, "Decision Trees: a Recent Overview," *Artificial Intelligence Review*, vol. 39, no. 4, pp. 261–283, 2013.

[27] M. Hassaballah and A. I. Awad, *Deep Learning in Computer Vision: Principles and Applications*. CRC Press, 2020.

[28] N. Akhtar and A. Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," *IEEE Access*, vol. 6, pp. 14 410–14 430, 2018.

[29] D. Heaven *et al.*, "Why deep-learning AIs are so easy to fool," *Nature*, vol. 574, no. 7777, pp. 163–166, 2019.

[30] X. Cai and D. C. Wunsch, "Computer Go: A grand challenge to AI," *Challenges for Computational Intelligence*, pp. 443–465, 2007.

[31] M. AlQuraishi, "AlphaFold at CASP13," *Bioinformatics*, vol. 35, no. 22, pp. 4862–4865, 2019.