Applying Principles from Medicine Back to Artificial Intelligence

Howard Schneider

Contents

Abstract

As artificial intelligence (AI) advances emerge, they are often incorporated into the field of medicine. However, the reverse of this process is an important one also—many of the breakthroughs in the field of AI are due to knowledge and inspiration *from* biology, psychology, and encompassing medical sciences. The artificial neuron, one of the first achievements in AI, is discussed. Knowledge about the visual cortex inspiring the creation of the neocognitron leading to the convolutional neural network, leading to the deep learning revolution in AI, is discussed. Cognitive architectures, which intentionally incorporate knowledge from the medical sciences, including biology and psychology, into a form that can add to the field of AI, are reviewed. While deep learning models can recognize patterns and play games at a human-like level, they are poor at causal understanding and solutions, especially if there are limited numbers of training examples. The causal cognitive architecture is reviewed. This architecture stores information in the form of navigation maps, which it can update and link/retrieve to/from other navigation maps, and produces precausal

Supplementary Information: The online version of this chapter (https://doi.org/10.1007/978-3-030-58080-3_289-1) contains supplementary material, which is available to authorized users.

H. Schneider (✉)
Sheppard Clinic North, Toronto, ON, Canada
e-mail: hschneidermd@alum.mit.edu

N. Lidströmer, H. Ashrafian (eds.), *Artificial Intelligence in Medicine*,
https://doi.org/10.1007/978-3-030-58080-3_289-1

behavior. If intermediate maps from its navigation center are fed back, stored, and operated on again, then full causal behavior can emerge. The architecture provides hypotheses and predictions about the emergence of schizophrenia in humans. The ability to navigate a world of concepts, explainability, and analogies also emerge from this architecture. Inspiration from the medical sciences can continue to provide AI with future breakthroughs.

Keywords

Artificial intelligence · Causal · Causality · Cognitive Architecture · Deep Learning · Explainability · Machine Learning · Perceptron · Psychosis · Schizophrenia

1 Introduction

As advances in artificial intelligence have developed over the last half century, they have been applied to the field of medicine with the hopes of automating many processes in medicine, resulting in better research towards an understanding of diseases and their treatments and a lower cost, higher quality health care service to patients. Consider, for example, the development of expert systems and deep learning and their application to medicine.

An expert system obtains knowledge, essentially if-then rules, from a human expert about some field and contains an inference engine that logically uses the if-then rules to answer questions or solve problems posed to it. Buchanan and colleague's DENDRAL expert system, which helped to identify an unknown organic chemical, was created in the late 1960s [1]. By the early 1970s, a medical application of an expert system, MYCIN, had been developed [2]. In response to answers to questions it asked, MYCIN could identify bacteria causing severe infections and recommend specific antibiotic treatments [3]. MYCIN had several hundred rules, used a straightforward inference engine, and produced reasonable responses for its very narrow range of expertise. However, it never was used in a practical clinical sense. Expert systems such as MYCIN raised hopes for the practical use of artificial intelligence in all aspects of science, technology, and commerce. By the 1980s, expert systems had become a major focus of the field of artificial intelligence, and international commerce and competition in the field had started. However, by the early 1990s, it had become apparent that while it was possible to build interesting demonstration applications with expert systems, they largely failed to produce useful results for the complex tasks they had been promised for, medical applications included. The early 1990s period is often called the "AI winter," with analogy to the term "nuclear winter," in which the expert system companies failed, and institutional research funding was cut back in the field [4].

Deep learning generally refers to machine learning using many layers of artificial neural networks (ANNs); i.e., there are middle layer(s) of artificial neural networks between the input and output layers. The theory and technology behind various machine learning approaches, including ANNs with multiple hidden middle layers, had started to greatly improve in the mid-2000s. In 2012, work by Krizhevsky, Sutskever, and Hinton using deep learning won a computer vision competition by a large margin over older methods [5]. In the ImageNet contest, a computer-based system needed to classify as accurately as possible over a million different images into some thousand different classes. Hinton and colleagues used a deep convolutional neural network. In such a neural network, there are multiple layers of artificial neurons connected with layers where the artificial neurons act as convolutional layers where such layers extract features from the previous layers, preserving the spatial relationships but essentially mapping into a small-size receptive field and extracting features as such. This achievement of Hinton and colleagues is regarded as an approximate start of what is called the "deep learning revolution" and propelled the utilization of deep learning into many domains, including, of course, just about every branch of medicine. For example, consider the field of schizophrenia research and patient care. A review by Veronese and colleagues in 2013 gives an overview of machine learning approaches in schizophrenia but describes little of deep learning [6]. However, within a few years, deep learning was being widely used in the field. In 2016, Kim and colleagues noted that deep neural networks (DNNs) with multiple hidden layers were performing much better in classification tasks compared to support vector machines (SVMs) and earlier AI models. Kim and colleagues used a DNN classifier of resting-state functional magnetic resonance imaging to diagnose schizophrenia patients from healthy controls and showed better results than the same analysis performed with an SVM classifier [7].

As artificial intelligence enhancements and technologies emerged, they eventually became incorporated into the field of medicine, reaching the point where there is now a scientific discipline of "artificial intelligence in medicine." The journal *Artificial Intelligence in Medicine* defines the field as "the scientific discipline pertaining to research studies, projects and applications that aim at supporting decision-based medical tasks through knowledge- and/or data-intensive computer-based solutions that ultimately support and improve the performance of a human care provider" [8]. However, in this chapter, the reverse idea is considered. Rather than advances in the field of artificial intelligence being applied to medicine, there is examination here of the principles and advances in the medical sciences, including biology and psychology, being applied *back* to the field of artificial intelligence.

