Does Deep Learning Have Epileptic Seizures? On the Modeling of the Brain

Damien Depannemaecker, Léo Lopez, Christophe Gauld

 doi:10.20944/preprints202207.0229.v1

Does deep learning have epileptic seizures? On the modeling of the brain

Damien Depannemaecker[1,*], Léo Pio-Lopez[2,*], Christophe Gauld[3]

[1] Paris-Saclay University, Centre National de la Recherche Scientifique (CNRS), Institute of Neuroscience (NeuroPSI), 91198 Gif sur Yvette, France
[2] Allen Discovery Center, Tufts University, Medford, MA, USA
[3] Child Psychiatry Departement, University Hospital Lyon, 59 Bd Pinel, 69000 Lyon, France
[*] These authors contributed equally

damien.d@cnrs.fr

Abstract

If the development of machine learning and artificial intelligence plays a role in many fields of research and technology today, it has a special relationship with neurosciences. Indeed, historically inspired by our knowledge of the brain, deep learning shares some vocabularies with neurosciences and can sometimes be considered a brain's model. Taking the particular example of epileptic seizure, which can develop in any biological neural tissue, we raise the question if and how the models used for deep learning can capture or model these pathological events. This particular example is a starting point to discuss the nature, limits, and functions of these models, and what we expect from a model of the brain. Finally, we argue that a pluralistic approach leading to the integrated coexistence of different models is necessary to study the brain in all its complexity.

Keywords: Neurosciences, Artificial Intelligence, Brain Modeling, Deep Learning

1 Introduction

Over the past decades, immense enthusiasm for deep learning (DL) has brought new insight into our understanding of how the brain works. It has found a large field of application in neurosciences [1–3]. In many cases, DL is used as a tool for data analysis. Indeed, using it to optimize some operations such as classification on large and complex datasets is very practical and useful, and very important results have been achieved in different fields ranging from biological data analysis, object or face recognition to robotics, just to name a few [4–6].

However, a growing literature seeks to use DL as a model of brain functions [1, 2]. In this literature, by digging into the machinery of the DL, a number of researchers aim to understand how the human brain works, and in this way of thinking, DL has been considered a model of the functions of the human brain. In this vein, the comparison of the brain and DL is not only a metaphor but presents itself as a programmatic challenge: DL could develop until being able to simulate all levels of brain functions. By refining itself, DL could simulate consciousness [7] or social interactions [8]. In such a computationalist approach, the brain is conceived as a set of algorithms, and the progression of the understanding of these algorithms will be able to approximate and reproduce (at the phenomenological level) the functions of the human brain. But the question remains: what do we mean by modeling the brain?

 doi:10.20944/preprints202207.0229.v1

In this article, we aim to demonstrate that DL cannot be considered a complete model of the brain. To achieve this aim, we use a methodological tool corresponding to the counterfactuals: if the brain no longer functions as it should function, can DL, in parallel, no longer work the same way? This question avoids any fallacious and contingent comparison or analogy, and avoid especially the list of programmatic arguments explaining why the DL will be able to raise its current technical limits. In other words, if DL is considered a model of the brain, pathological states such as epileptic seizure (a stereotyped and simplistic disorder of brain disease) may also exist as an intrinsic property of DL. Epileptic seizures are related to groups of neurons that synchronously generate repeated trains of action potentials, precluding the functions considered as normal in certain areas of the human brain. Certainly, seizures can occur in any neuronal system [9], and models can reproduce these paroxysmal dynamics from a single neuron model to the whole brain [10]. Thus, the question should have been "Is DL able to model a pathological human brain affected by an epileptic seizure, in the sense of a synchronous and repeated discharge precluding the functions considered as normal in certain areas of the human brain?". However, as we will see right away, Artificial Neural Network (ANN) used for DL cannot have epileptic seizures in this sense. The purpose of this article is to discuss these apparent triviality reasons and if the DL can be a model of human brain functions. We will provide clues to explain how DL should be considered and interpreted in the context of this comparison with brain functions.

