This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2020.DOI

Deep Learning Trends Driven by Temes: A Philosophical Perspective

YUHONG ZHANG[1], (Member, IEEE), UMER NAUMAN [2]

[1]School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, China, 460001 (e-mail: zhangyuhong001@gmail.com)

[2]Information Science and Engineering College, Henan University of Technology, Zhengzhou, China, 460001 (e-mail: stormy.umer@gmail.com)

Corresponding author: Yuhong Zhang (e-mail: zhangyuhong001@gmail.com).

This work was supported in part by the National Natural and Science Foundation of China under Gran 61705061, 61975053 and U1904120, and Research and Equipment Development of New Grain Situation Measurement and Control Technology under Gran 2017YFD0401004

ABSTRACT In up-to-date years deep learning has accrued diverse state-of-the-art results in many fields. However, most of the research efforts chiefly focus on pure technologies or engineering optimizations of deep learning. Inadequate consideration has been paid to it from the panorama of the philosophy of technology. Given the universal Darwinism panorama, technology is regarded as the seventh life form, which, like the other six well-known life forms, has the driving forces of survival and evolution. Similar to genes, *temes* are the inheritors of technology. If *temes* do not yield to the law of "survival of the fittest", they will be discarded by the evolution. In this paper, we refined the essence hidden behind deep learning's uproar emergence from a technical panorama. We then rendered its importance and philosophical perspective of deep learning. Based on the point of view of "*teme*" (the basis of technology), we moreover analyzed the inherent *teme*-related defects of deep learning. Finally, we addressed the possible teme-driven directions for deep learning.

INDEX TERMS Deep learning, philosophy of technology, universal Darwinism, teme

I. INTRODUCTION

As the most crucial advancement in the realm of artificial intelligence (AI), deep learning has dramatically widened the state-of-the-art in many fields like recognizing speech [1], translating texts [2], labeling images [3], playing strategy games [4], predicting protein folds [5], intrusion detection [6], and driving automobiles [7] in the recent years.

Artificial neural networks, especially with multiple hidden layers (hence the term *deep*), are remarkably good at learning the mapping between input (i.e., raw data) and output (a set of categories). For example, in handwriting recognition images, a neural network learns a mapping between input images and numerically classified (e.g., 0, 1, 2, ...). This kind of computing paradigm is called "*end-to-end*" [8].

Deep learning is known for its power to self-generate intermediate representations. Technically, it is a multistage way to learn data representations, discovering the underlying structure in big data-sets using various back-propagation variants. For instance, as shown in Figure 1, the network transforms the digital image into a representative image that is way more different from the original image, and the figure also portrays the growing number of information about the final result. These representations can be obtained by composing nonlinear but straightforward modules, which transform the representation at a low level into a higher, usually more abstract level.

For classification tasks, the higher level of representation amplifies the input parts, eliminating discrimination while suppressing the irrelevant variations. We can regard a deep neural network as multiple information-distillation levels, where information goes through successive filters and comes out increasingly purified [9].

Deep learning implementations can provide greater accuracy, better predictive performance, greater flexibility, and reconfigurability [10]. To a large extent, its success is chiefly ascribable to its inclination, which learns novel features or patterns and understands data representation in both a supervised and/or unsupervised hierarchical manner [11], [12]. Although having played a pivotal role in earlier AI and neural networks investigation [13]–[15], philosophy heretofore has been principally soundless on deep learning [16].

Much AI research is still focused on producing useful and profitable information processing, whether or not the results provide philosophical understanding [17]. Specific to deep

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access

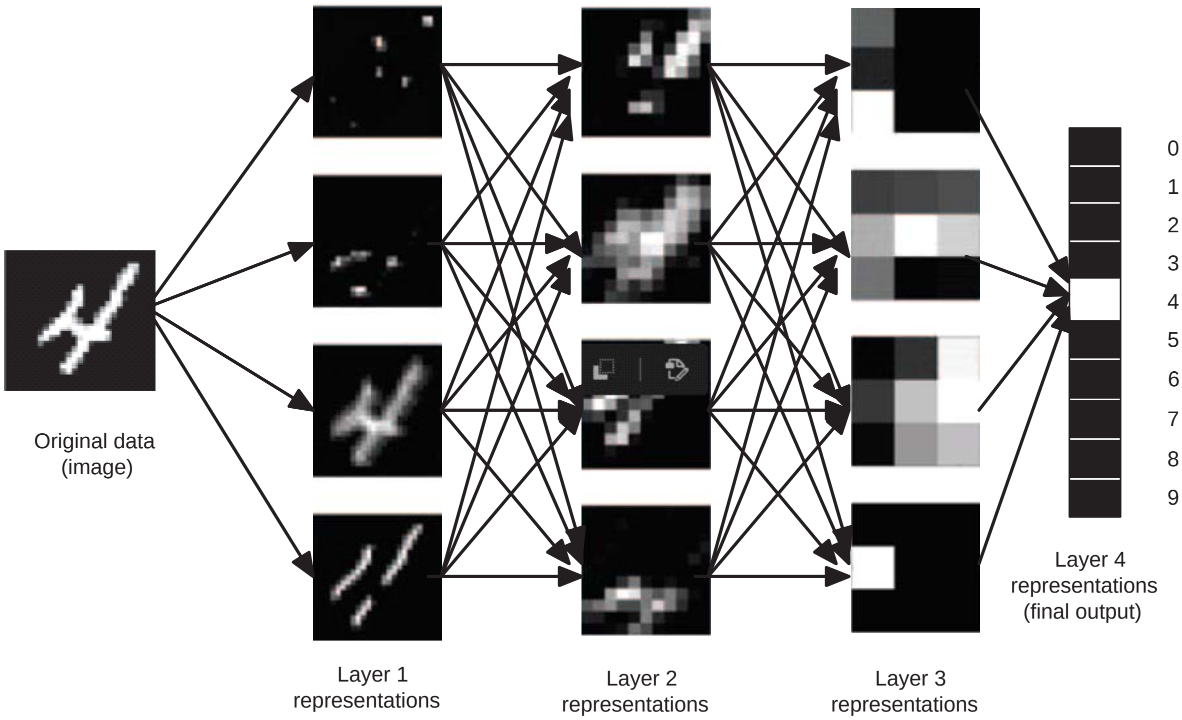


FIGURE 1. Hierarchical representations learned by deep learning from massive handwritten digital images

learning, most research endeavors are leveled toward pure technology or engineering. As far as we know, few comparisons have contemplated the bearings in deep learning from the perspective of philosophy of science and technology. If we could re-examine deep learning know-how from this viewpoint, we would find captivating insights into it.

Based on the observation of science and technology philosophy, we are engaged in analyzing deep learning trends in this paper. There is no doubt that deep learning plays an essential role in AI. Nevertheless, deep learning itself has many limitations and may well be approaching walls [18]–[21]. This paper makes a tentative discussion on the following problems from a novel philosophical perspective – *temes* : (1) what are the potential "*teme-related*" defects of deep learning? Furthermore, to survive, (2) what are its trends driven by *temes*?

The rest of this paper is organized as follows. In Section II, a formal definition of deep learning is provided, and its essence is refined. Then, the relations and differences of *gene*, *meme* and *teme* are discussed in more detail in Section III. Followed by Section IV, an intuitive thought experiment is carried out to prove the significance of this paper. In subsequent Section V , three *teme-related* defects in deep learning, including brute force (overeating data), low intelligence-energy density and superficial understanding, will be thoroughly explained. Then, in Section VI, a datasim perspective will be utilized to take an in-depth inspection of deep learning technologies. At last, some workable solutions, namely future trends, will be explained in Section VII.

II. FORMAL DEFINITION AND ESSENCE OF DEEP LEARNING

The essence of an artificial neural network is to approximate a function through the fitting of massive data. This function is expected to have a smart mapping between input and output. The mathematical basis of deep learning is the universal approximation theorem [22]. The theorem says that any continuous function can be approximately obtained by combining a series of simple affine function sets (such as *sigmoid*, *tanh* or ReLU [23]).