In this chapter, some examples of concepts from the medical sciences being applied *back* to the field of artificial intelligence are reviewed (Video 1):

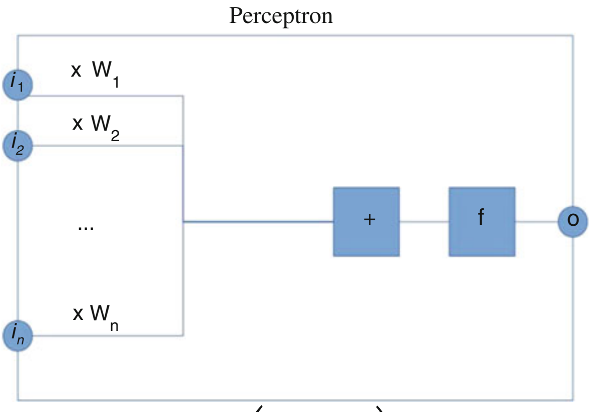
(i) Neural networks
(ii) Cognitive architectures
(iii) Causal cognitive architectures

2 Concepts from the Medical Sciences Being Applied *Back* to the Field of Artificial Intelligence: Neural Networks

The basis of the "deep learning revolution" described above is the artificial neural network. As is obviously implied by the name, neural networks were inspired by biological neurons. Neural networks were one of the earliest contributions to the field of AI. In 1943, Warren McCulloch and Walter Pitts took the basics of biological neuronal functioning and applied this knowledge to artificial neurons, which could be on or off and could activate other neurons [9]. They described a "logical calculus" for neural networks of their artificial neurons. A few years later, Donald Hebb described a method on how neurons could make connections with each other, i.e., "neurons that fire together wire together" [10], and an early learning mechanism arose that could be used with artificial neurons.

The artificial neuron, of course, does not simulate even a fraction of the complexity of an actual real neuron, but what it does do was sufficient for the emergence of neural networks. Rosenblatt's "perceptrons" continued the use of these simplified neurons as functional elements of networks that could learn to classify inputs as belonging to a particular class [11]. Figure 1 shows a functional schematic of a perceptron. Inputs i_1 to i_n are multiplied by weights $\mathbf{w}$, which are summed, and then passed through a function producing output o. Unfortunately, due to limitations in the way the perceptron networks were constructed (e.g., single-layer perceptron machine), it was shown that these simple neural networks could not function well enough to solve interesting problems [12]. Years later, more powerful and functional multilayer deep learning neural networks emerged, all with roots in the biological origins of the work of McCulloch and Pitts and which, as noted above, at the time of this writing are responsible for many of the modern applications of artificial intelligence, including applications in medicine.

Work by Hubel and Wiesel on the receptive fields of simple and complex cells in the visual



$$o = f\left(\sum_{k=1}^{n} i_k \cdot W_k\right)$$

Fig. 1 Functional diagram of a perceptron, i – input, w – weight, f – function, o – output. (Creative Commons License BY-SA. Credit to Mat W.)

cortex [13] inspired Fukushima’s neocognitron [14]. The neocognitron has simple (S) and complex (C) cells, with local features discerned by the simple cells and integrated into higher layers, and as such detects patterns. The neocognitron in turn inspired the convolutional neural network, which, as noted above, was used by Krishevsky, Sutskever, and Hinton in their deep learning neural networks [5], which is often taken as the start of what is called the “deep learning revolution” and resulted in the use of deep learning in many domains, including medicine.

3 Concepts from the Medical Sciences Being Applied *Back* to the Field of Artificial Intelligence: Cognitive Architectures

A cognitive architecture attempts to formally put a myriad of results from the medical sciences, including biology and psychology, into a model that gives a structure and functioning of the mind. As a result of the formal nature of the model, it can be implemented as a computer program, i.e., a computer program that exhibits varying degrees of intelligence. Concepts from the medical sciences, including biology and psychology, thus end up being applied back to the field of artificial intelligence.

ACT-R is a long-standing popular cognitive architecture [15]. ACT-R models the human mind as declarative knowledge (“chunks” which represent properties and are held in buffers) and procedural knowledge (“productions” which represent procedures to do various operations). There are two main types of long-term memory modules—declarative memory (i.e., facts, e.g., the pencil is brown) and procedural memory (i.e., productions, e.g., how to sharpen a pencil). Unlike typical databases in computer science at the time which stored mainly facts, note that ACT-R’s memory is or can be substantially comprised of productions, i.e., procedures, as well. Just as humans have a short-term working memory, ACT-R also has a working memory. Depending on the internal state of ACT-R, which depends on the values in its various modules, including modules that interface with the real world, a best production will be selected and executed, and this in turn will change the internal state of the ACT-R, and then another production will be selected and executed, and so on. A cognitive architecture such as ACT-R is not just a collection of concepts but exists as a computer program, which can be run as well as modified. ACT-R has evolved over the years from 1973 to 2019 incorporating additional, modified, or newer theories of cognition, and also, its software implementation has changed [16].

The operation of a cognitive architecture such as ACT-R is inspired by the medical sciences. For example, in a recent ACT-R model, in its production system, the matching of productions is inspired by the striatum, the selection of a production is inspired by the pallidum, and the execution of a production is inspired by the thalamus [16]. In turn, the ACT-R contributes to the field of artificial intelligence, in which there are many diverse applications, e.g., human-robot interactions [17] or flying an airplane [18].

A variety of cognitive architectures exist. Samsonovich in 2010 and Kotseruba and colleagues in 2016 reviewed many of these different architectures [19, 20]. Ritter and colleagues organized cognitive architectures into five main groups [16]:

1. Cognitive architectures using advanced knowledge structures such as plans, with an emphasis on performance, e.g., JACK multiagent systems [21].
2. Symbolic architectures with emphasis on modeling human cognition, e.g., Soar cognitive architecture [22] (although some versions of Soar also have subsymbolic operations and would be thus classified in the hybrid group below)
3. Subsymbolic/connectionist architectures with information distributed across multiple nodes, e.g., Leabra cognitive architecture [23].
4. Hybrid architectures with both symbolic and subsymbolic components, with emphasis on modeling human cognition, e.g., ACT-R, since it has symbolic production and declarative memory components, but it also has subsymbolic operations in the operation of the architecture
5. Nongenerative architectures that do not produce behavior but are largely used as design tools to predict time and other requirements to accomplish some set of procedures.

Although O'Reilly and colleagues' work describes in particular the Leabra architecture, their work applies more broadly to the design of other cognitive architectures [23]. They note that cognitive architectures represent a complex way to do computational modeling, but the reason they are used is that modeling human cognition does not seem possible with a simpler generalized algorithm. O'Reilly and colleagues give a list of principles that they used to develop the Leabra architecture and are broad enough to be used for other architectures. Some of these principles are:

1. Balance the tradeoffs associated with different approaches, often using a compromised approach that may integrate multiple solutions.
2. Biology is important. The human brain is the only working truly successful cognitive system, so there is merit in trying to understand its details.
3. Occam's razor—regardless of the complexity required, it is best to use the simplest model that is sufficient for the modeling required.
4. Again, there is a need to balance tradeoffs between biological constraints, cognitive constraints, and computational constraints, depending on the model desired.
5. Experience-driven learning mechanisms are essential.
6. Microstructural mechanisms will affect the macrostructural mechanisms.
7. Changes in neural firing, i.e., activation, can occur more quickly than changes in synapses.
8. Meaning, in a connectionist system, is due to the patterns across the entire system, not individual neural messages.