2 Why deep learning does not have seizures?

The neuron model (also called "node" in machine learning) used in DL can be described as follows. The node receives inputs from external data or from the previous nodes. Input values are weighted (multiplied) by values of "synaptic weights", and then, they are summed before going through a so-called activation function. The value obtained at the output of this function can be used as an input value for the next node. This is a static description using multiplications, a sum, and an activation function. This description is sufficient to reproduce a "functional" aspect of the processing of incoming information by a neuron. It has been called a formal neuron [11]. Once formed into a network for a particular application, the "learning rules" determine the evolution of synaptic weights depending on the output obtained. The historical and most used method for updating synaptic weights is the backpropagation [12], but others such as evolutionary computation [13] or direct feedback alignment [14] can be very efficient. These rules are applied to the different layers of the network. Indeed, networks are constituted of different layers of nodes, often projecting from one to the other. Different architectures have been developed, as for example recurrent neural networks (RNN), convolutional neural networks (CNN), generative adversarial networks (GAN), and autoencoders, for a large range of applications [15].

In the brain, seizures are characterized by paroxysmal electrophysiological activity. In DL, it would therefore be tempting to think that the state corresponding to the seizure could appear when all the nodes have a very high activation level (i.e. all activation functions are at their maximum). However, such an activation level is only one possible state of the network, like any other, not presenting any pathological aspect. More precisely, whether or not the state of the network allows the desired output, the model is considered relevant and useful for an assigned task, or the synaptic weights are updated until the desired output is obtained. But in any case, the differentiation between normal and pathological in human epileptic seizures cannot be transposed to the question of normal and pathological in DL. DL can only be a seizures classifier, detector, or eventually be used for prediction, given some specific input data [16–18].

To go further in the comparison between these two completely different systems (brain seizures and "DL seizures"), we have to highlight that in the brain, seizures can be propagated while synaptic communication fails [19, 20]. Indeed, homeostatic disturbances, leading to unhealthy con-

 doi:10.20944/preprints202207.0229.v1

centrations, will disturb membrane excitability, and reinforce the inability of glial cells to help to contain the seizure. In addition to the electro-diffusion mechanisms, there are also ephaptic interactions that participate in the propagation [10,19–21]. Many complex interactions and mechanisms are involved, going beyond the simplistic reductionist framework of interacting neurons only through synapses. Thus, seizures can spread in the brain even if all synaptic connections fail. In DL, such spreading is not conceivable.

It should be added two points precluding comparison of brain seizures and DL seizures. First, seizures are a dynamical phenomenon [9, 22]. Indeed, time evolution is a central element. But unlike the brain, the evolution of the state and the synaptic weights in DL takes place step by step. If a dynamical process does exist, it is a discrete process. This time is therefore that of the model, and not the real biological time that the model would have captured. Thus, the scale of the dynamics is not intrinsic to the nature of the model. However, a correspondence between the interval between steps and the real-time of the biological process can be established. In order to consider these temporal aspects, models based on temporal differential equations (i.e., models from the field of dynamical systems) could be more appropriate to capture the seizure phenomena at different scales [10,23]. Such models can also include learning mechanisms [21,24,25]. However, these models are not initially developed to achieve complex classification tasks, while DL captures this aspect of the brain's capacities.

Secondly, brain epileptic seizures correspond to a synchronous discharge at the level of the biological neurons, and more particularly at the level of biological tissue (e.g., ion channels). However, the DL does not necessarily have such a physical implementation, but only constitutes information exchange networks. The question that arises is that of DL implemented within, for example, a robot: could this physical (e.g., electrical) implementation of the DL have epilepsy, corresponding to the fact that the electrical system "burns"? We do not believe it. For instance, at the molecular level, the brain relies on an exchange of ions depending on the state of the system, and both together provide information to the brain [26–28]. According to Marr description [29], this happens at the implementation level. DL is based on an exchange of (numerical) values, at the algorithmic and computational levels. In terms of pathology, if an overflow of information is present in the brain, there can be a synchronous discharge, i.e., an epileptic seizure. Conversely, there is no change at the implementation level for DL, so if too much information is presented to the network, there will be no synchronous burst, but classical information processing. However, this finding does not mean that the (unimplemented) DL model itself may have "seizures". Apart even from the problem of embodied DL models, this impossibility of DL being overtaken by a flood of information removes any comparison between DL and the brain.

