A Prolegomenon on the Philosophical Foundations of Deep Learning as Theory of (Artificial) Intelligence

Skansi, Sandro; Kardum, Marko

Source / Izvornik: **Disputatio philosophica : international journal on philosophy and religion, 2021, 23, 89 - 99**

Journal article, Published version
Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

https://doi.org/10.32701/dp.23.1.6

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:111:956092

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2025-01-23**

SVEUČILIŠTE U ZAGREBU
HRVATSKI STUDIJI
UNIVERSITY OF ZAGREB
CENTER FOR CROATIAN STUDIES

Repository / Repozitorij:

Repository of University of Zagreb, Centre for Croatian Studies

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

A PROLEGOMENON ON THE PHILOSOPHICAL FOUNDATIONS OF DEEP LEARNING AS THEORY OF (ARTIFICIAL) INTELLIGENCE

Sandro Skansi, Marko Kardum

UDC 004.85:16
https://doi.org/10.32701/dp.23.1.6
Original scientific paper
Received: 15.11.2021
Accepted: 30.12.2021

Abstract

This paper examines the philosophical foundations of deep learning. By pointing to the beginnings of deep learning and artificial neuron as a logical model of a human neuron, it is possible to claim that artificial intelligence was developed even before its official creation and that it was strongly connected to propositional logic. Bearing in mind some major setbacks in the development of neural networks, we show that deep learning can be treated as the theory of artificial intelligence and that it falls under artificial intelligence paradigm by claiming that everything can be done with learning alone and that all intelligent behavior is learnable. Thus, deep learning is a philosophical or an epistemological approach in which a form of radical empiricism must be advocated. Therefore, there is nothing in the mind that was not in the senses, and there cannot be anything in the mind that is not learnable.

KEYWORDS: artificial intelligence, cybernetics, deep learning, empiricism, neural networks

Introduction

In this paper we explore the philosophical foundations of deep learning. Even though deep learning seems to be only an engineering approach in

✦ The article is a revised and extended version of the paper presented at the international symposium held in Zagreb on 20th November 2020, titled 'The Impact of Technology on Human Being and Its Self–understanding'.

* Sandro Skansi, PhD, Assistant Professor, Faculty of Croatian Studies, University of Zagreb, Borongajska cesta 83d, 10 000 Zagreb, Croatia. E–mail: sskansi@hrstud.hr
ORCID iD: https:// orcid.org/0000–0002–3851–1186

** Marko Kardum, PhD, Assistant Professor, Faculty of Croatian Studies, University of Zagreb, Borongajska cesta 83d, 10 000 Zagreb, Croatia. E–mail: mkardum@hrstud.hr
ORCID iD: https:// orcid.org/0000–0002–0797–6677

artificial intelligence, its rich history is permeated with philosophy, since deep learning aims to be a wholly encompassing theory of artificial intelligence, and by doing so, it becomes a theory of (artificial) minds. The deep insight of cybernetics as its original philosophy, which generated the man–machine equivalence, was never really popular, for it provided a rejection of the much beloved Cartesian dualism. Connectionism had its moments, but it was quite a shallow philosophical reflection of artificial neural networks, since it was mainly a theory which wanted to "import" neural networks in philosophy to fix computationalist theories in philosophy of mind. In this paper we will show that today's connections are even simpler, and that by revisiting its philosophical origins, it can be seen that the foundations of deep learning are purely empiricist in nature, as opposed to older paradigms of artificial intelligence.

1. *The Beginnings of Deep Learning*

Deep learning draws its beginnings from the work of Warren McCulloch and Walter Pitts (1943), where the authors described an artificial neuron for the first time in history. McCulloch and Pitts wanted to develop a "learning logic", or a logical model of how a human neuron learns. Initially, they considered the idea of neurons whose connections would not be directed, but it turned out that technical orientation greatly simplifies the construction of neurons. McCulloch and Pitts wanted to show how their neuron could learn any logical function, which they thought was the first step in establishing thought (de facto classical propositional logic) in this new artificial neuron (which is a de facto logical operation).