An interesting cognitive architecture is the very hybridized OpenCog, which has the goal of creating a human-equivalent artificial general intelligence (AGI) [24–26]. OpenCog uses a graphical database that links to a myriad of different cognitive processes, including forward and backward chaining of essentially production rules, Bayesian inference (i.e., update probabilities as more information becomes available), an economic attention allocation mechanism where there are numerous attention values attached to various pieces of information, probabilistic logic networks (allow reasoning with uncertain information in the real world), meta-optimizing semantic evolutionary and probabilistic evolutionary mechanisms (where "evolutionary" mechanisms are not actual biological genetics but refer to the field of genetic programming, i.e., of applying fitness selection to large numbers of, for example, random-like programs and selecting and evolving the more fit programs), a natural language input and output processing system, and the use of emotions. While at this time of writing OpenCog has obviously not even come close to its goal of a human equivalent AGI, and while OpenCog contains a potpourri of cognitive concepts, many of these do stem originally from animal and human minds and show the potential of applying principles from the medical sciences *back* to the field of artificial intelligence.

4 Concepts from the Medical Sciences Being Applied *Back* to the Field of Artificial Intelligence: The Causal Cognitive Architecture

As noted above, a cognitive architecture incorporates results from the medical sciences, including biology and psychology, into a system that models a mind, whether natural or artificial. This model is formal enough that it can be implemented as a computer program. Like other cognitive architectures, what is termed the "causal cognitive architecture" arises from observations in the medical sciences [27–32]. A simplified block diagram of an implementation of a causal cognitive architecture, the Causal Cognitive Architecture 1 (CCA1), is shown in Fig. 2 [32]. Some of the many observations and concepts from the medical sciences, including biology and psychology, that the causal cognitive architecture incorporates will be discussed first. Then its basic properties as well as its emergent properties will be considered.

In most of the animal world, in both invertebrates and vertebrates, there is the ability to move and the ability to navigate. In mammals, Hafting and colleagues [33] showed neural maps of the spatial environment in the dorsocaudal medial entorhinal cortex, with location activation of "grid cells" in these maps. Schafer and Schiller note that it is possible that these maps intended originally for navigation in the physical world can be used to navigate concepts [34]. Evolutionary precursors to the mammalian cortex go back to the earliest vertebrates [35]. Thus, in the design of the causal cognitive architecture, it was decided that information would be stored as navigation maps. Operations on these maps could be effected in a "navigation module" where maps would be modified, updated, stored adjacently, retrieved, linked to other maps, and so on. Except for some typical neural network-like hierarchical processing of input signals and output signals, all information would be stored and manipulated in map-like data structures, and much of the higher processed information would pass through the navigation module [32].

The mammalian cortex appears to be made up of large numbers of repeating cortical minicolumns [36]. Also, there is a generative aspect in how the mind processes information. As a result, a basic pattern recognizing circuit, a Hopfield-like network (HLN), was chosen for many of the causal cognitive architecture's circuits. HLNs can be dynamically reconfigured with other HLNs to extract what is described in the architecture as maximal "meaningfulness" from input vectors, where meaningfulness is the reciprocal of the Shannon entropy, favoring activation of the maximal number of HLNs further downstream [27, 31, 32].

Just as a child is not a tabula rasa, neither is the causal cognitive architecture. The instinctive primitives module shown in the architecture in Fig. 2 contains procedural vectors, which can be triggered by processed sensory inputs, by the navigation module, or by other modules of the architecture. If triggered, an instinctive primitive will be applied to the navigation map currently in the navigation module and can, for example, move along the map to trigger another instinctive or learned primitive or change the map or cause another linked map to be retrieved and so on. The instinctive primitives include physics primitives, mathematical primitives, psychology primitives, and planning primitives. Also, there are learned primitives that as the name implies are learned from experience. There is a letter "D" in many of the boxes in the architecture in Fig. 2. "D" stands for the internal developmental timer; i.e., as the