It is worth mentioning that recently scholars have also proved the universality of deep convolutional neural network (CNN) strictly in mathematics [24]. It means that CNN also can approximate any continuous function when the neural network is *deep* enough. At a very high level of abstraction, deep learning is actually a series of nested functions. In a feed-forward neural network, the function $F : \mathbb{R}^{n_0} \to \mathbb{R}^{out}$ can be defined as the Eq. 1:

$$F(x;\theta) = f_1(x) \circ g_1 \circ \cdots \circ f_L \circ g_L \circ f_{out} \qquad (1)$$

where f_l(l is the layer number of network) is a linear function, which sums up all the weighted inputs from the previous layer, and g_l is a nonlinear activation function, which is used for nonlinear transformation to improve the representation ability of the neural network. The parameter θ is composed of $input$ weights matrix $W_l \in \mathbb{R}^{k \cdot n_l \times n_{l-1}}$ and *bias* vectors $b_l \in \mathbb{R}^{k \times n_l}$ for each layer $l \in [L]$.

The output of the l-th layer is a vector $X_l = [x_{l,1} \cdots x_{l,n_l}]^T$ of neuron activation functions $x_{l,i}$ of the units

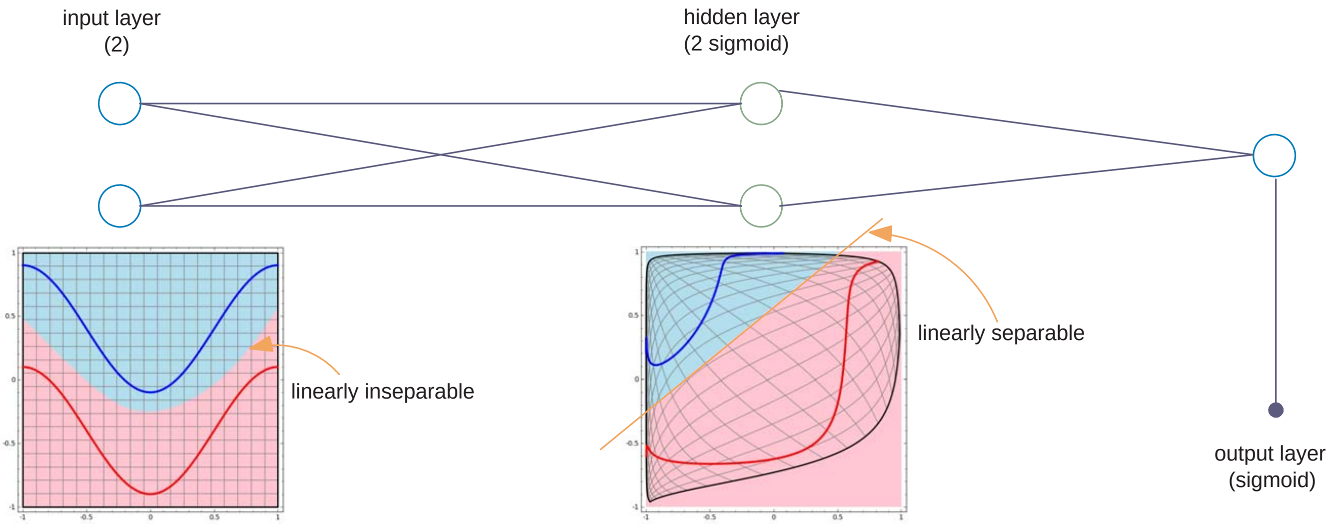


FIGURE 2. Hidden layer learns a representation so that the classes of data are linearly separable

$i \in [n_l]]$ in that layer. The l-th layer can be computed from the neuron activation functions of the antecedent layer by $X_l = h_l(f_l(X_{l-1}))$. The linear layer of l is given by

$$f_l(x_{l-1}) = W_l X_{l-1} + b_l \tag{2}$$

where $f_l = [f_{l,1}, f_{l,2}, \ldots, f_{l,n_l}]$ is an array composed of n_l previous vectors $f_{l,i} \in \mathbb{R}^k$.

Given the activation of the i-th unit in the l-th layer, the output is given by

$$x_{l,i} = h_{l,i}(f_{l,i}(X_{l-1})) \tag{3}$$

For simplicity, $g_l \circ f_l$ can be abbreviated as h_l. The most general form of machine learning, deep or not, is supervised learning. An objective function of supervised learning is often used to measure the error (or distance) between the actual output and the expected one. If there is an error, then the neural network's linked weights will be adjusted properly by the learning algorithm like back-propagation [25], consequently reducing the error to the tolerable range.

There are hundreds of millions of adjustable parameters and millions of labeled samples in a typical deep learning system, which trains the model to learn optimal or sub-optimal parameters and makes the loss function reach a minimum.

The advantage of the deep neural network is that, from the first layer, it can distort the input space hierarchically. This feature is handy for the classification. Each hidden layer can be regarded as a folding operator. After multiple successive layers are cleverly folded, the original linear indivisible space (the classes of data) become linearly separable [26].

As illustrated in Figure 2, the role of hidden layer neurons of a neural network can be interpreted as a function that transforms input patterns (examples of which are on the red and blue lines) from a nonlinear separable space (shown in the middle panel) to a linear separable space [27].

Figure 2 is a demonstrative example with only two input units, two hidden units, and one output unit. However, the current neural network for natural language processing (NLP) or object recognition may contain tens or hundreds of thousands of neurons. Through innumerable hyper-plane distortions, neural networks are still useful for categorizing tasks, but they are no longer interpretable to humans.

III. RELATED WORKS : FROM GENE, MEME TO TEME

There have been many academic articles discussing trends in deep learning [10], [28]–[30]. Young et al. [28] discussed the development trend of deep learning by taking natural language processing as an example. More technically, Nwankpa et al. [29] highlighted the recent trends in the use of the activation functions for deep learning applications. In a niche area, such as personality detection, Mehta et al. [30] summarized significant deep learning models that have been employed for personality detection and offered the future development direction of deep learning in this application field. At the higher macro-level, Hatcher et al. [10] provided a comprehensive reference for trends in deep learning in terms of implementations, platforms, and algorithms.

One of the above studies' commonality is that they focus on a fundamental technical or engineering perspective. This perspective is valid, but it may not be in-depth and comprehensive enough. If we want to gain a more in-depth insight, we have to pull away from the technical details to another novel viewpoint: the philosophical perspective. It helps us to observe the trends of deep learning in a more general and more thorough manner.

In this philosophical perspective, three concepts will be involved, namely *gene*, *meme*, and *teme*. They are described below, respectively. According to Darwin's theory of evolu-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access

tion, all life evolves, and the fittest survives. On the other hand, Kevin Kelly, a philosopher of technology, argued that in addition to archaea, protists, eubacteria, fungi, plants, and animals, the *technium* had matured the seventh kingdom of life [31]. The concept of the '*technium*' was coined to embody the vast techno-social system, and the *technium* incorporates all the machines, society, culture, and philosophies connected with technologies. As '*technium*' evolves, it develops its own dynamics [32].

The concept of *teme* emerges from *meme*. So, what is *meme*? Builds upon the principal hypothesis of George C. Williams's *Adaptation and Natural Selection* (1966) [33], the biologist Richard Dawkin published his influential book the *Selfish Gene* (1976) [34]. In 1983, Dawkins may have first neologized the term "*universal Darwinism*" [35]. It is utilized to demonstrate his hypothesis that any possible life forms outside the solar system would evolve by natural selection just as they do on Earth.

Dawkins clarified that the replicator at the base of Darwinian selection does not have to be DNA. According to this insight, he isolated the second type of replicator from the *gene*, named *meme*(short for *memtics*), which is a basic unit of cultural transmission or imitation. *Meme* is a newly developed theory aiming to interpret the evolutionary mechanisms of culture from the perspective of Darwinism [36].

Like the other six organisms, technology also has a driving power of evolutionary, and the principle of "*survival of the fittest* " also applies to it. If the evolution of real-life depends on *genes*, and the inheritance of culture relies on *meme*, what is the carrier of technological evolution? Yes, it is *teme*.

Following Dawkins' lead, another British psychologist Susan Blackmore (known by a successful science book named *The Meme Machine*) demonstrated that there exist three replicators – genes (the basis of life), memes (the basis of human culture), and *temes* (the basis of technology) on the Earth, as Figure 3 shows.

In Blackmore view, a *teme* is a sort of special meme (short for technological memes). By *temes*, the information is stored, reproduced, modified and selected by electronic (or photon, or quantum) machines [37].