Considering all these differences, should we think that DL are bad models? Of course not, but when and why can it be considered a good model? First, there are just not adapted to capture seizure phenomenon. Their function is diverted being used for a task that cannot correspond to the nature of such models. However, if DL cannot grasp pathological states, it cannot grasp the counterfactual model of the brain, which is the epileptic brain. Thus a correspondence between the model and the phenomenon depends on the state of the considered object. Indeed, if the communication between neurons is altered (i.e. during epilepsy), then DL failed in capturing this aspect. In addition, one particular aspect of DL must be mentioned. The model itself, based on formal neurons, constitutes networks with particular properties and architectures, and the complex functions approximated by this type of network, after their implementation.

The ability to view DL as a good model of the brain would therefore be limited to the study of a healthy brain, and only that. In this case, if the DL may be a model of the brain, the question is therefore to know what this model captures. If the DL can process information and approximate functions that the brain is able to perform, does it do so the same way?

Indeed, if we consider the DL beyond a black box that provides us interesting outputs depending

 doi:10.20944/preprints202207.0229.v1

on the inputs, and if we look at what is happening inside, we may wonder how what is there, has a connection with the brain. A recent and fast-developing field of research on explainable neural network offers new directions [30–32]. The aim is to understand and make interpretable what is happening in the network from input to output. This does not mean that the brain works the same way.

Certainly, DL has network properties allowing the processing of information that can help us build knowledge about the brain. If we established the correspondence between (biological) neurons and nodes, it is therefore situated at a computational level and at the neural network scale. These notions of what models capture, and of scale and level are discussed in the next section.

3 What do we expect from a model of the brain?

We propose in this last section two necessary axioms in order to consider DL capable of explaining certain functions of the human brain: the double relativity of the models, and the distinction between scale and levels.

What would mean a model of the brain, whether or not dynamic? The brain is an extremely complex system with a manifold of mutual interactions, from microscopic to whole-brain interactions, with a large number of high-level cognitive functions and the emergence of consciousness. Consequently, due to the impossibility of considering all these levels and phenomena, the construction of a model implies a reduction. It necessarily considers only the relevant elements to a particular phenomenon applied to a particular study objective. Such a contingency has been called the "double relativity": DL are constrained by the context of use and by the objective [33].

The relativity corresponding to the objective of the study questions the usefulness of the model. Does the model bring us knowledge, understanding, and predictions? These questions require specifying exactly what are the functions of a model during its construction. Therefore, functions are limited by the intrinsic nature of models, because it has been designed to reproduce only certain aspect. Epistemological approaches have been able to detail these different aims organized in 21 specific functions [34]. The mains and most frequent epistemic functions of models are explanation (regarding underlying mechanisms, rules, and causality), prediction (interpolation, extrapolation, case-specific prediction), understanding (share a common "vision", mental reconstruction of a considered phenomenon), and data representation and generation. In most situations, DL can be used for data classification, representation, and generation, or for prediction [15,35–37] but hardly for explanation (as it captures too few aspects of the original object of study, the brain). However, there are some efforts in this direction showing for example a possible neural implementation of backpropagation [38,39].