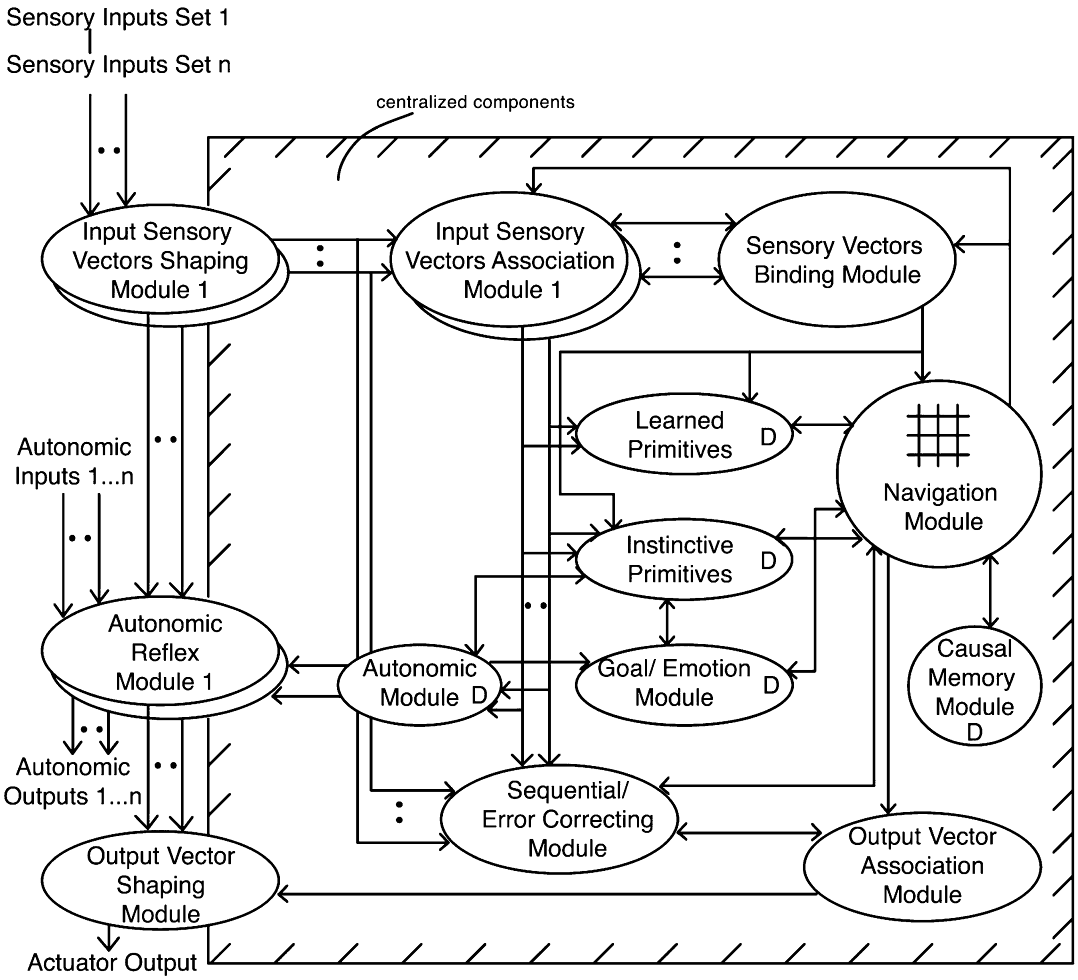


Fig. 2 Causal Cognitive Architecture 1 (CCA1) Not all connections are shown. D – internal developmental timer. (License – Original work for this publication)

CCA1 shown in Fig. 2 gains more experience, different levels of instinctual primitives or operations in these modules will be executed.

Examining Fig. 2, note that the sensory inputs are processed by hierarchies of input sensory vector association modules and then a binding module. Inputs are then sent to the navigation module. (They also go to the learned and instinctive primitives where they can trigger primitives to be executed, to the sequential/error correcting module which is useful for sequential data, to the autonomic modules, and to other modules.) Applying primitives, i.e., vectors that cause a modification of the current map in the navigation module, allows the input signal to be modified in a precausal fashion. The map already has some causality built into it; i.e., this goes to that, which goes to that and so on. The output of the navigation module is processed, in its most basic form, via the output vector association and shaping modules, into a movement of an actuator. As such, precausal behavior emerges from the Causal Cognitive Architecture 1 shown in Fig. 2 [32].

Deep learning can recognize patterns and use reinforcement learning to operate at human-like levels for certain skills such playing games [37, 38]. However, if deep learning needs to causally make sense of a problem, especially if there are not large numbers of training examples, its performance is poor compared to a four-year-old child [39, 40]. In the causal cognitive architecture, instinctive and learned primitives applied against map-like memories or input signals represented in map-like structures will produce outputs that the maps lead to, which as noted above results in a precasual complex associative behavior. Many birds as well as mammals can exhibit precausal behavior, i.e., not a full understanding of cause and effect but a complex association of causes with outcomes that is more than just a simple reflex or association. However, it is unclear whether *any* nonhumans can effectively understand cause and effect and can truly use full causality in making decisions.

Asian elephants have brains much larger than humans, but in experiments by Nissani [41] behavior was due to associative learning rather than causal learning or behavior. New Caledonia crows are frequently claimed to show causal behavior in solving physical problems. However, in research by Neilands and colleagues where the crows must drop a weighted object down a tube, the crows did not understand causally the idea of force, and there was little causal understanding overall [42]. Tayler and colleagues examined crows solving string pulling problems and came to the conclusion that there was little causal understanding but that the problems were solved instead by a perceptual-motor feedback loop [43]. If the behavior of primates, obviously closer in origin to humans, is considered, there still is not a clear indication that robust causal behavior occurs. For example, Visalberghi and Limongelli showed that while capuchin monkeys were able to use a stick to push a reward out of tube, if a hole and a trap are added to the tube, the monkeys showed poor understanding of cause-and-effect relations in not pushing the food into the trap [44]. This tube with a trap and food-reward-type problem is also difficult for chimpanzees. However, work by Seed [45] and colleagues showed that if the tube problem is simplified a bit, there appears to be some understanding by the chimpanzees about the functional properties of the problem. However, while the chimpanzees were able to solve this simplified problem, it still required dozens of trials [45].

While there remains controversy concerning the extent of causal behavior possible in nonhumans, note that chimpanzees are the closest extant relatives to humans and yet are not capable of a robust understanding of cause and effect or causal behavior. Thus, a design requirement of the causal cognitive architecture was that there should be a simple mechanism to explain how the architecture can transition from the precausal behavior described above to a fully causal behavior [32].

Almost all the circuits of the causal cognitive architecture have feedback pathways. Feedback from downstream hierarchical levels as well as lateral levels send back signals which will allow the CCA1 to better anticipate what the next input sensory vector should be and thus better recognize such input signals. In Fig. 2 showing the Causal Cognitive Architecture 1, if the feedback pathways from the navigation module to the input sensory various modules are enhanced, then full intermediate navigation maps from the navigation module can be stored temporarily in the input sensory modules. In the next input cycle, rather than considering actual sensory input signals, the intermediate results can be fed back to the navigation module and operated on again by the instinctive and learned primitives [32].

The ability of the causal cognitive architecture to temporarily store intermediate values of the navigation module and then process these map-like structures and values again, and again as many times as necessary, results in the ability of the causal cognitive architecture to have robust causal behavior [31, 32]. Note that the transition from precausal to full causal ability merely requires enhanced feedback pathways and a slight change in how input signals are processed, in keeping with the requirement above that the architecture be able to provide a mechanism for a simple transition from precausal to causal abilities.

In a simulation of the Causal Cognitive Architecture 1 (CCA1) controlling a search and rescue robot in a rain forest searching for a lost hiker [32], in precausal mode, if the robot comes to a fast-moving, noisy river, it will cross the river since it is able to cross rivers. As sensory inputs are processed in the navigation module, intuitive and learned primitives are activated and will not prevent the robot from entering the river. However, in this rain forest, the fast-moving, noisy river is the start of a waterfall, and unfortunately the robot is swept over the edge of the waterfall and is damaged. If the robot is repaired and goes out into the rain forest and senses another fast-moving, noisy river, then by associative mechanisms, i.e., via the learned primitives (as well as the goal/emotion module), it will avoid the river.

In the same situation in full causal mode [32], "fast moving" + "water" will cause an intuitive physics primitive to cause the navigation module to produce the navigation map "push" + "water" (where there are not actually discrete symbols but instead a map of the robot moving in water). The intermediate navigation map "push" + "water" is fed back to the sensory circuits, and in the next cycle processing input signals, the "push" + "water" navigation map loads back in the navigation module and is operated on by the intuitive and learned primitives. The result is pulling up or creating a new map with the robot underneath water. This intermediate result is fed back and then used as the input signal in the next cycle. An intuitive primitive is triggered by this map of the robot underneath the water, with an output from the navigation center to not go in this direction. As a result, in full causal mode, the search and rescue robot does not enter this fast-moving, noisy river and become damaged, even though it has never seen a waterfall or a noisy fast-moving river before or trained on such examples.

