Open Peer Review on Qeios

Strong Machine Learning: a Way Towards Human-Level Intelligence

Danko Nikolic

Funding: My company Robots Go Mental works on strong ML technology.
Potential competing interests: No potential competing interests to declare.

Abstract

Machine learning has achieved remarkable success with deep learning technologies. However, these methods are often inefficient in terms of resources; they require large datasets, many parameters and consume much computational power. In this paper, I define a general strategy for machine learning, named *strong machine learning*, which aims to create resource-effective machine learning models. Under strong machine learning fall all the approaches that learn inductive biases during an initial phase and later apply those inductive biases to make models more effective learners. Several strong machine learning methods already exist and some are very popular exactly due to their effectiveness. However, strong machine learning is in its infancy and a lot more can be done. In order to further advance AI, we need to direct our effort toward developing even better, more powerful strong machine learning methods.

Danko Nikolić[1,2,*]

[1] *Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Germany*

[2] *Robots Go Mental UG, Germany*

[*]Correspondence to: Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, 60528 Frankfurt am Main, Germany. E-mail address: danko.nikolic@gmail.com.

Introduction

Machine learning has made remarkable progress in the last decades. Machine learning algorithms can learn from data and perform various tasks that otherwise require human intelligence or creativity. These factors have enabled a number of creative applications in various domains. For example, machine learning algorithms can generate realistic images of faces, landscapes, or artworks (e.g., Goodfellow et al., 2014). They can also produce coherent and fluent texts for different purposes, such as summarization, translation, or storytelling (Brown et al., 2020). Moreover, they can power autonomous driving systems that can navigate complex and dynamic environments (Grigorescu et al., 2020). Furthermore, the algorithms based on machine learning can play board games such as chess or Go at a superhuman level, even surpassing the best human players (Silver et al., 2016). These are just some of the examples of the amazing applications that machine learning has made possible.

However, machine learning algorithms also depend on data and computational power to train and run effectively. They need large and diverse data sets that can provide them with enough examples and feedback to learn from. Unfortunately, these techniques are approaching the limits of what is possible with the current machine learning technology (Thompson et al., 2020; Kaplan et al., 2020; Meir et al. 2020). Improving these applications in terms of accuracy and intelligence is increasingly challenging due to the exploding demands on data and model sizes. Machine learning models are excessively data-hungry the biggest models having close to or even more than a trillion parameters. Both data and model sizes have potentially reached the maximal acceptable levels.

Moreover, the current machine learning approaches may be suboptimal and thus, resource-wasteful. For each skill that a machine learning model learns, there is possibly a much smaller model that can achieve the same or better performance. However, our current technology for learning cannot find such smaller models. This is especially a problem if data are insufficient. An extreme example is the algorithms for multiplying two numbers: There are highly optimal multiplication algorithms created by humans. And there are also deep learning models that learned how to multiply solely from examples of correct multiplications. Critically, deep learning cannot ever discover the optimal algorithms. Deep learning is limited to highly suboptimal, resource-hungry models. In contrast, a human mind can effectively extract a multiplication rule from a small number of examples and apply it.

One of the reasons why we need to improve machine learning technologies is that human (and animal) brains still have an edge over machines in many domains of intelligence. Granted, machines can perform very well in some domains that require specific skills or rules, such as chess or calculations. However, humans are still superior in other domains that require more flexibility and creativity, such as language, art, or social interaction. Humans can also learn from just a few examples and transfer their knowledge to new situations that they have never encountered before. Machines, on the other hand, often need a lot of data and feedback to learn and struggle to generalize beyond their training data (Kaplan et al., 2020). Therefore, there is a lot of potential for making machine learning technologies more human-like and more adaptable.

In this paper, I conceptualize an approach to machine learning, which can be referred to as *strong machine learning*. This approach aims to overcome the limitations of the current machine learning approaches, which depend on large amounts of data and parameters to achieve intelligence. Instead of following this path, we can explore technologies that enable machines to learn more effectively. Strong machine learning enables machines to use less data and resources, and to leverage their prior knowledge and experience to learn new skills and tasks. Strong machine learning makes machines more adaptable and creative, and more similar to human or animal learners.