In neurosciences, different scales are considered from molecular, cellular, tissue, and anatomical regions to whole-brain. At each of these scales, it can be considered three different levels of description, that can be understood in terms of Marr's levels [29]. They correspond to the implementation, algorithmic, and computational levels. The first would correspond to the biological substrate, the second to the (complex) relationship between the different elements, and the last to the emerging function. Thus, we have three levels of description that can be captured by the models. These levels of description may apply at different scales. The relationship between scale and level is not trivial, there is maybe not a correspondence such as lower level corresponds to lower scale (and vice versa). For example, there is a cellular scale and a cerebral connections scale: there are levels of Marr allowing to understand a problem in terms of implementation (biological) or computation (function). This aspect should be considered while interpreting results from modeling studies. Scales and levels undertake to design a phenomenological model (i.e., modeled toward a pragmatic objective) or mechanistic (i.e., modeled towards a resemblance of properties, often biophysical) [40]. Thus, when observing the apparent global behavior of a model, it may be

 doi:10.20944/preprints202207.0229.v1

interesting to look a the underlying scales within the model to explain the emergence phenomenon. However, many different underlying mechanisms may lead to similar apparent behaviors. This is known as neural degeneracy in biology or multiple realizability in epistemology, and it happens even in computational models [41, 42]. These concepts are particularly important because they may strongly limit the capacity for explanation or the generalization of a model.

However, minimal or necessary conditions for a given phenomenon can be found thanks to the model. For example, in the case of seizures, the consideration of two different time scales, a fast one, corresponding to the electrophysiological activity and a slow one, that enables the transition between ictals and inter-ictals periods, can be understood thanks to dynamical systems-based models [9, 10, 23]. This time scale separation does not exist in neural networks used for DL, and it is not appropriate to capture such a phenomenon. This characteristic of time scale separation can be identified in both objects, the biological nervous tissue, and the models. This also opens the question of biological plausibility. As models are non-biological objects that capture some level(s) of description (biological representation) for specific scale(s), models boundaries will necessarily cut or approximate the underlying mechanisms as they are in biology. But they can share, for the considered scales and levels, some common aspects with biology.

It is thus of first importance, to analyze our model through this prism of the optimal extraction of knowledge about the object of study. In the case of DL, the model of synapse used may not inform us about the biological synapses. But this model of synapses plays the role of functional interactions between the nodes, and we can identify from such model this necessity of evolution of the strength of this interaction for the information processing, and, enable to describe the network of interaction between cells. Non-neural tissue may also support different cognitive functions such as learning, memory, or decision-making [43–45].

This aspect relates to the theoretical level (see Fig.1), some models are built to mainly describe cognitive functions or a cognitive architecture (such as the generative adversarial networks (GANs) [37], Generative Pre-trained Transformer 3 (GPT-3) [46], Adaptive Control of Thought-Rational (ACT-R) [47], convolutional neural network (CNN) [48]) while others have the primary objective of capturing the electrophysiological activity of the brain (such as the Hodgkin-Huxley neuron model (HH) [49], the Izhikevich spiking neuron model (Izh) [50], or the virtual brain using mean-field model (TVB) [51]). In our figure, we can observe that we have no models in the upper right corner. We still don't have whole-brain models with the aim to be biophysical and functional. The free-energy principle is an attempt to link both approaches [52]. Artificial general intelligence is also a field where we could see the emergence of such an integrative model [53]. Being able to model these two aspects is particularly challenging, interesting approaches in this direction are based on spiking neural networks designed to reproduce functional tasks.

The existence of double relativity (models are constrained by the context of use and by their objective) and the consideration of separation between scales and levels allows considering a form of scientific pluralism. As this could be modeled within the framework of the Research Domain Criteria (RDoC) proposed by the National Institute of Mental Health [54]. it seems necessary to consider the different functions of the brain at all their scales of understanding. To model each of these functions at all these scales, a large number of DL models can be constructed. However, in order that this matrix could link models reproducing the functions at all scales, it is still necessary to add the different levels of description for each of the functions, for all scales. In other words, pluralism in terms of modeling is necessary to achieve faithful reproduction of the (non-pathological) functions of the brain. A pluralism of approaches (using other forms of models than DL) is necessary to complete the picture at different scales and levels. However, this observation does not remove the possibility that some brain pathology cannot be modeled intrinsically by DL (by modeling the information flow).