Here it is well worth looking deeper into their neuron. The McCulloch–Pitts neuron is a de facto "normalization" of the voting function. A logical voting function is a Boolean expression that receives as input N Boolean values and if most of them are 1, returns 1, otherwise returns 0. Interestingly, for a fixed number of inputs, this function is definable as a disjunction of conjunctions of all pairs of variables. However, as we noted, the McCulloch–Pitts neuron has a moment of "normalization". This normalization exists at the input level, so that each Boolean value receives a coefficient of importance, the so–called weight. Good weights guarantee that the output will be as it should be. For example, if it is a matter of learning a conjunction, which has a value of 1 when both inputs have a value of 1, the weights are 0.5, if the result is taken without decimals. In an analogous way, weights for other logical functions can be found on the same neuron, and this is precisely the importance of their discovery: the neuron remains (structurally) the same, but different weights create different functions. However,

the authors go a step further, de facto creating artificial intelligence 11 years before it was officially created in 1954. Is it possible to give the neuron the expected results and ask it to automatically find weights? Is it possible to expect neurons to learn any function on their own? In order to successfully illustrate the complex opportunities for the emergence of artificial intelligence, we must first consider the emergence of a much more important science: cybernetics.

2. *Cybernetics. The Original Philosophy of Deep Learning*

The philosopher Norbert Wiener, who is today considered the father of cybernetics, defined cybernetics as a new science of control (1948), emphasizing in the introduction the importance of logicians in this new discipline. In this period, there were no mathematicians or electrical engineers in cybernetics, but it was primarily a philosophical discipline that should have been completely formal, like logic, and yet experimental in the way that Pitts and McCulloch, Wiener's longtime collaborators, devised five years earlier. But perhaps the best definition was given by Ross Ashby (1958, pp. 1–6), where it is noted that cybernetics is similar to physics in its generality. Physics never takes energy in the system for granted, but it ignores the information components. Cybernetics should be just the opposite: a general and formalized study of the world with an emphasis on information change, not energy.

In this regard, it is very clear that cybernetics, in its original form, is in fact a metaphysical theory, more precisely, it is a formal ontology of processes. We refer the interested reader who is interested in more about cybernetics as metaphysics to Skansi and Šekrst (2021). Over the years, cybernetics would lose most of its luster and be degraded to ordinary computer engineering. In parallel with the rise of cybernetics, artificial intelligence emerged. The term was coined by John McCarthy in 1956 at the Dartmouth Conference (Russell and Norvig 2010), which was attended by mathematicians and electronics engineers. Interestingly, McCarthy, according to later statements, deliberately decided not to invite Norbert Wiener so as not to annex this "new" science to "his" cybernetics. The sad irony of fate is that today everything cybernetically attached to "artificial intelligence" has de facto expelled philosophers from its development and its actual history.

3. *The XOR: Is This Really a Mind?*

One of the greatest blows to artificial neural networks was dealt by Marvin Minsky and Seymour Papert (1969), because they showed that an elementary logical function cannot even be learned in principle by an artificial neuron. It was about equivalence. The problem was that artificial neurons are linear separators, and no line in a two–dimensional system can separate (0,0) and (1,1) on the one hand and (0,1) and (1,0) on the other. But their result is even stronger than that. Linear equivalence can be achieved by adding a third variable, which is the conjunction (or brain) of x and y. Then the equivalence becomes linearly separable in the newly formed three–dimensional Cartesian system. This consequently means that the artificial neuron is not capable of learning something that would simulate this third feature. In other words, an artificial neuron is limited by inputs and in principle lacks the ability to detect new "views". All in all, an artificial neuron is just an ordinary, "stupid" machine, which is no different from Turing's machine, and that is a catastrophic blow to the very idea of artificial intelligence, which should be "smarter" than an ordinary computer, or Turing's machine. In a sense, the above mentioned Papert and Minsky's argument buried both artificial neural networks and cybernetics, as the argument viewed them in the same way in which traditional philosophy sees logic: they were to be the main tool which shapes cybernetics into a formal discipline and, consequently, makes it scientific. It will take 17 years of almost non–existent and unfinished research for artificial neural networks to return.

The problem that Minsky and Papert saw is unsolvable, but few cyberneticists of this period tried the Nazi detour. The idea has existed since Pitts: instead of one neuron, what could we do if we had a whole network of neurons, where the output of one is the input to the others? A few things should be noted here. First, we mentioned that neural networks are de facto specified by weights, but there is one even more important detail here. If the weights are initialized randomly, and during the adaptation of the neurons to the task (training) they change from random to "perfect," why should we not delegate the tasks? Surely the weights that recognize the cat in the picture are better at recognizing the dog than random weights. This is a good idea, but unfortunately it will not be reached until the early 2010s because no one remembered it. But besides the weights, there are inputs — what if we used something better here than pure inputs? What if we used the output of another network as inputs?