Several methods for strong machine learning already exist. In the present paper, I am trying to unify them under the same concept. What may seem as unrelated methods may have something important in common: all these methods may be forms of strong machine learning. The goal is to put these efforts under a common vision, such that further efforts in the same direction are facilitated. Also, these efforts should be guided by identifying elementary requirements for an activity to be categorized as strong machine learning.

The concept of strong machine learning is inspired by John Searle's notion of strong artificial intelligence (strong AI). Searle distinguished between strong AI and weak AI. He defined strong AI as machine intelligence that works in ways similar to that human or animal intelligence. He claimed that only strong AI would have the same ability to extract the meaning and to understand of the world as humans and animals do. He also challenged the idea that weak AI, which is machine intelligence that can only simulate human or animal intelligence, could ever achieve strong AI. To this end, he introduced the famous Chinese room thought experiment (Searle, 1980). One goal of strong machine learning is to develop technologies that are more similar in intelligence to humans, and eventually reach strong AI, as Searle envisioned it.

Weak vs. strong machine learning

The difference between weak and strong machine learning lies in how they achieve higher levels of intelligence. Suppose we have an existing model with a certain performance, and we want to improve it. We can measure the improvement by the number of categories classified or by the overall accuracy of the model. How do we achieve this improvement? If we add more raw resources, such as more data or more parameters in the model, then we are using a weak machine learning approach. On the other hand, if we use the knowledge from the previous learning and transfer it to be more effective in the new learning, then we are using a strong machine learning approach. Therefore, strong machine learning is simply defined in terms of the resources used: if the primary resource is not a raw ingredient (data, parameters) but a more refined ingredient, which is previously acquired knowledge, then we are applying strong machine learning.

Starting a machine learning task, not from scratch, but with prior knowledge can be beneficial (Bozinovski & Fulgosi, 1976; George Karimpanal & Bouffanais, 2019; Brown et al., 2020; Yuan et al, 2021). Learning can be faster and may require less data. Also, less computational power may be needed. Moreover, models that start with prior knowledge may need fewer parameters to learn a new task than models that start from the beginning. Weak machine learning relies on raw resources and that strong machine learning relies on existing knowledge. The existing knowledge results in either smaller models, or less data needed, or both. These key differences between weak and strong machine learning are listed in Table 1.

Animals and humans are born with capabilities for strong learning. We inherit powerful learning abilities from our ancestors, the necessary knowledge being passed on by genes. This past knowledge allows us to quickly learn various domains such as motor skills, object manipulation, hunting, escaping predators, and so on. Humans are also effective in learning to use language, mathematics, and so on. Clearly, our learning is a lot more effective than that of machines. For example, the amount of language an adult human is exposed to in a lifetime is only a small fraction of the amount of language needed to train a state-of-the-art machine learning model (the ratio is more than one to a million). We can say that this genetic knowledge gives *strength* to our learning. The goal of strong machine learning is to equip machines with similar capabilities.

One of the sources of the strong learning capabilities of our genes is the evolution by natural selection, which has gradually modified our genes over millions of years. Evolution by natural selection is an example of weak learning, as it requires a lot of random trials through mutations, which are not guided by any prior knowledge or feedback. Many generations of individuals and many mutation experiments were needed to evolve new species and eventually, to give rise to human intelligence. This slow and inefficient form of learning can be described as weak, and was required in order to developed our strong learning capabilities.

This brings us to the point that weak and strong machine learning have to be combined. Strong machine learning cannot appear out of thin air. Strong machine learning requires knowledge and for this knowledge, weak machine learning is needed. That is, the creation of strong machine learning must begin with some form of weak machine learning. The knowledge that strong machine learning relies on has to be somehow acquired. This knowledge is acquired through raw resources: data, parameters, intensive computation. Only after a successful acquisition of initial machine learning knowledge, one we start relying on the existing knowledge and hence, start applying the principles of strong machine learning.

Table 1. Properties of weak vs. strong machine learning.