As we all know, the "*Elements*" is the most successful and influential mathematical book, written by the ancient Greek mathematician Euclid. It is the foundation of European mathematics. Literally, the "*element*" is the atomic unit of mathematical reasoning. Analogously, life, culture and even technology are their basic units about information transfer and inheritance , which are *gene*, *meme* and *teme*, respectively.

A gene is the fundamental physical and functional unit of heredity. Genes can be passed from ancestor to offspring and contain the information needed to specify traits [38]. Not all genes are retained and passed on to future generations over the long course of evolution. From the Darwinist perspective, people widely believed that natural selection of genes could be summarized as "*survival of the fittest*".

As mentioned earlier, Richard Dawkins put forward a new theory to explain the evolution of culture [39]. Dawkins argued that if an individual in any system ultimately produces or depends on a kind of information body, which has three traits of *heredity*, *variability* and *selectivity*, then evolution is bound to happen. Therefore, evolutionary is not limited to biological genes, but any replicator with the three characteristics described above.

As a result, Dawkins first proposed the framework "*meme*". A *meme* is an idea, style, or behavior, which spreads employing imitation from people to people within a culture and often conveys the symbolic meaning of representing a specific phenomenon or theme [40], [41]. According to the Oxford English Dictionary, *meme* is that an *element* of a culture that may be considered to be passed on by non-genetic means, esp. imitation.

There are obvious differences between *gene* and *meme*. *Gene* can move only vertically, from one generation to the next, via meiosis. Memes also can move vertically, but more often horizontally, within generations.

As for technology, Susan put forward the concept of *teme*, the third replicator, which is derived from *meme* [37]. In a nutshell, *temes* are the information that technologies carry as they evolve [42].

In the framework of *temes*, as one of the cutting-edge technologies in AI, deep learning (a kind of electronic algorithm) naturally belongs to the kind of new *life*, so it needs *temes* to maintain its survival and inheritance. As a result, it has a strong desire to reproduce themselves and have a robust internal evolutionary drive to weed out the lousy *temes*, which are not conducive to reproduction. In a sense, this is a more generalized "Universal Darwinism" [43].

IV. AN INTUITIVE THOUGHT EXPERIMENT

Philosophers often use thought experiments (mostly logical reasoning) to argue their points. The earliest logical thought experiment in history was Aristotle's syllogism [44]. A syllogism is a shred of logical evidence that applies deductive reasoning to reach a conclusion based upon two or more propositions asserted or assumed to be true [45].

Syllogistic arguments are usually represented in a three-line form. For example:

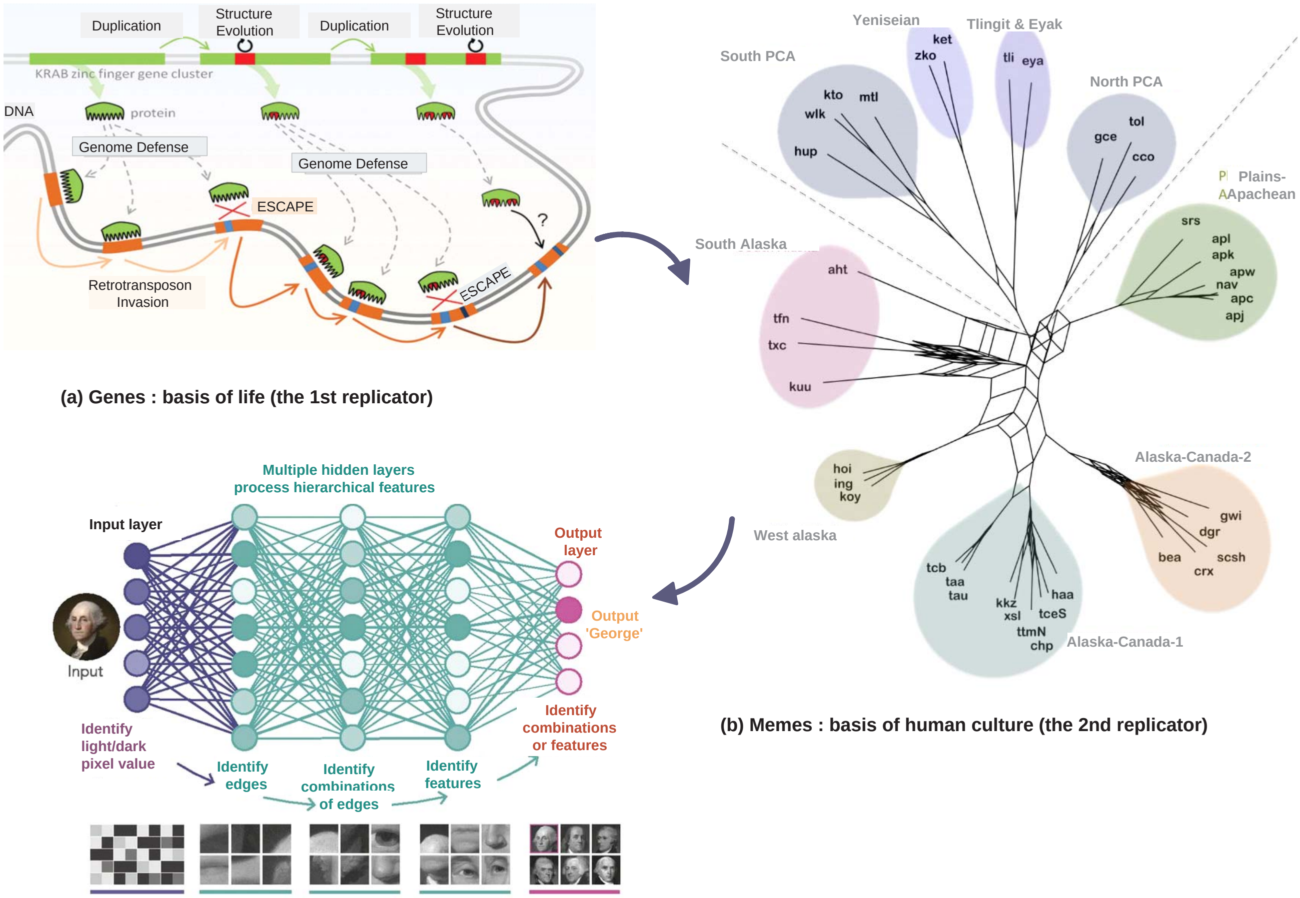
- *Major premise*: All humans are mortal, $\forall x(S(x) \rightarrow M(x))$.
- *Minor premise* : Aristotle is human, $S(p)$.
- *Conclusion* : Aristotle is mortal, $M(p)$.

There is a similar thought experiment of "syllogism" for deep learning technology:

- *Major premise*: All life follows the law of "*survival of the fittest*" (A).
- *Minor premise* : Deep learning technology is also a kind of "*life*" (B).
- *Conclusion* : Deep learning follows the law of "*survival of the fittest*" (C).

For major premise A, we have Darwin's theory of evolution as a guarantee. For minor premise B, we have made

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access



(a) Genes : basis of life (the 1st replicator)

(b) Memes : basis of human culture (the 2nd replicator)

(c) Temes : basis of technology (the 3rd replicator)

FIGURE 3. Three replicators on the Earth

a deduction in the related work of Section III. As a result, according to the logic of "syllogism", we can conclude that deep learning technology can also be bound by the law of "*survival of the fittest*". Therefore, there is a driving force of the "technical evolution" for deep learning; otherwise, it will be eliminated in the future of technology competition.

Now, the question is, (1) what are the *teme-related* defects that would put it at a disadvantage in evolutionary competition? Furthermore, (2) how to avoid or improve these "teme-related" defects. The answer to the latter question is the trends in deep learning driven by "*teme*". We will address these issues in the subsequent section.

V. TEME DEFECTS OF DEEP LEARNING

It seems that with unlimited data and countless computational resources, there might be a little need for any other theories in deep learning [46]. However, as François Chollet (the author of Keras, a well-known deep learning framework) said,

> "*For most problems where deep learning has enabled transformation with better solutions, we have entered diminishing returns territory in 2016-2017.*"

As with all influential technologies in their early-flowering, deep learning also needs to surmount a range of severe hurdles [47]. Consequently, concerns remain about its *teme-related* defects that may hinder deep learning from evolving for survival. Here, teme-related defects are used to describe the inherent imperfections of deep learning technology.