4. *Thinking with Backpropagation*

Technically, there is one big problem, how to deliver the necessary weight corrections to a neuron that is further from the end result. This was solved by the idea of backpropagation, which was discovered by Hinton, Rumelhart, and Williams in 1986 (1986). The solution was in fact very simple: the weights are refined by adding to them a partial derivative of the function that measures the error in the direction of that weight. Surprisingly, partial neural network derivatives are extremely simple for the main error functions. However, as one problem is solved, another has arisen: the application of partial derivations to complex networks depends largely on the so–called chain rule, which basically says that going from the finite neuron to the initial error (between 0 and 1) multiplies. This means that an error of 0.3 is distributed not only in the last layer of neurons, but the error of that layer is also distributed in the penultimate layer, e.g., 0.5, then 0.3 times 0.5 and 0.15, respectively. The layer in front is not distributed only 0.3 but 0.15 times 0.3 or 0.05. This is called the vanishing gradient, and the problem is that the weights on the initial layers change very little, to the point that they are already almost at the level of rounding error and sum in the fifth layer from the end. This is of course a big limitation because more complex problems cannot be solved by simply adding layers, although the equivalence problem has been successfully solved with two layers.

There have been very daring upgrade attempts, and it can be said that in the 1980s there was an unusually diverse creativity in making neural networks, as well as optimism that it was "technology of the future", so they are mentioned in *Terminator 2* and in the episode "Measure of a man" of *Star Trek: The Next Generation* as the central system of very compelling cyborgs. This period is also specific in the emergence of connectionism in the philosophy of mind: the mind is a neural network, and everything specific to the individual comes from experience and is stored in weights between real neurons. Any human ability should therefore be able to be simulated by sufficiently advanced artificial neural networks. It is interesting to note that the *reflection* on artificial neural networks in the philosophy of mind has never occurred before. The reason is that earlier, when it was cybernetics, no distinction was made between philosophy and technology: cybernetics was and remains a de facto philosophical methodology, which studied analogous processes in man, society and machine with the same methods. In this sense, Wiener's and Ashby's cybernetics had as its primary task the study of processes, and only secondarily their engineering replication, and consequently there was no gap between philosophy and engineering, or theory and practice.

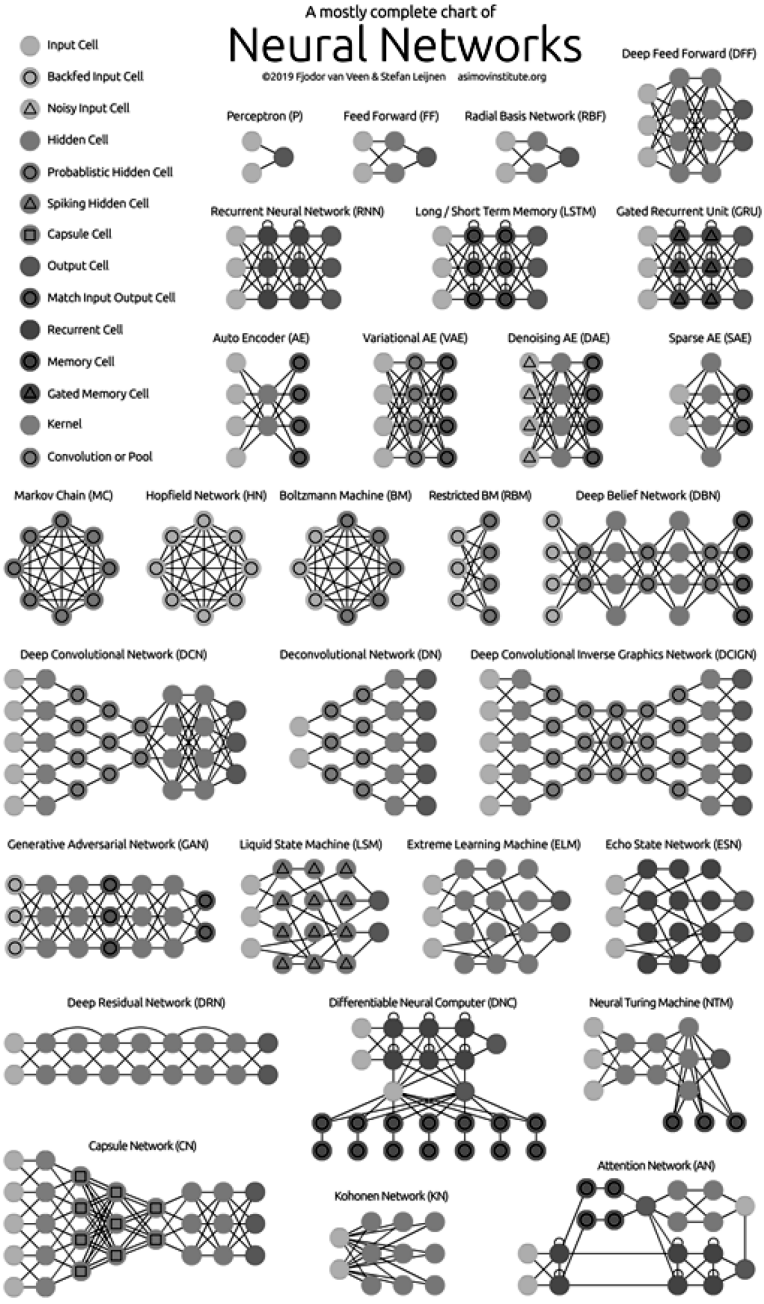
5. *Cybernetics for Dummies and Engineers*