Feature	Weak	Strong
Demand on training data sets	large	small
Demand on parameters	high	low
Inductive biases	scarce	abundant
Previous knowledge	scarce	abundant
Generality of learning capabilities	high	low
Domain of application	broad	narrow

The challenge for engineers or scientists is to devise techniques for strengthening machine learning, for evolving weak machine learning into strong ones. These techniques will require two sets of methods: methods for acquiring knowledge by weak machine learning, and methods for applying this knowledge later in a strong manner. Thus the engineers first need to devise methods for training i.e., the learning curriculum, the training steps, and so on. Once learning is performed successfully, then they can begin with strong machine learning. Strong machine learning is a discipline that devises methods for teaching machine learning systems to become smarter learners.

Inductive biases: The key to strong machine learning

Strong machine learning builds on several theoretical concepts in machine learning and cybernetics. One of these is the concept of *inductive biases*. Inductive bias is a term used in machine learning to describe the set of assumptions that a learning algorithm makes to predict the outputs of new inputs. Critically, the data alone is not enough to learn the best output for every possible input. The learning algorithm needs to have some prior knowledge or 'beliefs' about the nature of the problem and the solution. These prior knowledge or beliefs are commonly referred to as the inductive bias of the algorithm (Michell, 1980).

Inductive bias influences the tendency of a machine learning model to learn certain types of relationships between variables. For example, a model made of linear equations is good at learning linear relationships between variables, but not sinusoidal ones. In contrast, a model made of sinusoidal equations has a strong tendency to learn sinusoidal patterns in time series and has difficulties with linear relationships. One can say that these two types of models are each 'biased' towards learning certain types of patterns.

A well-chosen model for a certain task implies selecting a model with the right inductive biases for that task. Convolutional neural networks are better at learning to detect objects in images than vanilla deep learning models. This is because convolutional models have the appropriate inductive biases – namely, the convolutional layers. Similarly, transformer models have advantageous inductive biases for learning to complete sequential data.

Models with incorrectly chosen inductive biases have difficulties. They can often also learn the tasks that they are not suited for but they require more parameters and more data for training. For example, linear equations can approximate sinusoids to an arbitrary precision; one just needs many linear equations and thus many parameters need to be estimated. Similarly, a Fourier Transform – a model for approximating time series with sinusoids – can describe any arbitrary signal, not only sinusoidal signals. Even rectangular signals can be described by sinusoidal shapes; the problem is that the number of parameters in the model needs to be large. Often, the number of parameters used is the same as the size of the original data. Similarly, a simple (vanilla) deep learning model can learn to detect objects from images without applying convolution. However, such a model will necessarily require many more examples and parameters and thus, computational time compared to a model that employs convolution. If inductive biases are chosen well, the resulting models not only learn faster but can also be much smaller.

The current best models for computer vision or for language can also be considered as not being the most optimal models. They are just more optimal than simple neural networks but they are not the ultimate models that may ever exist. A lot more efficient models, with much better inductive biases, may be possible. We just do not know them. As in the above example of multiplication, millions of times more effective models for vision and language may exist.

The fact is that human DNA requires less than one gigabyte to be stored. And this DNA drives the human brain to deal with both vision and language unprecedently well. This indicates that there are potentially inductive biases that are a lot more effective than anything machine learning has to offer today.

The question is then how strong machine learning can be employed to discover these more effective inductive biases. Traditionally, inductive biases (linear, sinusoidal, convolution, attention in transformer models, etc.) are added to models by human engineers and scientists. These inductive biases are based on human insight. Strong machine learning can be understood as an approach in which machines are made able to learn inductive biases by themselves, without a direct human design of those inductive biases. Humans only define the conditions to learn inductive biases; machines extract the actual inductive biases from data. Thus, machine learning becomes strong once it has its own inductive biases learned earlier. A machine learning process learns inductive biases and stores them in a form for later use and effective learning.

A brief overview of several existing strong machine learning techniques

There are several existing methods in machine learning that can be qualified as strong machine learning. These methods begin with weak machine learning and then transfer the acquired knowledge to achieve strong machine learning. In other words, these techniques first learn and then apply inductive biases. The present list is not exhaustive; there are likely many other methods not mentioned here.