An emergent property of the causal cognitive architecture is the ability to not only handle navigation in the physical world but also use the map-like data structures in the navigation center to hold and link to more abstract concepts and to navigate this world of concepts. Causal operations can be performed on these concepts as they are modified and stored and cause retrieval of other concepts [32].

The navigation module shown in Fig. 2 stores the many navigation maps it constructs in the causal memory module. When a similar situation occurs again in the future, then similar navigation maps are triggered in the causal memory module and fed into the navigation module, thus providing a learned representation of which actions occurred in the past and which can be used for the new situation, with modifications as needed. An emergent property of this is explainability—the sequence of navigation maps used in an action or decision gives a very reasonable explanation of why a particular answer was given for a particular problem. (Note that the subsymbolic aspects of the causal cognitive architecture, such as the initial steps in processing an input signal through a hierarchy of Hopfield-like networks or equivalent, are not fully captured by the navigation maps.) Gilpin and colleagues [46] describe the property of explainability as a model being able to give the reasons for its behavior. The lack of explainability in conventional deep learning models is an important concern when using such models for critical applications, including those of AI in medicine.

Another emergent property of this architecture is the ability to generate and handle analogies [32]. For example, consider asking the almost philosophical question to the search and rescue robot in the example above after returning from its rescue mission of whether it wants to spend time with either person A or person B, who are both similar but person B is more smiley and more outgoing, i.e., very noisy. Person A and person B would be put onto a temporary map in the CCA1's navigation module (Fig. 2). The Causal Cognitive Architecture 1 has to essentially decide whether to navigate to person A or person B. Its instinctive primitives, which include psychology primitives, favor smiling people. However, person B is also noisy, and this results in the navigation center pulling up a previous navigation map—the river was noisy, and thus the link on the map with person B now ends up pointing to a dangerous situation. Intermediate results would be fed back to the sensory input stages and processed again, temporary maps would be switched back again, and the noisy person B is now associated with possible danger. Thus, there is a navigation output to navigate to person A. Note that without any elaborate, specialized algorithms and without any special central controlling stored program, other than the inherent and relatively simple operational cycles of the architecture, the CCA1 has made what would seem like a cognitively advanced decision. Analogies readily emerge from the causal cognitive architecture.

The causal cognitive architecture was not intended to model disease, and indeed no pathological predispositions were intentionally designed into its architecture. However, an interesting emergent property of the architecture is that when the mode switches from precausal to fully causal (i.e., send back intermediate results from the navigation module to the sensory stages so they can be operated on again in the next sensory input cycle), psychotic behavior may emerge.

Any number of varied imperfections or combinations of them can result in psychotic behavior (but in precausal mode usually would not have had catastrophic effects as such) [29, 31, 32]. The intermediate navigation maps that the navigation module stores temporarily in the input sensory vector circuits, under a faulty operation, can be interpreted as a real input sensory signal, i.e., effectively hallucination-like and then delusional-like responses when instinctive and learned primitives are applied to it in subsequent operations, and then cognitive dysfunction of the system. Hallucinations, delusions, and cognitive dysfunction are typical of psychosis. Note that it is not just one defect that can cause this, but many different defects in various circuits.

Since the causal cognitive architecture is inspired by biology (although *not* pathology), then if it predicts that psychosis easily emerges, why is it that only approximately 1% of the human population has schizophrenia and not a larger percentage of the population? Actually, as van Os and colleagues show in their research [47], more than 10% of the population (a large figure from a population point of view) will experience some sort of psychotic-like symptoms—there are many causes why humans may experience psychotic-like symptoms, as the causal cognitive architecture predicts. Work by Anttila and colleagues [48] looking at the genomes of 265,218 psychiatry patients and 784,643 controls found considerable genetic overlap between what should be very different formal psychiatric disorders, including schizophrenia, again in keeping with the causal cognitive architecture predicting that psychosis does not emerge from a single or small group of genes.

In the causal cognitive architecture, in the noncausal mode, i.e., in the precasual mode, the psychotic-like behavior does not happen as easily. Note that humans readily develop psychosis (i.e., the 10% figure above) but humans do have robust causal abilities. While there is still controversy concerning the extent of causal behavior possible in nonhumans, from all the evidence available, as discussed above, nonhumans do not have robust causal behavior. Should psychosis readily arise in nonhumans? The causal cognitive architecture would predict no. In fact, in just about all other mammals, psychosis is rare, and in psychopharmacological research settings, large efforts are needed to induce at best unreliable models of schizophrenia in research animals [49].

Causal aspects of the causal cognitive architecture are inspired by the abilities of human working memory. The causal cognitive architecture would predict that in asymptomatic relatives of patients with psychosis, there are also defects in the causal process and the working memory, even though perhaps in that person the cumulative effects of the defects do not result in psychosis. In fact, decreased working memory abilities are found not only in patients with schizophrenia but also in unaffected relatives [50].

It is interesting to look at the transition from nonhumans, where the closest extant relatives of humans, chimpanzees, do not have full causal abilities nor readily develop psychosis, and it would be assumed that the last common ancestors of both modern humans and chimpanzees did not either. The "schizophrenia paradox" is that schizophrenia reduces an individual's ability to reproduce and should be naturally eliminated from the population yet continues to be found throughout the world at a relatively high prevalence approaching 1%. There are all sorts of rationalizations for this [51–54]. However, the causal cognitive architecture rejects the schizophrenia paradox and these explanations. Instead, it predicts that the transition to a brain architecture that allows causality is also vulnerable to many different defects allowing psychosis to emerge. Thus, there should be many different genetic alleles that prior to the transition to full causality were not harmful, that now would allow psychosis to emerge more easily, and that removing a myriad of different genetic characteristics from a population will not occur as easily as a removing a single faulty allele [29, 31, 32].

Work by Liu and colleagues [55] published in 2019 compared single nucleotide polymorphisms (SNPs) associated with schizophrenia in modern humans with those present in the recovered genomes of extinct archaic Denisovan and Neanderthal humans (thought to have split from the modern human ancestors approximately 440,000 to 270,000 years ago [56]). Liu and colleagues showed that the risk alleles for schizophrenia appear to have been gradually removed from the modern human genome due to the negative selection pressure, as the causal cognitive architecture predicted.