The years 1997 and 1998 saw two different events that could be characterized as the birth of deep learning, and both could be seen as ways to fight the vanishing gradient. The first was the joint work of Hochreiter and Schmidhuber (1997), the long short–term memory. The idea was a lifelong obsession for Schmidhuber — to improve recurrent neural networks with an explicit memory, which was a successful way to fight the vanishing gradient. Today, more than twenty years later, LSTMs are still widely used, especially for tasks that require an ongoing output. A wholly different idea came to Yann LeCun in 1998 (LeCun, Bottou, Bengio and Haffner 1998), for solving the vanishing gradient in multiple layers. The idea was to use a simple nine input logistic regression (neuron) to scan the image. This constitutes one feature map, smaller than the original image. Note that just like a new logistic regression can be used for each color channel for the image, we can use multiple regression on the same channel, thereby creating a number of so–called feature maps. Structures made in this fashion would become known as convolutional neural networks.

The events from 1997 and 1998 marked the beginning of a new era for artificial neural networks, which would become known under the cryptic name of "deep learning". This ends the "yesterday" and begins the "today". As a last note, we would like to point to a great injustice. The 2018 Turing Award, considered by many to be the equivalent of the Nobel Prize in Computing, was awarded to Hinton, LeCun and Bengio, who undoubtedly deserve it, but not to Schmidhuber, who was a key person in developing recurrent neural networks, and should have been included. We believe that the only reason is that Schmidhuber did not have such marketing as the other ones, behind whom there are Google and Facebook with their propaganda machinery.

It can be argued that the "present" of deep learning starts in 2006 with Deep Belief Networks (Hinton, Osindero and Teh 2006), the first architecture to be called "deep". The architecture itself is irrelevant, and we point the reader who is interested in the technical details to the original paper. What is relevant is that before 2006 most research was done on developing alternative internal structures, as exemplified by LSTMs and convolutional networks. However, from 2006 to the present day, and especially after the invention of modern autoencoders in 2009 (Bengio 2009), the vast majority of non–derivative research focused on building external structures, i.e., combining simple neural networks, LSTMs, convolutional networks and autoencoders in larger models capable of more complex behavior. This can be seen in Figure 1, taken from Van Veen and Leijnen (2020), which also provides a great overview of different architectures in use today.

Figure 1. A mostly complete chart of Neural Networks (Taken from Van Veen and Leijnen 2020).
Slika 1. Gotovo potpuni prikaz neuronskih mreža (Preuzeto iz Van Veen i Leijnen 2020).