1) Transfer learning (a.k.a., pre-training)

Transfer learning is by far the most popular and most commonly used technique that can be characterized as having the properties of strong machine learning. This is also likely the most simple form of strong machine learning. Transfer learning is based on pre-training a deep learning network on a related task (Bozinovski & Fulgosi, 1976; George Karimpanal & Bouffanais, 2019). A network is pre-trained from scratch on a large dataset and then the parameters of the trained network are used as a starting point for the next training task. The state of the parameters serves as an inductive bias for subsequent learning. The later task usually can be achieved with a lot smaller training dataset. In other words, it would not be possible to successfully train a network from scratch on the later task with small amounts of data if it was not first pre-trained on a related task with a large amount of data. Thus, the learning has been transferred from one task to another, resulting in a reduced need for resources in later learning stages.

Transfer learning, also known as pre-training, is indispensable for today's applications of machine learning. Pre-trained models are widely available and frequently used for example, in computer vision (Yuan et al, 2021) and natural language processing (Brown et al., 2020) among others. Transfer learning enables a model to achieve higher levels of intelligence and adaptability by using multiple stages of learning. The transferred weights and biases play the role of inductive biases operating during subsequent learning.

2) One-shot learning

Lake et al. (2015) reported a modeling approach in which they attempted to mimic human concepts. Their model could learn a 'concept' of handwritten characters. This included not only the visual appearance of the characters but also the strokes needed to write a character. Most characters require multiple strokes to be written and the strokes need to be executed in a specific order and direction. The model first learned from a sample of writing systems the principles of writing. This can be qualified as the initial training phase of weak machine learning. After that, the model acquired a 'concept' of writing and was able to learn new handwritten characters from completely new writing systems. The model became now a lot more effective; it could learn from a single example. Hence, this approach can be referred to as one-shot learning. This later stage can be qualified as strong machine learning. The model used its 'concept' of writing as the inductive bias to learn new characters. The authors argued that this learned capability to learn quickly is similar to how humans acquire and use concepts. The work of Lake et al. inspired other researchers to explore more methods for one-shot learning (Fei-Fei et al., 2006; Vinyals et al., 2016; REFs).

3) Zero-shot learning

In some cases, models can deal with new classes or new problems without any additional learning i.e., without updates of their parameters. For example, a model may be trained on images of animals and on texts describing those images. We can have a case in which a model was not trained on zebras and yet it can recognize zebras if told that they look like striped horses. This is possible because the model learned the visual appearance of horses and also learned what it means for an animal to be striped by learning for example about tigers. The initial training on horses and tigers can be considered weak machine learning. The later application in which the model recognizes striped horses, can be considered a strong application of knowledge acquired previously. Now new data are needed for training and yet the model can perform novel tasks by providing a single sample of auxiliary information. The knowledge of the model serves as the inductive bias to recognize zebras even without any updates of weights. Zero-short learning is popular in image classification and natural language processing (e.g., Brown et al., 2020).

4) Guided Transfer Learning

Guided transfer learning is a technique that expands on the capability of transfer learning. While traditional transfer learning carries over the parameters of the model, guided transfer learning carries over information on which of those parameters should be allowed to change during subsequent training and which parameters should better stay unchanged. Guided transfer learning also starts with a weak machine learning component in which a set of scout networks is trained on easy tasks and with sufficient data. During this process, knowledge is collected about which of the parameters tend to change and which tend to remain unchanged in subsequent learning. This knowledge about parameter changeability is then transferred to later tasks and is thus used as inductive biases. Much like for transfer learning, the initial tasks with scout learning and a later learning task need to be related for guided transfer learning to work well.

Guided transfer learning further reduces the need for data and computation beyond what transfer learning can reduce. Wherever the limits of traditional transfer learning are, guided transfer learning pushes these limits further. Guided transfer learning makes the inductive biases even stronger i.e., even more specialized for a given type of task than what can be done by traditional transfer learning. Guided transfer learning has been shown to help with extremely small amounts of data, with reducing catastrophic interference (forgetting old information by novel learning)(Nikolić et al., 2023), and with tasks normally extremely difficult for deep learning such as finding solutions to problems based on a mixture of logical operations (OR, AND, XOR)(Schmidhuber and Hochreiter, 1996; Mansour, 1994; Linial et al., 1993; Nikolić, 2023).