Finally, an important point is that the brain contains enough information to develop by itself

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2022 doi:10.20944/preprints202207.0229.v1

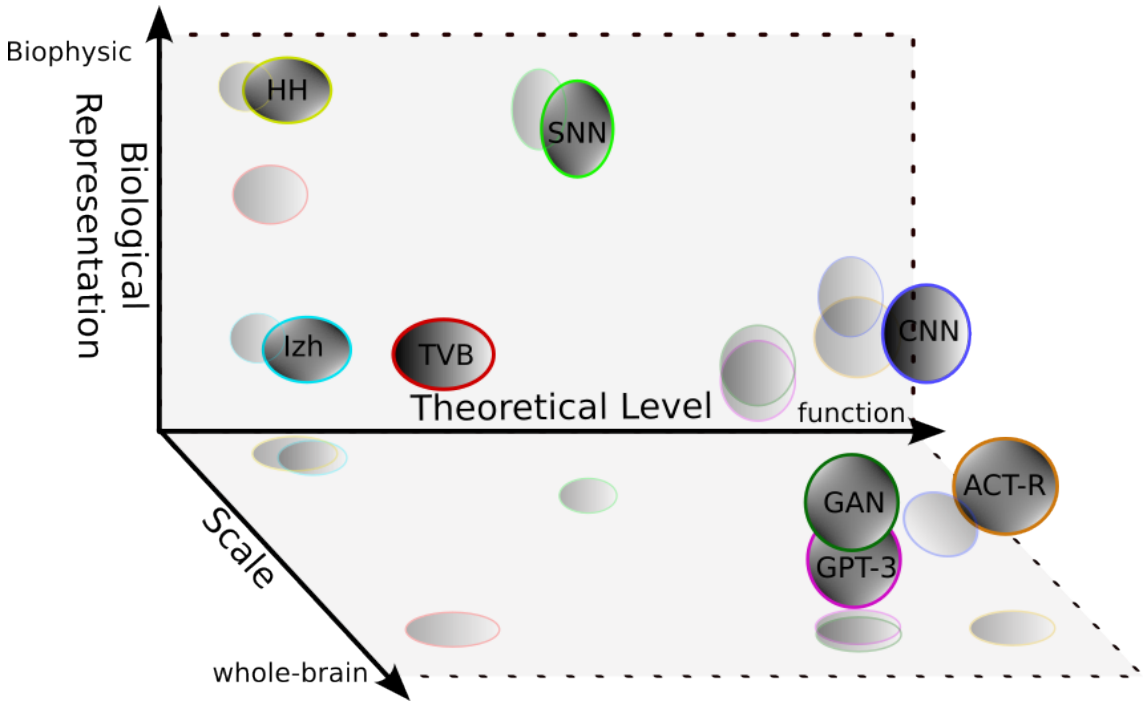

Figure 1: Different models related to the brain are characterized by their biological representation (how much is the description close to the biophysics), the theoretical level (how much the model captures the function), for different scales (from the molecule to the whole brain). Few examples are shown here: (HH) Hodgkin-Huxley neuron model [49], (Izh) Izhikevich spiking neuron model [50], (TVB) The Virtual Brain [51], (GANs) Generative adversarial networks [37], (GPT-3) Generative Pre-trained Transformer 3 [46], (ACT-R) Adaptive Control of Thought—Rational [47], (CNN) Convolutional neural network [48]. Two clusters appear, one integrating models of functions, which do not capture the electrophysiological activity, and on the other part of the model space we have models designed to reproduce the dynamics of electrophysiological activities.