6. *Philosophical Underpinnings of Deep Learning as It Sits*

Thus far, we have dealt quite a bit with how deep learning developed, but very little has been said about how it works, what it can be applied to, and what constitutes its philosophical importance. We have explained in brief how neural networks, i.e., deep learning works, but we have said nothing about what makes it special. If one takes even a cursory look at all of the different machine learning architectures, like the ones in Burkov (2019), one might get the feeling that deep learning is only a subset of machine learning, and in a sense, one would be right. Deep learning *is* machine learning, and all limitations of machine learning apply to deep learning as well. But its ambitions are much greater than just a formalization of learning.

However, what would it mean to have greater ambitions? It means that deep learning thinks of itself as a new *artificial intelligence* and *not a machine learning* paradigm. If the reader objects to this, claiming that there are things which are not just learning, we should remind them that, for instance, *every* language task may be viewed as a translation task, and even more generally, every language task can be viewed as a question–answering task. In a sense then, all of the natural language processing is then just question answering. Deep learning is trying to make the same claim about itself and artificial intelligence.

As the Turing test (Turing 1950) was once considered the benchmark of machine intelligence, today we could look at Searle's Chinese room (Searle 1980) as a test to determine intelligence. In a sense, the two tests use the same idea: there is a process behind the curtains, and a "spectator" interacts with the process by using some controls on his side. In the case of the Turing test, there is a machine behind the curtain, and if the machine is deemed to be human by the spectator, the machine is intelligent, and the test is passed by imitation. In the case of Searle's test, there is a human behind the curtain, and if the human is deemed to know Chinese by the spectator, the test is failed by imitation. The Turing test equates intelligence to imitation, while Searle's test supposedly shows that imitation alone is not intelligence. The question might be asked whether imitation forms (a) the kernel of intelligence or (b) the basis of intelligent behavior. There is a fine difference, best shown by Weston et al. (2015) that the kernel of intelligence is not a module, but a set of tasks, and any module that accomplishes them might be viewed as rudimentarily intelligent, regardless of its structure. As this set of tasks cannot be simulated by imitation, we could say that the answer to (a) is negative, but the answer to (b) is positive. In such a world, it is easy to see how deep learning hopes to circumvent Searle's test by redelegating the complex tasks to training. The idea of deep learning is that the "man" in the room is to "learn" Chinese and only then process the inputs.

In a sense, deep learning claims that everything can be done with learning alone. All intelligent behavior is, in principle, learnable. And this makes deep learning not just a technical, but also a philosophical approach, more precisely an epistemological approach. To see this, consider the question "What is learning?" It is, in essence, induction — and the claim that induction can be used to form a complete (artificial) mind is a radical form of empiricism. Not only there cannot be anything in the mind that is not in the senses, but there cannot be anything in the mind which is not learnable. What makes learnability different from sensing is that we can sense stuff without being able to explicate, let alone explicitly internalize: I can sense and acknowledge a sensation as "nice" without a commitment of making an explicit mental drawer where nice things are to be placed. However, if I learn what is nice, I will always be able to place similar sensations in the same compartment.

Conclusion

The present state of deep learning is plagued with applications, and one might be tempted to dismiss the idea of deep learning as a mere technological novelty. But as we have argued, it is in fact not philosophically neutral. By equating machine learning with induction, we were able to show that deep learning has a strong philosophical position, empiricism. Even though this might seem strange at first, one ought to wonder what induction is. If one agrees that the most basic form of induction is the recognition of a "species" by means of shared attributes (whatever they might actually be), then machine learning is not just a formalization of a subtype of inductive inference but *is* a formalization of induction as such. Deep learning presupposes that everything needed to recreate the mind or some of its faculties is not just given as sensory data but is in fact learnable by a sufficiently advanced deep learning algorithm. In a sense, given enough technological slack, everything that is needed is sensory data, and any intelligent behavior can be produced. And not just that: the behavior will not just be (re)produced, but it will be learned, i.e., explicitly internalized, and then reproduced as required.

References

Ashby, William Ross. 1958. "Requisite Variety and Its Implications for the Control of Complex Systems." *Cybernetica* 1 (2): 83–99. https://doi.org/10.1007/978–1–4899–0718–9_28.

Bengio, Yoshua. 2009. "Learning Deep Architectures for AI." *Foundations and Trends in Machine Learning* 2, (8): 1795–1797. http://dx.doi.org/10.1561/2200000006.

Burkov, Andriy. 2019. *The Hundred–Page Machine Learning Book Paperback*. Published by Andriy Burkov. ISBN–13: ‎978–1999579500.