The tradeoff between generality and specificity

There is a tradeoff associated with every inductive bias. Inductive biases only help machine learning perform better within a certain domain. Critically, inductive biases make it harder to perform well in other, unrelated domains. For example, a convolutional network will be effective in computer vision but will very much struggle with sequences of words in natural languages. By making learning effective in one domain inductive biases make it more difficult to learn in other domains.

This leads to a tradeoff between the generality of what a machine learning algorithm can learn in principle on one hand and the specificity for efficient learning that the inductive biases offer. More powerful inductive biases lead to more specificity which means that such a model can be applied to a narrower set of problems. For example, we may not only narrow down a model to computer vision but to even more specific vision problems. For example, a machine learning algorithm may become an expert at learning to recognize animals only or to recognize hand-written characters only. Such more specialized models become even better learners of these domains but, at the same time, have even more trouble learning outside those domains, for example, to detect cars from images. The stronger the capability to learn within a certain domain, the narrower the domain of learning. In other words, there is a clear tradeoff between the learning power of strong machine learning which is characterized by high specificity, and the range of the problems to which a given strong machine learning algorithm can be applied, which is characterized by reduced generality. The higher the specificity, the lower the generality.

As an extreme example of specificity, we can consider a human-made model in a form of a single equation that describes a certain property of the universe: $E = m\,c^2$. This equation is a simple 'model' with a single parameter, c. In principle, the model can be applied to different universes with different speeds of light. In each of these universes, one can 'train' the model to predict energy (E) given the mass (m). The model just needs to learn one parameter which is the speed of light in that universe, c. Thus, a single data point with E and m may be enough to learn the parameter c and thereby enable to model to calculate energies correctly for a new universe. This model is an extreme example of domain specificity: The model can learn one thing only – converting mass to energy – but it can learn these conversions most effectively, from a single data point. This model cannot be applied to anything else, but only to this one problem. This 'model' was not learned through strong machine learning, but was created by human insight (Albert Einstein's insight, to be precise), which is a strong learning capability of the human brain. Nevertheless, the famous Einstein's equation illustrates how extreme specialization comes with an extreme learning efficiency but necessarily also with an extreme lack of generalization.

This tradeoff between generality and specificity is related to the no-free-lunch theorem in machine learning (Adam et al., 2019). The no-free-lunch theorem states that there is no single best learning algorithm that can perform well on all possible problems. Different algorithms have different strengths and weaknesses, and their performance depends on the characteristics of the problem and the data. This implies that there always will be a tradeoff between generality and specificity: An algorithm that is good at one domain is necessarily bad at some other domain. Strong machine learning is subjected to the no-free-lunch theorem. Strong machine learning models have to sacrifice generality for specificity in order to achieve high learning efficiency in terms of data and model sizes.

The relationship between strong machine learning and human intelligence

A possible strategy to create AI with human-like intelligence is by further developing strong machine learning techniques. These techniques should emulate the learning abilities of humans and animals. Living organisms are strong learners who can quickly acquire new concepts, generalize from a single event, and adapt to novel situations with minimal data and guidance. To achieve similar performance in machines, we likely need to enhance the existing machine learning models making by providing them with *strength*. They should be able to learn human- and animal-type of tasks with less data and fewer parameters.

For such developments, we can draw inspiration from how human brains are composed of specialized modules each being optimized for a certain task. The cerebral cortex consists of about 46 areas each being specialized for a specific function playing a role in vision, language, motor commands, and so on (Strotzer, 2009; Ardila et al, 2016; Kawachi 2017;). Each of these areas is probably a strong learner that can learn from relatively small amounts of data. After we are born, and as we learn, we add even more specializations to our brain modules. This means that we also lose some generality. Adults cannot learn as flexibly as babies can. But adults can learn more effectively than babies within the specializations that they developed since.

What makes a difference between human and animal intelligence is probably the inductive biases that we are already born with. These inductive biases may be expressed in the total number of brain areas, the general connectivity of those brain areas, and the learning rules by which we further learn after birth.