5 Discussion

As artificial intelligence advancements and technologies emerge, they are often incorporated into the field of medicine. However, it is noted that the reverse of this process is an important one also. Many of the large improvements and breakthroughs in the field of artificial intelligence are in fact due to knowledge and inspiration *from* the medical sciences, including biology and psychology. A number of examples were discussed above.

The basis of neural networks, the artificial neuron, was one of the first achievements in the field of artificial intelligence. It was derived from knowledge in the biological and medical fields. Similarly, scientific knowledge about the visual cortex inspired the creation of the neocognitron. In turn, he neocognitron led to the convolutional neural network, which in turn led the start of what is sometimes called the deep learning revolution. Deep learning neural networks have at this time been applied to almost every field of academia, industry, and, of course, medicine.

Cognitive architectures intentionally attempt to take knowledge from the medical sciences, including biology and psychology, and incorporate this knowledge into a formal structure, which then can be implemented as a computer program. A variety of cognitive architectures were discussed above. While many cognitive architectures make only small contributions to the field of AI but may be more useful to provide insight into the operation of the human mind, some are intended as bona fide contributions to artificial intelligence. The OpenCog cognitive architecture, for example, has the goal of becoming a human equivalent artificial general intelligence, although admittedly at the time of writing, it is far from this achievement.

As was noted above, deep learning can recognize patterns and use reinforcement learning to play games at a human-like level [37, 38]. However, if a deep-learning-based system needs causal understanding to solve a problem, especially if there are not large numbers of training examples, its performance is poor compared to a 4-year-old child [39, 40]. For many situations, it simply is not possible or practical to have extensive training on every type of example that could arise. This certainly happens in the medical applications of AI as well as in almost all other fields as well. Work by Amodei and Hernandez [57] note that the famous Moore's law has a 2-year doubling period (i.e., the number of transistors on integrated circuits fabricated by manufacturers doubles approximately every 2 years), which means from 2012 to 2018 there should have been an approximately eightfold improvement in the power of computer hardware. A 800% increase is, of course, significant and impressive. However, what really happened in the training of deep learning networks from 2012 to 2018 was that a 300,000-fold increase (i.e., a 30 million % increase) in hardware computational power was used to train deep learning networks. This number has only increased in the years since 2018. These exponential increases in computing power are not sustainable, nor is the ever-increasing need for more training examples. The resources and costs simply become prohibitive, despite the improvements by hardware manufacturers and despite the improvements in deep learning training algorithms. Thus, deep learning neural networks may start approaching pragmatic limits in their ability to do more complex tasks at a human level, certainly those requiring causal decisions and limited training examples.

Others in the field of AI have started to recognize the problems that current implementations of deep learning face and the need to consider alternative approaches. Discussed above, for example, was the interesting cognitive architecture OpenCog, which has the goal of creating a human-equivalent artificial general intelligence [24–26]. Graves and colleagues [58] discuss using a neural network that can read and write to an external memory, i.e., a hybrid system. Huyck [59] describes work on a neuromorphic-like cognitive architecture with a fast implicit subconscious system and a slow explicit conscious system. Epstein [60] discusses cognitive modeling of spatial navigation. Lake and colleagues [61] discuss building causal models of the world and discuss intuitive physics and psychology present in infants. Hawkins and colleagues [62] and others, such as Schafer and Schiller [34], discuss how abstract concepts can be represented in a spatial framework. Laird and Mohan [63] discuss using architectures that contain more innate learning mechanisms. Taatgen [64] discusses using multiple levels of abstraction for learning. In a recent paper discussing deep learning for higher-level cognition [65], Goyal and Bengio write: "Have the main principles required for deep learning to achieve human-level performance been discovered, with the main remaining obstacle being to scale up? …. We argue that having larger and more diverse datasets is important but insufficient without good architectural inductive biases."

While the causal cognitive architecture discussed above is largely at an experimental level, it is a technology that is essentially derived from the medical sciences, including biology and psychology [32]. As noted above, the causal cognitive architecture creates navigation maps from input sensory vectors and produces a navigation output, i.e., related to movement. A precausal behavior can result from this architecture. There is no understanding of cause and effect, but the maps navigate from problem to solution that may seem to be related to this or that cause. However, intermediate solutions to a problem, i.e., the maps created in the navigation module, can be fed back along existing but enlarged feedback pathways and stored as an input sensory value, and in the next input cycle processed again, and repeatedly so as needed. Maps of intermediate parts of the problem solution can be created and stored, new maps made, previous maps retrieved, and so on. In this fashion, true causal behavior can result with an understanding of cause and effect.

The causal cognitive architecture, inspired and derived by the medical sciences, including biology and psychology, offers features, albeit in toy models tested to date [27–32], useful for the field of AI. It allows problems to be mapped and processed causally. It stores the intermediate maps it creates in solving a problem or performing a behavior—retrieving and replaying the maps allow explainability about a decision the architecture made. While solving physical navigation problems is obvious, the causal cognitive architecture can similarly navigate a world of concepts. Another emergent property of the architecture is the ability to handle analogies automatically.

As shown above, many of the large improvements and breakthroughs in the field of artificial intelligence are in fact due to knowledge and inspiration *from* the medical sciences. An appreciation for the origin of the technologies in AI can help with their application in general as well as to the medical fields.

6 Cross-References

- ▶ AIM and Medical Innovation
- ▶ AIM and the Emergence of Deep Medicine
- ▶ AIM in Schizophrenia
- ▶ Basic Concepts of AI
- ▶ Basic Concepts of AIM
- ▶ Introduction to AIM and "Lessons for all doctors"

References

1. Buchanan BG, Sutherland GL, Feigenbaum EA. Heuristic DENDRAL: a program for generating explanatory hypotheses in organic chemistry. In: Meltzer B, Michie D, Swann M, editors. Machine Intelligence 4 – Proceedings of the Fourth Annual

Machine Intelligence Workshop. Edinburgh: Edinburgh University Press; 1969.