within its environment, and at the same time is a central organ of internal regulation, perception, action, and behavior. But it also can have seizures, migraines, or strokes and can reorganize itself at different scales, while, to a certain extent, maintaining its functions. In silico models, built by humans are not capable of this variety of functions at such different time scales. These models cannot be exactly equivalent to a biological brain, being built differently on a different substrate it will not perfectly reproduce the same functions using the same exact underlying mechanisms. Again, the models are centered on a level of description at a given scale. If DL can be interesting from a computational point of view, it will not explain and understand the underlying biological mechanism at smaller scales (subcellular, molecular). Thus if it may be useful to reproduce, predict or understand certain aspects of the brain, we should not restrict our vision to this approach to consider the brain. Such a broader pluralist vision also enables to improve models by relating them to their neighbors in terms of level and scale and questioning their possible integration. In Fig. 1 we show the relative position of different types of models to each other. It seems particularly difficult to integrate all these models into a single simulation framework. The resulting system would be so complex that it would be extremely difficult to extract any knowledge about the biological brain. Therefore, it seems necessary to consider the coexistence of different models, each developed for

specific objectives. This pluralism is isolationist because explanations are understood at a given level of analysis, relatively impervious to explanations at other levels [55]. The possibility of direct integration between models can only be very local.

4 Conclusion

DL is very powerful to model computational or cognitive aspects of the brain, and also for data analysis. However, to build a theory of the brain, the modeler has to keep in mind the two axioms we described earlier: the double relativity of the models, and the distinction between scale and levels. The corollary is that an explicative model needs to be enough constrained to grasp and explain neural phenomena. DL is a very good example of that, it can model functions such as learning, classification, prediction, or other cognitive functions, but lacks the resolution to model brain pathologies such as epilepsy. Scientific pluralism is therefore a key element for constructing a theory of the brain.

Overcoming this potential problem in the research practices, constant questioning about the function and the boundaries of the used model is absolutely necessary to understand the provided results. Finally, models are built on specific assumptions to answer specific questions. Indeed considering a general model of the brain is very different. We build a model of some aspects of the brain to fulfill given functions. DL offers a specific and computationalist point of view that can be interpreted at different scales and levels. By refining the knowledge of the boundaries of our model, we may better understand their best range of applications from which we would be able to extract a genuine and deeper knowledge about this complex object of study: the brain.

5 Compliance with Ethical Standards

This article does not contain any studies with human participants or animals performed by any of the authors. No funds, grants, or other support were received. The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Saxe A, Nelli S, Summerfield C. If deep learning is the answer, what is the question? Nature Reviews Neuroscience. 2020 Nov;22(1):55–67. Available from: https://doi.org/10.1038/s41583-020-00395-8.

[2] Richards BA, Lillicrap TP, Beaudoin P, Bengio Y, Bogacz R, Christensen A, et al. A deep learning framework for neuroscience. Nature neuroscience. 2019;22(11):1761–1770.

[3] Marblestone AH, Wayne G, Kording KP. Toward an integration of deep learning and neuroscience. Frontiers in computational neuroscience. 2016;10:94.

[4] Yan LC, Yoshua B, Geoffrey H. Deep learning. nature. 2015;521(7553):436–444.

[5] Tang B, Pan Z, Yin K, Khateeb A. Recent advances of deep learning in bioinformatics and computational biology. Frontiers in genetics. 2019;10:214.

[6] Schmidhuber J. Deep learning in neural networks: An overview. Neural networks. 2015;61:85–117.

[7] Reggia JA. The rise of machine consciousness: Studying consciousness with computational models. Neural Networks. 2013;44:112–131. Available from: https://www.sciencedirect.com/science/article/pii/S0893608013000968.

[8] Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon JF, Breazeal C, et al. Machine behaviour. Nature. 2019 04;568(7753):477–486.

[9] Jirsa VK, Stacey WC, Quilichini PP, Ivanov AI, Bernard C. On the nature of seizure dynamics. Brain : a journal of neurology. 2014 8;137(Pt 8):2210–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24919973 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4107736.

[10] Depannemaecker D, Destexhe A, Jirsa V, Bernard C. Modeling seizures: from single neurons to networks. Seizure. 2021 Jun;Available from: https://doi.org/10.1016/j.seizure.2021.06.015.

[11] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics. 1943;5(4):115–133.

[12] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. 1986 Oct;323(6088):533–536. Available from: https://doi.org/10.1038/323533a0.

[13] Miikkulainen R, Liang J, Meyerson E, Rawal A, Fink D, Francon O, et al. Evolving deep neural networks. In: Artificial intelligence in the age of neural networks and brain computing. Elsevier; 2019. p. 293–312.