The strategy to approach human intelligence in AI should therefore, not only be to increase the number of modules but also to apply strong machine learning techniques to train those modules to learn effectively for their specific domains. Recently, I proposed that these modules should apply transient selections of pathways (Nikolić, 2023) as a part of mechanisms for the effective implementation of cognitive operations, founded in the theory of practopoiesis (Nikolić, 2015). This creates an open field for the application of strong machine learning, as inductive biases can be stored not only in connections between neurons but also in those mechanisms that transiently open and close pathways.

An important contrast can be made between the approach to intelligent machines based on strong machine learning and another one based on Artificial General Intelligence (AGI) (Goertzel, 2014). Both approaches share the vision of reaching human-level intelligence. However, they imply different ways to reach that goal. AGI emphasizes a search for one (or more) ‘general’ algorithms that can apply to any domain or any problem, without much need for specialization (Hutter, 2004). The central idea of AGI is generality. In contrast, strong machine learning emphasizes specialization; it calls for developing effective specialized algorithms that can learn specific domains well. Instead of attempting to find a general algorithm, we may be better off searching for effective forms of strong machine learning. The only general ‘algorithm’ may be a widely applicable method for learning inductive biases for the various specialized modules of strong machine learning. Maybe this hypothetical general method for learning specializations is what AGI is looking for.

The next frontier is strong machine learning

AI development can clearly be boosted by advancing strong machine learning techniques. At the beginning of solving a new problem, deep learning always relies on weak machine learning (more data and parameters). As we have seen, however, there are a few machine learning techniques that are being regularly applied to deep learning that can be classified as strong.

As of today, AI still relies too much on weak machine learning: the addition of data and parameters to the models. Strong machine learning is still in its infancy. The problem with weak machine learning is that one quickly exhausts the potential of that approach. The models quickly grow to the maximum sizes that are feasible in terms of computational costs and environmental concerns. Also, the models quickly exhaust all of the training data and the generation of new data becomes prohibitively expensive.

As a consequence, in some cases, only big organizations with large resources are able to push weak machine learning models further. Smaller organizations are unable to compete. For example, large language models can only be created by big organizations that have the resources to build and maintain them. This gives them an unfair advantage over smaller players and reduces the competition and creativity in the field of AI. The concept of strong machine learning, promises to create more efficient and intelligent models that can learn from less data and with fewer parameters. This promises to level the playing field and allow more people and organizations to participate and contribute to the advancement of AI.

Another reason for making models smaller is the need for intelligent machines at the edge. Robotics in general and especially the field of autonomous driving are clear examples of this. Huge models that run on supercomputers perform better in those robotic tasks than small models that can be actually fitted under the car hoods. Autonomous driving and robotics in general are in need of smaller models that nevertheless exhibit high performance as the big ones. This can potentially only be achieved through strong machine learning.

Equally so is the scarcity of data a reason for strong machine learning. Large language models could be created in part due to the easy access to a vast amount of texts stored in an electronic form, covering almost everything that has ever been written by mankind. This played well with another property of language: Language is highly efficient in storing information – the language is highly compressed. Images, on the other hand, are much more wasteful with resources. For example, a language-based story and a picture-based movie may be telling the same story but may differ by a factor of a million in memory requirements. This high memory demeans for images has implications for the further development of AI based on computer vision: The use of images in a weak machine learning manner reaches technical limits much faster than language does. Building 'large vision models' that would generate movies with similar quality and flexibility as language models generate texts would require resources that probably not even the largest organizations can afford. This makes pressure to develop strong machine learning techniques for computer vision. Only strong machine learning may overcome these limitations.

Conclusions

Strong machine learning is an approach to machine learning that focuses on learning inductive biases. Inductive biases are assumptions that are built into the learning algorithm that help the algorithm learn more efficiently in a given domain.

The idea is that by learning these inductive biases, machines can later learn more effectively, with less data and with smaller models.

We need to further develop techniques for machines to learn their own inductive biases. Instead of focusing on building large powerful models, we should focus on advancing the capabilities for strong machine learning. Several strong machine learning techniques already exist. However, more work is needed in this direction. The field of strong machine learning is in its infancy.

While developing ever better strong machine learning algorithms, we have to keep in mind the generality-specificity tradeoff. Strong learners are specialized. A challenge that comes with strong machine learning is how to combine multiple specialized learners into a fully functioning system that exhibits high levels of intelligence.