Hinton, Geoffrey E., Simon Osindero, and Yee–Whye Teh. 2006. "A Fast Learning Algorithm for Deep Belief Nets." *Neural Computation* 18 (7): 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short–term Memory." *Neural Computation* 9 (8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient Based Learning Applied to Document Recognition." *Proceedings of IEEE* 86 (11): 2278–2324.

Minsky, Marvin, and Seymour A. Papert. 1969. *Perceptrons. An Introduction to Computational Geometry*. Cambridge (MA): The MIT Press. https://doi.org/10.7551/mitpress/11301.001.0001.

Perkov, Tin. 2020. "The McCulloch–Pitts Paper from the Perspective of Mathematical Logic." In Guide to Deep Learning Basics. Logical, Historical and Philosophical Perspectives edited by Sandro Skansi, 7–12. Cham: Springer. https://doi:10.1007/978–3–030–37591–1_2.

Pitts, Walter, and Warren S. McCulloch. 1943. "A Logical Calculus of the Ideas Immanent in Nervous Activity." *The Bulletin of Mathematical Biophysics* 5: 115–133. https://doi.org/10.1007/BF02478259.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. "Learning Representations by Back–propagating Errors." *Nature* 323: 533–536. https://doi.org/10.1038/323533a0.

Russell, Stuart, and Peter Norvig. 2010. *Artificial Intelligence. A Modern Approach* (3rd edition). Hoboken (NJ): Prentice Hall.

Searle, John. 1980. "Minds, Brains and Programs." *The Behavioral and Brain Sciences* 3, pp. 417–424.

Skansi, Sandro, and Kristina Šekrst. 2021. "The Role of Process Ontology in Cybernetics." *Synthesis Philosophica* (accepted for publication)

Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind*, LIX (236): 433–460, doi:10.1093/mind/LIX.236.433

Van Veen, Fjodor, and Stefan Leijnen. 2020. "The Neural Network Zoo." *Proceedings* 47 (1), 9: 2–6. https://doi.org/10.3390/proceedings2020047009.

Weston, Jason, Antoine Bordes, Sumit Chopra, Alexander M. Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. 2015. "Towards AI–Complete Question Answering: A Set of Prerequisite Toy Tasks." https://arxiv.org/abs/1502.05698v10

Wiener, Norbert. 1948. *Cybernetics; or Control Communication in the Animal and the Machine.* New York: The Technology Press; John Wiley and Sons.

Sažetak

PROLEGOMENA FILOZOFIJSKOG UTEMELJENJA DUBOKOG UČENJA KAO TEORIJE (UMJETNE) INTELIGENCIJE

SANDRO SKANSI, MARKO KARDUM

U radu se ispituju filozofski temelji dubokog učenja. Ukazivanjem na početke dubokog učenja i umjetnog neurona kao formalnog modela ljudskog neurona moguće je tvrditi da je umjetna inteligencija razvijena i prije njezinog službenog imenovanja te da je bila snažno povezana s propozicionalnom logikom. Imajući na umu neke velike zastoje u razvoju neuronskih mreža, pokazujemo da se duboko učenje može tretirati kao teorija umjetne inteligencije te da potpada pod paradigmu umjetne inteligencije jer je za nju dovoljno samo učenje i jer se inteligentno ponašanje uči. Dakle, duboko učenje je filozofski ili epistemološki pristup u kojem se mora zagovarati radikalni empirizam. Prema tome, ne samo da ne postoji ništa u umu što nije bilo u osjetilima, već u umu ne postoji ništa što se ne može naučiti.

Ključne riječi: empirizam, duboko učenje, kibernetika, neuralne mreže, umjetna inteligencija

* Doc. dr. sc. Sandro Skansi, Fakultet hrvatskih studija Sveučilišta u Zagrebu, Borongajska cesta 83d, 10 000 Zagreb, Hrvatska. E–adresa: sskansi@hrstud.hr
ORCID iD: https:// orcid.org/0000–0002–3851–1186

** Doc. dr. sc. Marko Kardum, Fakultet hrvatskih studija Sveučilišta u Zagrebu, Borongajska cesta 83d, 10 000 Zagreb, Hrvatska. E–adresa: mkardum@hrstud.hr
ORCID iD: https:// orcid.org/0000–0002–0797–6677