[14] Nøkland A. Direct Feedback Alignment Provides Learning in Deep Neural Networks. arXiv. 2016;.

[15] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016. http://www.deeplearningbook.org.

[16] Thangavel P, Thomas J, Peh WY, Jing J, Yuvaraj R, Cash SS, et al. Time–frequency decomposition of scalp electroencephalograms improves deep learning-based epilepsy diagnosis. International journal of neural systems. 2021;31(08):2150032.

[17] Ullah I, Hussain M, Aboalsamh H, et al. An automated system for epilepsy detection using EEG brain signals based on deep learning approach. Expert Systems with Applications. 2018;107:61–71.

[18] Sun M, Wang F, Min T, Zang T, Wang Y. Prediction for high risk clinical symptoms of epilepsy based on deep learning algorithm. IEEE Access. 2018;6:77596–77605.

[19] Pumain R, Menini C, Heinemann U, Louvel J, Silva-Barrat C. Chemical synaptic transmission is not necessary for epileptic seizures to persist in the baboon Papio papio. Experimental Neurology. 1985 Jul;89(1):250–258. Available from: https://doi.org/10.1016/0014-4886(85)90280-8.

[20] de Almeida ACG, Rodrigues AM, Scorza FA, Cavalheiro EA, Teixeira HZ, Duarte MA, et al. Mechanistic hypotheses for nonsynaptic epileptiform activity induction and its transition from the interictal to ictal state-Computational simulation. Epilepsia. 2008 Nov;49(11):1908–1924. Available from: https://doi.org/10.1111/j.1528-1167.2008.01686.x.

[21] Depannemaecker D, Santos LEC, Rodrigues AM, Scorza CA, Scorza FA, de Almeida ACG. Realistic spiking neural network: Non-synaptic mechanisms improve convergence in cell assembly. Neural Networks. 2020 Feb;122:420–433. Available from: https://doi.org/10.1016/j.neunet.2019.09.038.

[22] Jiruska P, de Curtis M, Jefferys JGR, Schevon CA, Schiff SJ, Schindler K. Synchronization and desynchronization in epilepsy: controversies and hypotheses. The Journal of Physiology. 2013 Jan;591(4):787–797. Available from: https://doi.org/10.1113/jphysiol.2012.239590.

[23] Depannemaecker D, Ivanov A, Lillo D, Spek L, Bernard C, Jirsa V. A unified physiological framework of transitions between seizures, sustained ictal activity and depolarization block at the single neuron level. Journal of Computational Neuroscience. 2022 Jan;50(1):33–49. Available from: https://doi.org/10.1007/s10827-022-00811-1.

[24] Tavanaei A, Ghodrati M, Kheradpisheh SR, Masquelier T, Maida A. Deep learning in spiking neural networks. Neural Networks. 2019 Mar;111:47–63. Available from: https://doi.org/10.1016/j.neunet.2018.12.002.

[25] Nicola W, Clopath C. Supervised learning in spiking neural networks with FORCE training. Nature Communications. 2017 Dec;8(1). Available from: https://doi.org/10.1038/s41467-017-01827-3.

[26] Kolb B, Whishaw IQ. BRAIN PLASTICITY AND BEHAVIOR. Annual Review of Psychology. 1998;49(1):43–64. PMID: 9496621. Available from: https://doi.org/10.1146/annurev.psych.49.1.43.

[27] Sapolsky R. Behave : the biology of humans at our best and worst. New York, New York: Penguin Press; 2017.

[28] Kandel ER, Schwartz JH, Jessell TM, editors. Principles of Neural Science. 3rd ed. Elsevier; 1991.

[29] Marr D. Vision. The MIT Press. MIT Press; 1982.

[30] Vaughan J, Sudjianto A, Brahimi E, Chen J, Nair VN. Explainable Neural Networks based on Additive Index Models; 2018.