A. BRUTE FORCE

It is a well-known fact that the significant improvements in high-performance computing and the abundant supply of big data provide a strong basis for verifying empirical hypotheses. However, one of the most problematic aspects of deep learning is that it has immeasurable performance but not enough interpretability. In other words, deep learning itself is hard to articulate why it performs so well in specific fields, and it even has the nickname "*brute force reasoning*". There is another simple and straightforward interpretation of the nickname: "*quantity is quality*".

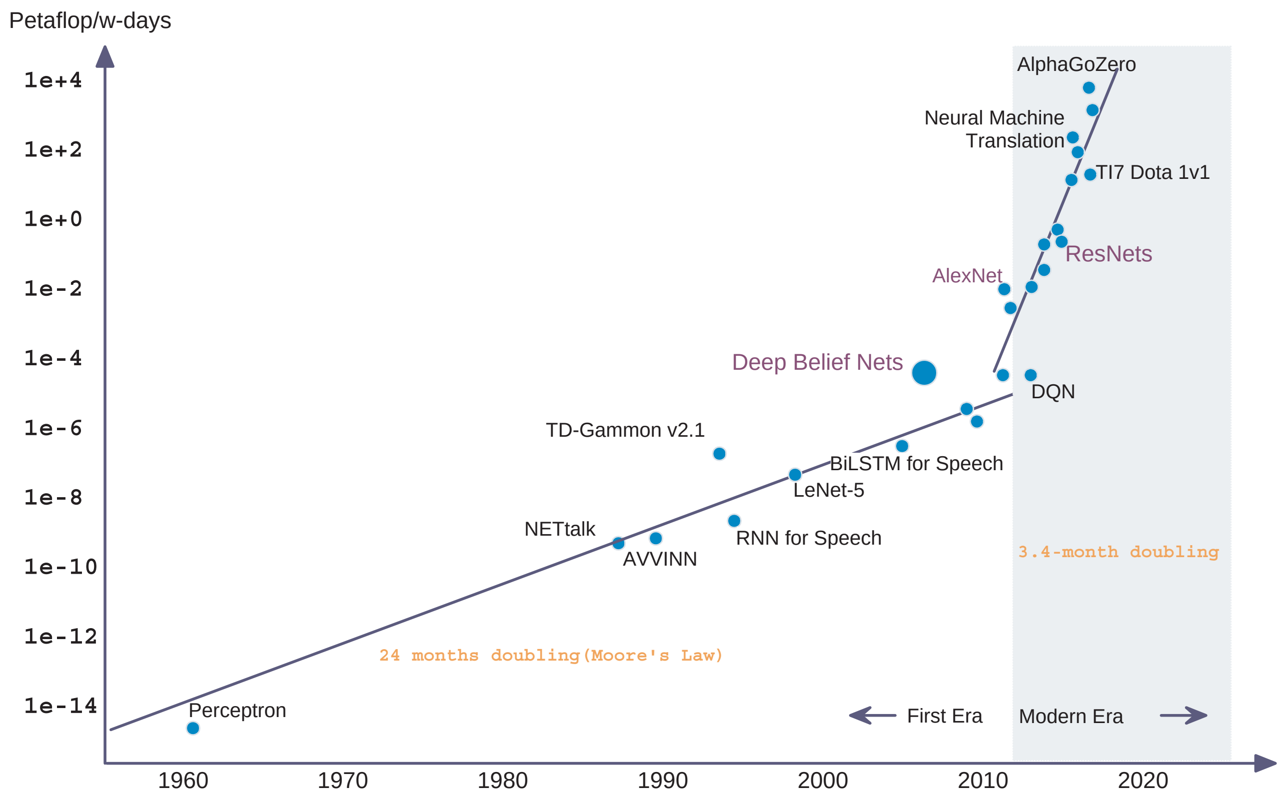


FIGURE 4. Two distinct eras of computation usage in training AI systems(credit : OpenAI [48])

Notwithstanding, deep learning performance depends mostly on the "parameter tuning" with algorithm engineering expertise. Even with the same algorithm, performance can vary widely due to diverse parameters. As a result, algorithmic engineers have to spend a lot of time and energy (that is, using brute force), like alchemy, on exploring with different neural network parameters hoping for better performance. Deep learning with "brute force" *teme* may not be the kind of intelligence we want. Machine learning's kernel task is to understand "learning" itself, but the current deep learning technologies have moved away from it.

At the International Conference on Machine Learning: Perspectives and Applications (ICML 2015) in Berlin, Vladimir Vapnik, a notable AI expert, offered a philosophical perspective: "ideas and intuitions come either from God or from the devil. The difference is that God is clever, while the devil is not. " [49] Vapnik proposed that both big data and deep learning have the *teme* of brute force.

Furthermore, one of the big slogans of deep learning is that it can be easily extended. AlexNet in 2012 had 60M parameters [3], but now many deep learning-based models have at least 1000 times the number of parameters compared to AlexNet. For example, recently, OpenAI has trained a Generative pre-trained Transformer 3 (GPT-3) model, an autoregressive language model that utilizes deep learning to generate human-like text. The model has 175 billion parameters [50].

Therefore, the questions remain, are these current models with massive parameters delivering more than a thousand-fold performance improvement? How about even 100 times? Another of OpenAI's research comes in handy [48].

As we can see from Figure 4, in the field of artificial intelligence, there is a doubling of computing power roughly every two years since the introduction of the *perceptron* in 1959. Nevertheless, since 2012, the trend has been more pronounced, with performance doubling every *3.4* months.

Figure 4 manifests that deep learning-based models cannot be scaled up to get better performance. A beneficial machine learning model does not require a large sample of train data or tremendous computing power. Nonetheless, the existing AI models do the opposite.

In the evolutionary chronicle of life, it is a liability to consume many resources to survive. The extirpation of the dinosaurs, for instance, is a piece of vivid evidence. For preaching the importance of saving living resources, some religious sects even specify *gluttony* as "*deadly sin*".

Cherishing living resources is not only a virtue for human beings but also technology. However, deep learning at present also contains some "*deadly sins*" - such as *overeating data*. Indeed, suppose deep learning remains complacent about its addiction to "big data". In that case, it may be a dead-end, and we will miss the opportunity that nature has given us to explore the fundamental principles of *learning*. When children accomplish cat recognition or dog, we all know that

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access

they only need some pictures or object observation to finish the learning. Children acquire knowledge based on small samples rather than big data. In contrast, models based on deep learning often expect hundreds of millions of sample training to achieve so-called learning tasks. Furthermore, overeating data also reflects that the current deep learning algorithms have a great "misunderstanding" of learning nature.

B. LOW "INTELLIGENCE-ENERGY DENSITY"

Based on the previous analysis, we can see that, from the perspective of "*temes*", the current deep learning technology has some "*teme-related*" defects such as "brute force" and overeating data. As stated above, technology can be regarded as the seventh *life* form. By the nature of sustaining a living being, any *life* requires energy expenditure. Deep learning also has another *teme-related* defect of overeating energy.

According to Google, AlphaGo's deep neural network model typically was trained over more than 16,000 processors. In the course of the match against humans, according to the energy consumed (set as parameter E), AlphaGo ran on 1920 CPUs and 280 GPUs. Estimated, on average, each CPU consumes 200 watts (W), and each GPU consumes $800W$, with a total consumption of more than $600,000$ W. Lee Sedol's brain, by contrast, consumes only $20W$. In terms of the record of Lee Sedol and AlphaGo, although Lee Sedol lost by $1:4$, on the whole, it can still be regarded as the same intellectual order of magnitude as AlphaGo (set as parameter I). Here we creatively propose a concept of intelligence-energy density (parameter set as D) :

$$D = \frac{I}{E} \tag{4}$$

With this simple paradigm, we can perceive that Lee Sedol's "*carbon-based*" brain can undoubtedly defeat "*silicon-based*" of AlphaGo. In terms of performance of intelligence-density, AlphaGo is just one in a hundred thousand of Lee Sedol. From this point of view, it may bring back a bit of human dignity.

Advances in techniques and hardware for training deep neural networks have recently impressive improvements across many fundamental Natural Language Processing (NLP) tasks, with the most computationally-hungry models.

Training the state-of-the-art NLP model based on deep learning requires many computing resources, which correspondingly necessitate considerable energy, appreciable financial burden, and even environmental cost. The estimated CO_2 emissions are listed in Table 1. It can be observed from Table 1 that the NLP model based on deep learning consumes huge energy [51].