2. Buchanan BG, Shortliffe EH. Rule based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project. Reading: Addison-Wesley; 1984.
3. Perry CA. Knowledge bases in medicine: a review. Bull Med Libr Assoc. 1990;78(3):271–82. PMID: 2203499
4. Russell S, Norvig P. The history of artificial intelligence. In: Artificial intelligence: a modern approach. 4th ed. Hoboken: Pearson; 2021. p. 17–27.
5. Krizhevsky, A, Sutskever, I, Hinton, GE. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 1 (NIPS'12). Curran Associates: Red Hook, NY, USA, 1097–1105; 2012.
6. Veronese E, Castellani U, Peruzzo D, Bellani M, Brambilla P. Machine learning approaches: from theory to application in schizophrenia. Comput Math Methods Med. 2013;2013:867924. https://doi.org/10.1155/2013/867924.
7. Kim J, Calhoun VD, Shim E, Lee JH. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. NeuroImage. 2016;124(Pt A):127–46. https://doi.org/10.1016/j.neuroimage.2015.05.018.
8. Sciencedirect.com/journal/artificial-intelligence-in-medicine. About the journal – aims and scope. 2020. [cited 2020 Dec 8]. Available from: https://www.sciencedirect.com/journal/artificial-intelligence-in-medicine
9. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys. 1943;5:117–37.
10. Hebb DO. The organization of behavior. New York: Wiley; 1949.
11. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65(6):386–408. https://doi.org/10.1037/h0042519.
12. Minsky ML, Papert SA. Perceptrons. Cambridge, MA: MIT Press; 1969.
13. Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. J Physiol Oct. 1959;148(3):574–91. https://doi.org/10.1113/jphysiol.1959.sp006308.
14. Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980;36:193–202. https://doi.org/10.1007/BF00344251.
15. Anderson JR. The architecture of cognition. Cambridge, MA: Harvard University Press; 1983.
16. Ritter FE, Tehranchi F, Oury JD. ACT-R: a cognitive architecture for modeling cognition. Wiley Interdiscip Rev Cogn Sci. 2019;10(3):e1488. https://doi.org/10.1002/wcs.1488.
17. Trafton JG, Hiatt LM, Harrison AM, Tamborello FP, Khemlani SS, Schultz AC. ACT-R/E: an embodied cognitive architecture for human-robot interaction. J Hum Robot Interact. 2013;2:30–55. https://doi.org/10.5898/JHRI.2.1.Trafton.
18. Schoppek W, Boehm-Davis DA. Opportunities and challenges of modelling user behavior in complex real world tasks. MMI Interaktiv. 2004;7:47–60.
19. Samsonovich A. Toward a unified catalog of implemented cognitive architectures. In: Biologically inspired cognitive architectures; 2010. p. 195–244. https://doi.org/10.3233/978-1-60750-660-7-195.
20. Kotseruba, I, Gonzalez, O, Tsotsos, JK. A review of 40 years of cognitive architecture research. arXiv: 1610.08062v3 [cs.ai]; 2016.
21. Busetta P, Howden N, Rönnquist R, Hodgson A. Structuring BDI agents in functional clusters. In: Jennings NR, Lespérance Y, editors. Intelligent agents VI. Agent theories, architectures, and languages. ATAL 1999. Lecture notes in computer science, Vol. 1757. Berlin: Springer; 2000. https://doi.org/10.1007/10719619_21.
22. Laird JE. The soar cognitive architecture. Cambridge, MA: MIT Press; 2012.
23. O'Reilly RC, Hazy TE, Herd SA. The Leabra cognitive architecture: how to play 20 principles with nature and win! In: Chipman SEF, editor. The Oxford handbook of cognitive science. New York: Oxford University Press; 2017. p. 91–115.
24. Hart D, Goertzel B. OpenCog: a software framework for integrative artificial general intelligence. In: Wang, et al., editors. Proceedings of the first AGI conference. Amsterdam, Netherlands: IOS Press; 2008. p. 468–72.
25. Goertzel, B, Duong, D. OpenCog NS: a deeply-interactive hybrid neural-symbolic cognitive architecture designed for global/local memory synergy. AAAI Fall Symposium: Biologically Inspired Cognitive Architectures; 2009.
26. Goertzel B, Pennachin C, Geisweiller N. Engineering general intelligence, Part 2: the CogPrime architecture for integrative, embodied AGI. Paris: Atlantis Press; 2014.
27. Schneider H. Meaningful-based cognitive architecture. In: Samsonovich AV, editor. Biologically inspired cognitive architectures BICA 2018. Procedia computer science, vol. 145; 2018. p. 471–80. https://doi.org/10.1016/j.procs.2018.11.109.
28. Schneider H. Subsymbolic versus symbolic data flow in the meaningful-based cognitive architecture. In: Samsonovich AV, editor. Biologically inspired cognitive architectures BICA 2019, Advances in intelligent systems and computing, vol. 948; 2020. p. 465–74. https://doi.org/10.1007/978-3-030-25719-4_61.
29. Schneider H. Schizophrenia and the future of artificial intelligence. In: Samsonovich AV, editor. Biologically inspired cognitive architectures 2019, Advances in intelligent systems and computing, vol. 948; 2020.

p. 475–84. https://doi.org/10.1007/978-3-030-25719-4_62.