[31] Yang Z, Zhang A, Sudjianto A. Enhancing Explainability of Neural Networks Through Architecture Constraints. IEEE Transactions on Neural Networks and Learning Systems. 2021 Jun;32(6):2610–2621. Available from: https://doi.org/10.1109/tnnls.2020.3007259.

[32] Wan A, Dunlap L, Ho D, Yin J, Lee S, Jin H, et al.. NBDT: Neural-Backed Decision Trees; 2021.

[33] Ruphy S. Scientific pluralism reconsidered: A new approach to the (dis)unity of science; 2016.

[34] Varenne F. From models to simulations. Abingdon, Oxon New York, NY: Routledge; 2019.

[35] Shrestha A, Mahmood A. Review of Deep Learning Algorithms and Architectures. IEEE Access. 2019;7:53040–53065.

[36] Rawat W, Wang Z. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput. 2017 Sep;29(9):2352–2449.

[37] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al.. Generative Adversarial Networks; 2014.

[38] Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G. Backpropagation and the brain. Nature Reviews Neuroscience. 2020;21(6):335–346.

[39] Whittington JC, Bogacz R. Theories of error back-propagation in the brain. Trends in cognitive sciences. 2019;23(3):235–250.

[40] Gauld C, Brun C, Boraud T, Carlu M, Depannemaecker D. Computational models in neurosciences between mechanistic and phenomenological characterizations; 2022. Available from: https://doi.org/10.20944/preprints202201.0206.v1.

[41] Miłkowski M. Computation and Multiple Realizability. In: Fundamental Issues of Artificial Intelligence. Springer International Publishing; 2016. p. 29–41. Available from: https://doi.org/10.1007/978-3-319-26485-1$_3$.

[42] Bickle J. Multiple Realizability. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy. Summer 2020 ed. Metaphysics Research Lab, Stanford University; 2020. .

[43] Levin M, Pezzulo G, Finkelstein JM. Endogenous bioelectric signaling networks: exploiting voltage gradients for control of growth and form. Annual review of biomedical engineering. 2017;19:353–387.

[44] Pezzulo G, Levin M. Re-membering the body: applications of computational neuroscience to the top-down control of regeneration of limbs and other complex organs. Integrative Biology. 2015;7(12):1487–1517.

[45] Pezzulo G, Levin M. Top-down models in biology: explanation and control of complex living systems above the molecular level. Journal of The Royal Society Interface. 2016;13(124):20160555.

[46] Floridi L, Chiriatti M. GPT-3: Its nature, scope, limits, and consequences. Minds and Machines. 2020;30(4):681–694.

[47] Anderson JR, Matessa M, Lebiere C. ACT-R: A theory of higher level cognition and its relation to visual attention. Human–Computer Interaction. 1997;12(4):439–462.

[48] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern recognition. 2018;77:354–377.

[49] Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of Physiology. 1952 Aug;117(4):500–544. Available from: https://doi.org/10.1113/jphysiol.1952.sp004764.

[50] Izhikevich EM. Simple model of spiking neurons. IEEE Transactions on neural networks. 2003;14(6):1569–1572.

[51] Goldman JS, Kusch L, Yalcinkaya BH, Depannemaecker D, Nghiem TAE, Jirsa V, et al. Brain-scale emergence of slow-wave synchrony and highly responsive asynchronous states based on biologically realistic population models simulated in The Virtual Brain. bioRxiv. 2020;Available from: https://www.biorxiv.org/content/early/2020/12/29/2020.12.28.424574.

[52] Friston K. The free-energy principle: a unified brain theory? Nature reviews neuroscience. 2010;11(2):127–138.

[53] Ullman S. Using neuroscience to develop artificial intelligence. Science. 2019;363(6428):692–693.

[54] Cuthbert BN. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. World Psychiatry. 2014;13(1):28–35.

 doi:10.20944/preprints202207.0229.v1

[55] Mayr E. Cause and Effect in Biology. Science. 1961 Nov;134(3489):1501–1506. Available from: https://doi.org/10.1126/science.134.3489.1501.