More importantly, if the theory of *temes* concerning "evolution of technology‘' makes sense, then the low level of "intelligence-energy density" of deep learning, such as AlphaGo, will undoubtedly be challenging to survive. The technology with low energy consumption is the eternal pursuance of humankind. If the "*temes*" of deep learning desire to pass from one generation to the subsequent, such a *teme*, it must be improved by "artificial selection". A natural evolutionary tendency is that it must be democratized and universalized to enhance its "intelligence-energy density".

Consumption	**CO_2e (lbs)**
Air travel, 1 passenger, NY↔ SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
tuning & experimentation	78,468
Transformer (big)	192
neural architecture search	626,155

TABLE 1. The estimated CO_2 emissions from training common NLP models, compared to common consumption

C. SUPERFICIAL UNDERSTANDING

We comprehend that the concept of "*deep*" in deep learning refers to technical architectural features, or more specifically, it implies that neural network stacks lots of hidden layers. Nevertheless, for now, the "*deep*" does not refer to the deep learning having a *deep* understanding of abstract concepts.

Most of deep learning networks rely heavily on a technique, called *convolution* [52]. Unfortunately, this powerful technique also has a critical inherent *teme-related* defect, known as *translational invariance*. Invariance to translation implies that, if we translate the inputs, Convolutional Neural Network (CNN) can still determine correctly the class to which the input should belong. Take Figure 5 as an example. Regardless of the translation, rotation, and scaling of the stone lion image, it can still be categorized as a lion by using the property of translational invariance.

Translation invariance is the output of pooling operations [53]. Through pooling, the output of the convolutional network is the statistical result of a specific local region. To put it simply, the output of max-pooling is the maximum of a specific region. In contrast, the output of average pooling is the mean of this region, as exhibited in Figure 6.

As we replace the output with the maximum in max-pooling, even if we swap the input moderately, it may not influence the final result. For example, suppose that one-dimension pixel vector [1,3,4] inadvertently is changed as [1,4,3]. Its maximum is still "4". Consequently, translational invariance is a beneficial property where the object's location is not required precisely.

As a result, it can improve the robustness of classification in CNN. For instance, if a CNN model is built to detect faces, features such as eyes, nose and mouth, they are required to be detected regardless of where they appear. The pseudo-code of face recognition workflow by using CNN can be simplified and described in Algorithm 1:

If the CNN network can detect the face's features as accurately as possible, it can determine that it is a face or not.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access

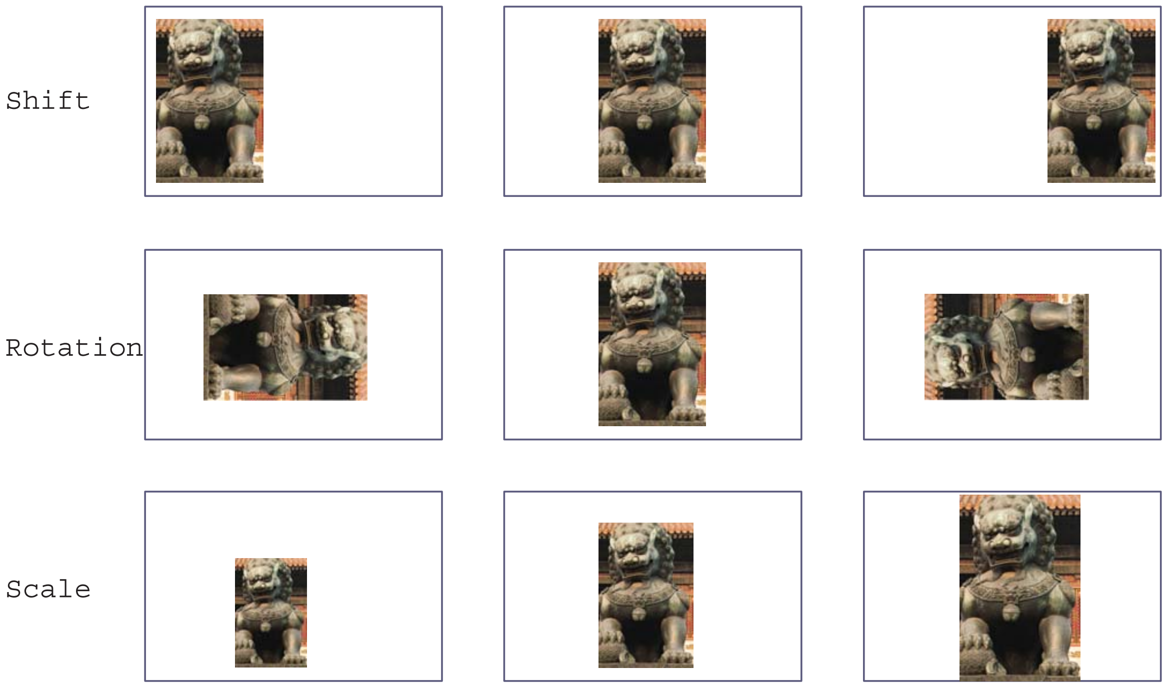


FIGURE 5. Translational Invariance Property in CNN

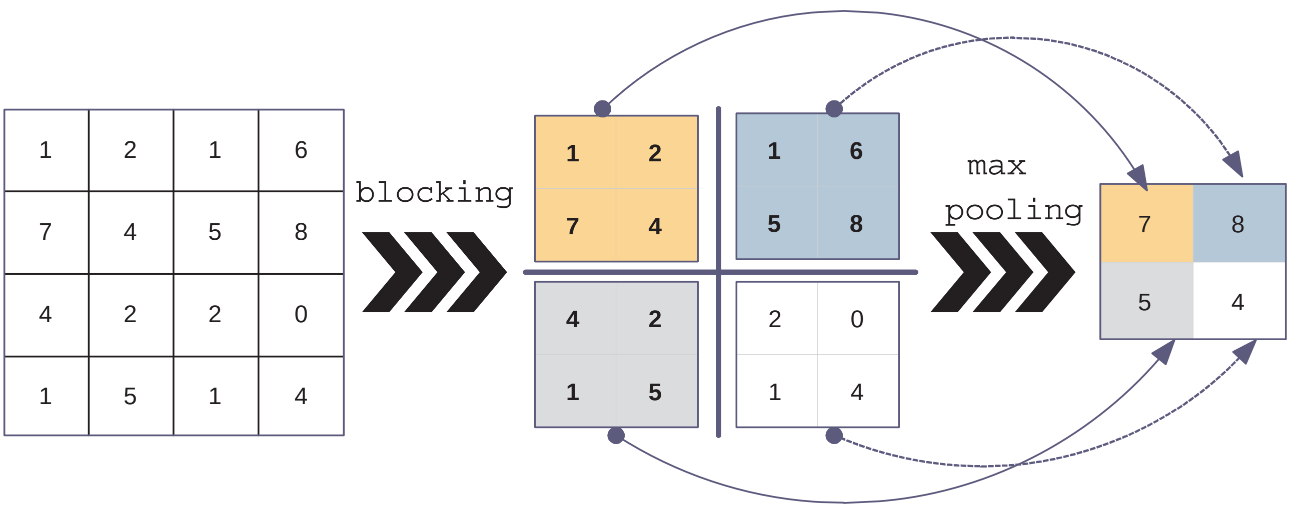


FIGURE 6. Pooling Operation in CNN

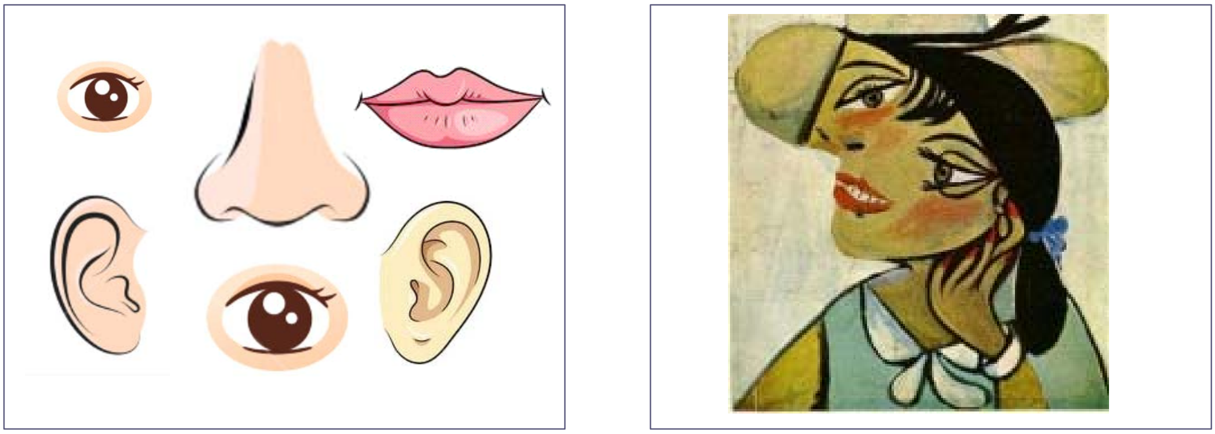
In most scenarios, the logic of face recognition described in Algorithm 1 is applicable.