30. Schneider H. Emergence of belief systems and the future of artificial intelligence. In: Samsonovich AV, editor. Biologically inspired cognitive architectures BICA 2019, Advances in intelligent systems and computing, vol. 948; 2020. p. 485–94. https://doi.org/10.1007/978-3-030-25719-4_63.
31. Schneider H. The meaningful-based cognitive architecture model of schizophrenia. Cogn Syst Res. 2020;5(9):73–90. https://doi.org/10.1016/j.cogsys.2019.09.019.
32. Schneider H. Causal cognitive architecture 1: integration of connectionist elements into a navigation-based framework. Cogn Syst Res. 2021;66:67–81. https://doi.org/10.1016/j.cogsys.2020.10.021.
33. Hafting T, Fyhn M, Molden S, Moser MB, Moser EI. Microstructure of a spatial map in the entorhinal cortex. Nature. 2005;436(7052):801–6. https://doi.org/10.1038/nature03721.
34. Schafer M, Schiller D. Navigating social space. Neuron. 2018;100(2):476–89. https://doi.org/10.1016/j.neuron.2018.10.006.
35. Suryanarayana SM, Robertson B, Wallén P, et al. The lamprey pallium provides a blueprint of the mammalian layered cortex. Curr Biol. 2017;27(21):3264–77. https://doi.org/10.1016/j.cub.2017.09.034.
36. Buxhoeveden DP, Casanova MF. The minicolumn hypothesis in neuroscience. Brain. 2002;125 (Pt 5):935–51. https://doi.org/10.1093/brain/awf110.
37. Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, MA: MIT Press; 2016.
38. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature. 2015;518:529–33. https://doi.org/10.1038/nature14236.
39. Waismeyer A, Meltzoff AN, Gopnik A. Causal learning from probabilistic events in 24-month-olds: an action measure. Dev Sci. 2015;18(1):175–82. https://doi.org/10.1111/desc.12208.
40. Ullman S. Using neuroscience to develop artificial intelligence. Science. 2019;363(6428):692–3. https://doi.org/10.1126/science.aau6595.
41. Nissani M. Do Asian elephants (*Elephas maximus*) apply causal reasoning to tool-use tasks? J Exp Psychol Anim Behav Process. 2006;32(1):91–6. https://doi.org/10.1037/0097-7403.32.1.91.
42. Neilands PD, Jelbert SA, Breen AJ, Schiestl M, Taylor AH. How insightful is 'insight'? New Caledonian Crows do not attend to object weight during spontaneous stone dropping. PLoS One. 2016;11(12):e0167419. https://doi.org/10.1371/journal.pone.0167419.
43. Taylor AH, Knaebe B, Gray RD. An end to insight? New Caledonian crows can spontaneously solve problems without planning their actions. Proc Biol Sci. 2012;279(1749):4977–81. https://doi.org/10.1098/rspb.2012.1998.
44. Visalberghi E, Limongelli L. Lack of comprehension of cause-effect relations in tool-using capuchin monkeys (*Cebus apella*). J Comp Psychol. 1994;108(1):15–22. https://doi.org/10.1037/0735-7036.108.1.15.
45. Seed AM, Call J, Emery NJ, Clayton NS. Chimpanzees solve the trap problem when the confound of tool-use is removed. J Exp Psychol Anim Behav Process. 2009;35(1):23–34. https://doi.org/10.1037/a0012925.
46. Gilpin, LH, Bau, D, Yuan, BZ, Bajwa, A, Specter, M, Kagal L. Explaining explanations: an overview of interpretability of machine learning. 2018 IEEE 5th international conference on data science and advanced analytics (DSAA), Turin, Italy, 2018;80–89. https://doi.org/10.1109/DSAA.2018.00018.
47. van Os J, Hanssen M, Bijil RV, et al. Prevalence of psychotic disorder and community level psychotic symptoms: an urban-rural comparison. Arch Gen Psychiatry. 2001;58(7):663–8.
48. Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, Duncan L, Escott-Price V, et al. Analysis of shared heritability in common disorders of the brain. Science. 2018;360(6395):eaap8757. https://doi.org/10.1126/science.aap8757.
49. Jones CA, Watson DJ, Fone KC. Animal models of schizophrenia. Br J Pharmacol. 2011;164(4):1162–94. https://doi.org/10.1111/j.1476-5381.2011.01386.x.
50. Zhang R, Picchioni M, Allen P, Toulopoulou T. Working memory in unaffected relatives of patients with schizophrenia: a meta-analysis of functional magnetic resonance imaging studies. Schizophr Bull. 2016;42(4):1068–77. https://doi.org/10.1093/schbul/sbv221.
51. Pearlson GD, Folley BS. Schizophrenia, psychiatric genetics, and Darwinian psychiatry: an evolutionary framework. Schizophr Bull. 2008;34(4):722–33. https://doi.org/10.1093/schbul/sbm130.
52. Benítez-Burraco A, Di Pietro L, Barba M, Lattanzi W. Schizophrenia and human self-domestication: an evolutionary linguistics approach. Brain Behav Evol. 2017;89(3):162–84. https://doi.org/10.1159/000468506.
53. Polimeni J, Reiss JP. Evolutionary perspectives on schizophrenia. Can J Psychiatr. 2003;48(1):34–9. https://doi.org/10.1177/070674370304800107.
54. Crow TJ. Schizophrenia as the price that *homo sapiens* pays for language: a resolution of the central paradox in the origin of the species. Brain Res Brain Res Rev. 2000;31(2–3):118–29. https://doi.org/10.1016/s0165-0173(99)00029-6.
55. Liu C, Everall I, Pantelis C, Bousman C. Interrogating the evolutionary paradox of schizophrenia: a novel framework and evidence supporting recent negative selection of schizophrenia risk alleles. Front Genet. 2019;10:389. https://doi.org/10.3389/fgene.2019.00389.
56. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, Viola B, Briggs AW, Stenzel U, Johnson PL, Maricic T, Good JM, Marques-Bonet T, Alkan C, Fu Q, Mallick S, Li H, Meyer M, Eichler EE, Stoneking M, Richards M, Talamo S, Shunkov MV, Derevianko AP, Hublin JJ, Kelso J, Slatkin M, Pääbo

S. Genetic history of an archaic hominin group from Denisova cave in Siberia. Nature. 2010;468(7327): 1053–60. https://doi.org/10.1038/nature09710.

57. Amodei, D, Hernandez, D. AI and Compute. OpenAI Blog; 2018. Retrieved Dec 10, 2020 from: https://openai.com/blog/ai-and-compute/
58. Graves A, Wayne G, Reynolds M, Harley T, Danihelka I, Grabska-Barwińska A, Colmenarejo SG, Grefenstette E, Ramalho T, Agapiou J, Badia AP, Hermann KM, Zwols Y, Ostrovski G, Cain A, King H, Summerfield C, Blunsom P, Kavukcuoglu K, Hassabis D. Hybrid computing using a neural network with dynamic external memory. Nature. 2016;538(7626):471–6. https://doi.org/10.1038/nature20101.
59. Huyck, CR. The neural cognitive architecture. AAAI 2017 fall symposium: technical report FS-17-05; 2017.
60. Epstein, SL. Navigation, cognitive spatial models, and the mind. AAAI 2017 fall symposium: technical report FS-17-05; 2017.
61. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. Behav Brain Sci. 2017;40:e253. https://doi.org/10.1017/S0140525X16001837.
62. Hawkins J, Lewis M, Klukas M, Purdy S, Ahmad S. A framework for intelligence and cortical function based on grid cells in the neocortex. Front Neural Circuits. 2019;12:121. https://doi.org/10.3389/fncir.2018.00121.
63. Laird, J, Mohan, S. Learning fast and slow: levels of learning in general autonomous intelligent agents. The Thirty-Second AAAI Conference on Artificial Intelligence (*AAAI* 2018). April 2018. Accessed at (Dec 10 2020): https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17261/16424
64. Taatgen, NA. Cognitive architectures: innate or learned? AAAI 2017 fall symposium: technical report FS-17-05; 2017.
65. Goyal, A, Bengio, Y. Inductive biases for deep learning of higher-level cognition. arXiv preprint arXiv:2011.15091; Dec 7, 2020.