The ultimate ambition of strong machine learning is to help create machines that match human intelligence i.e., to help develop strong AI. The idea is that by providing machines with the same inductive biases that humans have, we can create machines that are capable of learning and reasoning in the same way that humans do.

References

- Adam, S. P., Alexandropoulos, S. A. N., Pardalos, P. M., & Vrahatis, M. N. (2019). No free lunch theorem: A review. *Approximation and optimization: Algorithms, complexity and applications*, 57-82.
- Ardila, A., Bernal, B., & Rosselli, M. (2016). How localized are language brain areas? A review of Brodmann areas involvement in oral language. *Archives of Clinical Neuropsychology*, *31*(1), 112-122.
- Bozinovski, S., & Fulgosi, A. (1976). The influence of pattern similarity and transfer learning upon training of a base perceptron B2. (original in Croatian) Proceedings of Symposium Informatica 3-121-5.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P.,... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.
- Fei-Fei, L., Fergus, R., & Perona, P. (2006). One-shot learning of object categories.*IEEE transactions on pattern analysis and machine intelligence*, *28*(4), 594-611.
- Goertzel, B. (2014). Artificial general intelligence: concept, state of the art, and future prospects.*Journal of Artificial General Intelligence*, *5*(1), 1.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S.,... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.
- Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, *37*(3), 362-386. Hutter, M. (2004). *Universal artificial intelligence: Sequential decisions based on algorithmic probability*. Springer Science & Business Media.
- George Karimpanal, T., & Bouffanais, R. (2019). Self-organizing maps for storage and transfer of knowledge in reinforcement learning. *Adaptive Behavior*, *27*(2), 111-126.
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R.,... & Amodei, D. (2020). Scaling laws for

neural language models. *arXiv preprint arXiv:2001.08361*.

- Kawachi, J. (2017). Brodmann areas 17, 18, and 19 in the human brain: An overview. Brain and nerve= Shinkei kenkyu no shinpo, 69(4), 397-410.
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332-1338.
- Linial, N., Mansour, Y., & Nisan, N. (1993). Constant depth circuits, Fourier transform, and learnability.*Journal of the ACM (JACM)*, *40*(3), 607-620.
- Mansour, Y. (1994). Learning Boolean functions via the Fourier transform. In*Theoretical advances in neural computation and learning* (pp. 391-424). Boston, MA: Springer US.
- Meir, Y., Sardi, S., Hodassman, S., Kisos, K., Ben-Noam, I., Goldental, A., & Kanter, I. (2020). Power-law scaling to assist with key challenges in artificial intelligence. *Scientific reports*, *10*(1), 19628.
- Mitchell, T. M. (1980). The need for biases in learning generalizations. CBM-TR 5-110, New Brunswick, New Jersey, USA: Rutgers University.
- Nikolić, D. (2015). Practopoiesis: Or how life fosters a mind.*Journal of Theoretical Biology*, *373*, 40-61.
- Nikolić, D. (2023). Where is the mind within the brain? Transient selection of subnetworks by metabotropic receptors and G protein-gated ion channels. *Computational Biology and Chemistry*, *103*, 107820.
- Nikolić, D., Andrić, D., & Nikolić, V. (2023). Guided Transfer Learning. arXiv preprint arXiv:2303.16154.
- Schmidhuber, J., & Hochreiter, S. (1996). Guessing can outperform many long time lag algorithms. Technical Note IDSIA-19-96, IDSIA, 1996
- Searle, J. R. (1980). Minds, brains, and programs.*Behavioral and brain sciences*, *3*(3), 417-424.
- Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G.,... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484-489.
- Strotzer, M. (2009). One century of brain mapping using Brodmann areas. Clinical Neuroradiology, 19(3), 179.
- Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The computational limits of deep learning.*arXiv preprint arXiv:2007.05558*.
- Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one shot learning.*Advances in neural information processing systems*, *29*.
- Yuan, L., Chen, D., Chen, Y. L., Codella, N., Dai, X., Gao, J.,... & Zhang, P. (2021). Florence: A new foundation model for computer vision. *arXiv preprint arXiv:2111.11432*.