On closer inspection, the above logic, however, is flawed. Refer to the two sub-graphs (a) and (b) in Figure 7. On the left is a completely misplaced portrait of face features, and on the right is a portrait of a woman with "*plausible*" features (painted by Picasso). For the two images, each human faces' partial features, such as eyes, nose, and mouth, exist objectively.

If the parts are checked separately, it would not be ugly. However, when all of them are put together somehow, these so-called "face" are unacceptable to ordinary humans. That is not what we want, so what is the problem? Their mistakes are that the face features' relative positions in the two images differ from ordinary people.

If the CNN algorithm, shown in Figure 7, is used to detect the two images, at high probability, it would tell that both are faces. However, humans may not identify with it. It indicates that there is a sizeable cognitive gap between the CNN algorithm and the human brain. As a result, it proves

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access



(a) The dislocated "face"   (b) The famous grotesque figure by Picasso

FIGURE 7. 'Faces' that change feature locations

Algorithm 1 The working principle of Face recognition in CNN

Input: Massive amount of trained face images
Output: Binary classification: face or not

face feature detection by using CNN;
if has a nose & has a mouth & has two eyes & has two ears **then**
  classify it as a face;
else
  this is the other object;
end if

that even if CNN works in most cases, it cannot cover up its inherent defects of *temes*: a superficial understanding of the input data.

CNN usually has a high accuracy rate in the classification task. However, is it the right goal for a deep learning algorithm to continuously improve the accuracy rate? Professor Geoffrey Hinton did not think so. According to Hinton's idea [54], an ideal facial recognition algorithm would look something like an Algorithm 2.

Algorithm 2 The working principle of Face recognition in CNN

Input: Massive amount of trained face images
Output: Binary classification: face or not

face feature detection by using CNN;
if has two adjacent eyes & has one nose under eyes & has one a mouth under nose & has two symmetrical ears **then**
  classify it as a face;
else
  this is the other object;
end if

In addition to identifying features, the algorithm is also expected to understand the spatial hierarchy among features. That is to say, a higher level of classification algorithm should depend on the appropriate representation of content rather than the simple feature detection.

Once a good representation of the context is found, the context will have a solid understanding. This better representation can then be used for pattern recognition, semantic analysis, building abstract logic. A good understanding of representation can achieve a more advanced goal of AI. Although the translation invariance brought by CNN's pooling strategy makes classification more reliable, it pays the price for the loss of pose information, which is exactly conducive to the excellent representation of content. Here pose information refers to 3D orientation relative to the viewer but also the lighting and the coloring. CNNs are known to have a problem when objects are rotated in 3D space or when lighting conditions are changed.

As for Figure 8, it is easy for humans to identify the lion, no matter how it rotates in high dimensional "pose space". Simultaneously, it is difficult for the CNN classification algorithm to distinguish the lion in a 3D space rotation. Because people can understand the content of an image, whereas the CNN algorithms only can detect a few key features, once the features change, it becomes way above the pay grade.

Hinton ever commented that CNN has a high accuracy rate of classification. On the surface, it appears that it is a good situation, but a disaster. Accordingly, Hinton asserted: "*CNN is doomed to have no future*". Hinton's critique of CNN's shortcomings goes hand in hand with a possible solution, known as the Capsule Neural Network (CapsNet) [55]–[57].

VI. DEEP LEARNING FROM A DATAISM PERSPECTIVE

As can be seen from the previous analysis, deep learning is essentially a kind of electronic algorithm that "*overeats*" data and energy. In other words, deep learning is a data-hungry algorithm, which is undoubtedly one of the representative

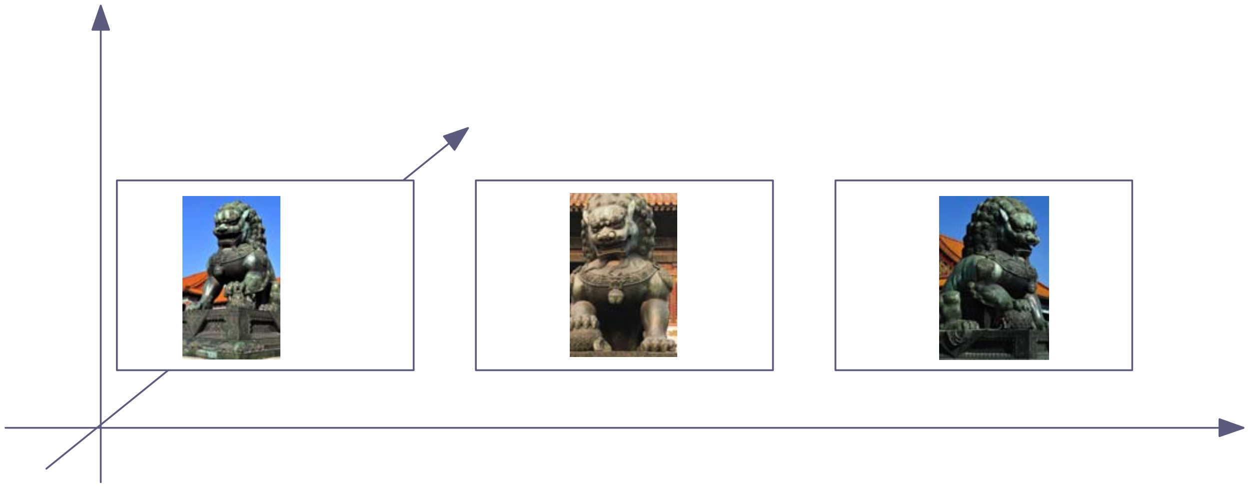


FIGURE 8. A lion rotating in 3D space

works of *dataism*.

Dataism is a terminology that was neologized for describing the emerging of big data and practiced to illustrate the significance of data philosophy [58]. *Dataism* asserts that the universe consists of data flows, and its contribution to data processing determines the value of any phenomenon or entity [59].

In the 150 years since Charles Darwin published *On the Origin of Species*, life sciences tend to regard biology as a biochemical algorithm. Simultaneously, in the 80 years since Alan Turing came up with the Turing machine concept, computer scientists have learned to design increasingly complex electronic algorithms.

Now, *dataism* tries to put the two together and points out that the same mathematical laws apply to both the biochemical and the electronic algorithms. That is to say, *dataism* collapses the barrier between life and machine. Logically, both the biochemical and the electronic algorithms share the same desires: survive, grow, and reproduce (evolution). To achieve this, they both have an intrinsic developmental drive and an evolutionary urge.

According to the view of universal evolution, "survival of the fittest", whether *genes* of biochemical algorithm or the *temes* electronic algorithm, whichever do not adapt to the *environment*, will be mercilessly ostracized by *nature*.

Technology is a complex system and an evolving process [60]. Any technology, including deep learning, is not perfect. It has its own inherent *teme-related* defects. As a consequence, it necessitates being on the road of self-reconstruction and evolution perpetually. The current deep learning is no exception to technology evolution. There is a rigorous demand for deep learning to improve its own "*temes*" and accept the baptism of evolution.

VII. THE EVOLUTIONARY DIRECTION OF DEEP LEARNING

At present, deep learning still has a utilization market, but it gradually parted away from science. The situation means that, if we neglect AI science, we are doomed to repeat the history of "AI winter". It is as if *genes* determine what we will look like in the future, and *temes* will define what technology will look like in the future, unless the *genes* or *temes* are being continuously improved. For doing this, technological evolutionary for deep learning may be carried out from the following three directions.

A. COMBINATION OF FUNCTIONALITY AND INTERPRETABILITY

The causation and interpretability of the world still play a decisive role for humans in achieving knowledge breakthroughs. Because of the data, the machine can only discover the unknown parts of the current knowledge domain. Without an explainable model, however, the upper boundary of the current domain of knowledge is restricted by the linearly increasing computational power of machines, which cannot be extended to a new knowledge domain.

The success of deep learning is, to no small extent, a triumph of utilitarianism in the era of big data. In typical implementation scenarios, deep learning models have millions or even billions of parameters. The deep learning models became more complex, losing their adaptability and transparency. As a result, these models are identifiable to their developers not in terms of human interpretable labels, but only in terms of their topology within a complex network [18].

Although these models, to a certain extent, can be double-checked in real data, the laissez-faire attitude of the pursuit of utilitarianism rather than the interpretability is a kind of degeneration of human beings, which carries substantial potential risks in the future.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access

On the other hand, adversarial samples' existence may lead to the failure of the pre-training model [61], and the irrationality of the training data will also be magnified locally. To further improve the algorithm performance, optimize the network parameters, and enhance the model's generalization ability, the deep learning model needs to be interpretable. In conclusion, explainable AI (XAI) is becoming critical for deep learning [62].

In terms of the interpretability of deep learning, some researchers have made some reasonable attempts. For example, some researchers have utilized the information bottleneck theory to open the black box of deep learning [63]. At the same time, some scholars take advantage of self-explanatory neural models that generate natural language explanations. In other words, deep learning models that have a built-in module that generates explanations for the predictions of the model [64].

In the future, to overcome the "*teme-related*" defect of superficial understanding, the incorporation of functionality and interpretability will be a promising research direction. Indeed, the current success of AlphaGo, to some extent, depends on both connectionism based on deep learning and interpretable logical reasoning based on the Monte Carlo Tree Search [65], [66].

B. ACTIVE LEARNING BASED ON SMALL DATA

Deep learning is the representative of connectionism. It should be noted that the original bionic subject of connectionism was the human brain. Now, human has evolved into the most intelligent life on Earth. Therefore, we should not forget the original intention of AI. Some of the outstanding human characteristics can still be a guide to AI research.

In evolution, human beings can create so many new things and are good at active learning based on small data [67] and excel at extrapolates based on transfer learning [68]. Active learning is any learning activity in which one participates or interacts with the learning process. For efficiency, transfer learning is often used to solve one problem and apply it to a different but related problem. In a relatively new field where there is not much data and experience to learn, transfer learning is vital.

In essence, whether active learning or transfer learning is an economical learning style, saving data, computational power, and training time. The two kinds of learning are the active choice of human beings under the condition of limited resources. For machine learning, "limited resource" means that learning can only be done with limited labeled data and computing resources.

At present, the ability of AI, in terms of transfer learning and active learning, is still wanted. For instance, even though AlphaGo can beat any human Go player, but if we slightly change the rules of chess, such as changing the board's shape from square to hexagonal, AlphaGo will suddenly be catastrophically forgotten. Its hard-earned "intelligence" will be useless immediately. If AlphaGo is expected to perform well in new environments, it must be trained again using new training data from scratch.

However, this "*revolutionary*" technological iteration does not conform to the evolutionary laws of nature. Therefore, a higher-order algorithm needs to have the learning ability such that it is based on small data and the initiative to find valuable data. It follows the continuous technological evolution strategy to achieve the purpose of self-active learning.

C. SPECIAL CHIP FOR ENERGY-SAVING

The energy consumption of data transfer among machines, CPU processes, and threads is a big deal [69]. To save energy, let us take some inspiration from the calculations of biological brains. There is a synaptic gap between the axon's end and the adjacent dendrites in the living brain's neural network, chemicals called neurotransmitters spread. In this way, the nervous system completes the signal (namely, *data*) transmission among neurons. This localized *data* transmission significantly saves a lot of the brain's energy, which is the body's most energy-consuming organ.

In contrast, the current artificial neural network system (deep learning) relies on computing power to transmit all data, and every neuron in the extensive network is traversed and accessed. This kind of data passage significantly consumes energy and increases the difficulty of neural network training. Therefore, it is sensible to conclude that if the current artificial neural network has found a brain-like data transmission mechanism, it will be much more productive and significantly lesser energy consumption.

May benefit from biological brain inspiration, a method called *dropout* was proposed by Srivastava and Hinton et al. [70]. Its key idea is to randomly drop units (along with their connections) from the neural network during training. To some extent, it reduces the number of neurons involved in the training process. Further, it is a way of preventing overfitting. However, it is a random disabling mechanism, far from being a conscious local connection of biological neurons. We still have a long way to go to take a cue from biology.

From the perspective of technological evolution, technologies with low "intelligence-energy density" are challenging to be reproduced and inherited. Therefore, to maintain the continuity of electronic algorithms, special energy-saving hardware is also called upon to save more energy consumption and enhance deep learning's "intelligence-energy density". In this respect, Chinese computer scientists have made a reasonable endeavor. They proposed a dedicated processor instruction set for deep learning and based on this instruction set. They designed the deep learning chip named Cambrian, whose work area and power consumption is only one-hundredth of mainstream GPUs [71].

VIII. CONCLUSIONS AND FUTURE WORK

Artificial intelligence, represented by deep learning, has made remarkable achievements in recent years. Nevertheless, deep learning has many *teme* defects, such as "brute force", low "intelligence-energy density", and superficial un-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3032143, IEEE Access

derstanding. From the perspective of technological evolution, these *teme-related* defects may lead to the stagnation of deep learning. For avoiding such a situation, it is essential to improve the defects in these aspects.

In this paper, we also discuss several workable directions for improving the *teme-related* deficiencies of deep learning. To this extent, many advances in computer science, brain science, cognitive science, psychology, and philosophy are also expected to combine with deep learning. Only in this way can we go further on the road to the technological evolution of AI.

In the book *Sapiens: A Brief History of Humankind* [72], Yuval Harari constructed a new historical evolution framework based on "common virtual" and dataism. This new framework reveals the laws of "evolution of evolution" related to *memes* and *temes*. In the future, we will use this novel perspective to investigate the technological evolution of deep learning further.

REFERENCES

[1] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, et al., "Deep speech 2: End-to-end speech recognition in english and mandarin," in International conference on machine learning, pp. 173–182, 2016.
[2] M. Popel, M. Tomkova, J. Tomek, Ł. Kaiser, J. Uszkoreit, O. Bojar, and Z. Žabokrtskỳ, "Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals," Nature communications, vol. 11, no. 1, pp. 1–15, 2020.
[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.
[4] Z.-N. Li, C. Zhu, Y.-L. Gao, Z.-K. Wang, and J. Wang, "Alphago policy network: A dcnn accelerator on fpga," IEEE Access, 2020.
[5] B. Liu, C.-C. Li, and K. Yan, "Deepsvm-fold: protein fold recognition by combining support vector machines and pairwise sequence similarity scores generated by deep learning networks," Briefings in bioinformatics, vol. 21, no. 5, pp. 1733–1741, 2020.
[6] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," IEEE Access, vol. 5, pp. 21954–21961, 2017.
[7] A. R. Fayjie, S. Hossain, D. Oualid, and D.-J. Lee, "Driverless car: Autonomous driving using deep reinforcement learning in urban environment," in 2018 15th International Conference on Ubiquitous Robots (UR), pp. 896–901, IEEE, 2018.
[8] Y. Wang, L. Wang, H. Wang, and P. Li, "End-to-end image super-resolution via deep and shallow convolutional networks," IEEE Access, vol. 7, pp. 31959–31970, 2019.
[9] C. Francois, Deep learning with Python. Manning Publications Company, 2017.
[10] W. G. Hatcher and W. Yu, "A survey of deep learning: Platforms, applications and emerging research trends," IEEE Access, vol. 6, pp. 24411–24432, 2018.
[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436–444, 2015.
[12] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," IEEE journal of biomedical and health informatics, vol. 21, no. 1, pp. 4–21, 2016.
[13] C. Glymour, R. Scheines, and P. Spirtes, Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling. Academic Press, 2014.
[14] J. Copeland, Artificial intelligence: A philosophical introduction. John Wiley & Sons, 2015.
[15] A. Sloman, The computer revolution in philosophy: Philosophy, science and models of mind. Author, 2019.
[16] C. Buckner, "Deep learning: A philosophical introduction," Philosophy Compass, vol. 14, no. 10, p. e12625, 2019.
[17] S. E. of Philosophy, "Computational philosophy," 2020.
[18] G. Marcus, "Deep learning: A critical appraisal," arXiv preprint arXiv:1801.00631, 2018.
[19] F. Chollet, "The limitations of deep learning," Deep Learning With Python, 2017.
[20] I. Wladawsky-Berger, "Deep learning: Is it approaching a wall?," The wall street journal, Sep 2018.
[21] M. M. Waldrop, "News feature: What are the limits of deep learning?," Proceedings of the National Academy of Sciences, vol. 116, no. 4, pp. 1074–1077, 2019.
[22] G. Cybenko, "Approximation by superpositions of a sigmoidal function," Mathematics of control, signals and systems, vol. 2, no. 4, pp. 303–314, 1989.
[23] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in ICML, 2010.
[24] D.-X. Zhou, "Universality of deep convolutional neural networks," Applied and computational harmonic analysis, vol. 48, no. 2, pp. 787–794, 2020.
[25] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in Advances in neural information processing systems, pp. 396–404, 1990.
[26] G. F. Montufar, R. Pascanu, K. Cho, and Y. Bengio, "On the number of linear regions of deep neural networks," in Advances in neural information processing systems, pp. 2924–2932, 2014.
[27] C. Olah, "Neural networks, manifolds, and topology," 2014.
[28] Y. Tom, H. Devamanyu, P. Soujanya, and C. Erik, "Recent trends in deep learning based natural language processing [review article]," IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55–75, 2018.
[29] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018.
[30] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent trends in deep learning based personality detection," Artificial Intelligence Review, 2019.
[31] K. Kelly, The technium. Electronic Industry Press.
[32] Z. Baber, "Society: The rise of the'technium'," Nature, vol. 468, no. 7322, pp. 372–373, 2010.
[33] G. C. Williams, Adaptation and natural selection: A critique of some current evolutionary thought, vol. 75. Princeton university press, 2018.
[34] R. Dawkins, The selfish gene. Oxford university press, 2016.
[35] R. Dawkins et al., Universal darwinism. Cambridge University Press, 1983.
[36] R. R. Nelson, "Universal darwinism and evolutionary social science," Biology & Philosophy, vol. 22, no. 1, pp. 73–94, 2007.
[37] S. Blackmore and R. Dawkins, The meme machine, vol. 25. Oxford University Press, 2000.
[38] A. J. Griffiths, S. R. Wessler, R. C. Lewontin, W. M. Gelbart, D. T. Suzuki, J. H. Miller, et al., An introduction to genetic analysis. Macmillan, 2005.
[39] L. Gabora, "Meme and variations: A computational model of cultural evolution," in 1993 Lectures in complex systems, pp. 471–485, Addison Wesley, 1995.
[40] R. Nelson, "Evolutionary social science and universal darwinism," Journal of evolutionary economics, vol. 16, no. 5, pp. 491–510, 2006.
[41] L. Gabora, "The origin and evolution of culture and creativity," Journal of Memetics: Evolutionary Models of Information Transmission, vol. 1, no. 1, pp. 1–28, 1997.
[42] L. Bull, O. Holland, and S. Blackmore, "On meme–gene coevolution," Artificial life, vol. 6, no. 3, pp. 227–235, 2000.
[43] G. M. Hodgson, "Generalizing darwinism to social evolution: Some early attempts," Journal of Economic Issues, vol. 39, no. 4, pp. 899–914, 2005.
[44] J. Levy, The Infinite Tortoise: The Curious Thought Experiments of History's Great Thinkers. Michael O'Mara Books, 2016.
[45] G. Patzig, Aristotle's theory of the syllogism: A logico-philological study of Book A of the Prior analytics, vol. 16. Springer Science & Business Media, 2013.
[46] C. Anderson, "The end of theory: The data deluge makes the scientific method obsolete," Wired magazine, vol. 16, no. 7, pp. 16–07, 2008.
[47] G. Marcus, "Innateness, alphazero, and artificial intelligence," arXiv preprint arXiv:1801.05667, 2018.
[48] D. Amodei and D. Hernande, "Ai and compute," 2018.
[49] Z. Lipton, "Does deep learning come from the devil?," 2015.
[50] W. Heaven, "Openai's new language generator gpt-3 is shockingly good—and completely mindless," 2020.

[51] E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in nlp," arXiv preprint arXiv:1906.02243, 2019.
[52] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural computation, vol. 1, no. 4, pp. 541–551, 1989.
[53] Y. Zhang, The Beauty of Deep learning: Data processing and best practices in the AI era. Electronic Industry Press, 2018.
[54] D. Elton, "Geoffrey hinton on what's wrong with cnns," 2018.
[55] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in Advances in neural information processing systems, pp. 3856–3866, 2017.
[56] G. E. Hinton, S. Sabour, and N. Frosst, "Matrix capsules with em routing," in International conference on learning representations, 2018.
[57] A. Kosiorek, S. Sabour, Y. W. Teh, and G. E. Hinton, "Stacked capsule autoencoders," in Advances in Neural Information Processing Systems, pp. 15512–15522, 2019.
[58] D. Brooks, "The philosophy of data," New York Times, vol. 4, p. 2013, 2013.
[59] K. Kelly, What technology wants. Penguin, 2010.
[60] W. B. Arthur, The Nature of Technology: What It Is and How It Evolves. 2009.
[61] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," arXiv preprint arXiv:1312.6199, 2013.
[62] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K. R. Muller, "Explainable ai: Interpreting, explaining and visualizing deep learning," Lecture Notes in Computer ence, 2019.
[63] R. Shwartz-Ziv and N. Tishby, "Opening the black box of deep neural networks via information," arXiv preprint arXiv:1703.00810.
[64] O.-M. Camburu, Explaining Deep Neural Networks. PhD thesis, University of Oxford, 10 2020.
[65] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., "Mastering the game of go with deep neural networks and tree search," nature, vol. 529, no. 7587, pp. 484–489, 2016.
[66] J. X. Chen, "The evolution of computing: Alphago," Computing in Science & Engineering, vol. 18, no. 4, pp. 4–7, 2016.
[67] B. Settles, "Active learning literature survey," Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
[68] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1345–1359, 2009.
[69] H. Xiong, D. Zhang, L. Wang, and H. Chaouchi, "Emc 3: Energy-efficient data transfer in mobile crowdsensing under full coverage constraint," IEEE Transactions on Mobile Computing, vol. 14, no. 7, pp. 1355–1368, 2014.
[70] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," The journal of machine learning research, vol. 15, no. 1, pp. 1929–1958, 2014.
[71] S. Liu, Z. Du, J. Tao, D. Han, T. Luo, Y. Xie, Y. Chen, and T. Chen, "Cambricon: An instruction set architecture for neural networks," in 2016 ACM/IEEE 43rd Annual International Symposium on Computer Architecture (ISCA), pp. 393–405, IEEE, 2016.
[72] Y. N. Harari, Sapiens: A brief history of humankind. Random House, 2014.

YUHONG ZHANG was born in Henan province of China in 1980. He received a B.S. in mechanical and electronics from Xi' an University of Technology, China, in 2001, and M.S. degrees in traffic control engineering from Chang'an University, China in 2004, a Ph.D. degree in computer science at the University of Electronic Science of China in 2012.

From 2009 to 2011, he was a research fellow at Northwestern University at the High-Performance Computing Laboratory, IL, USA. Since 2012, he has been an Assistant Professor with the College of Information Science and Engineering, Henan University of Technology, China. Since 2019, he has been a visiting scholar at Indiana University Purdue University at Indianapolis, USA, engaged in deep learning and philosophy of science and technology. He is the author of 10 books, more than 30 articles. His research interests include deep learning, big data, and technology philosophy.

UMER NAUMAN was born in 1992 in Pakistan. He is currently a first-year student at Henan University of technology's Ph.D. program, Grain Big Data Analysis. He is passionate about exploring the in-depth knowledge of data analysis and artificial intelligence. His data analysis skills and theoretical knowledge about big data and business intelligence set up apart from other professional and ethical students. After completing his Bachelors's degree in Computer Engineering, he joined UNICEF as a data analyst from 2014-2017 in Pakistan. The project he worked upon was the POLIO Eradication program, and his Data Analysis skills made the project a success. He came to China in September 2017 as a Master's student via the Chinese Government Scholarship (CSC) and successfully published three research papers during his Master's degree.