Deep Intelligence: What AI Should Learn from Nature's Imagination

Ali A. Minai[1]

Abstract

Artificial intelligence (AI) has recently seen explosive growth and remarkable successes in several application areas. However, it is becoming clear that the methods that have made this possible are subject to several limitations that might inhibit progress towards replicating the more general intelligence seen in humans and other animals. In contrast to current AI methods that focus on specific tasks and rely on large amounts of offline data and extensive, slow, and mostly supervised learning, this *natural intelligence* is quick, versatile, agile, and open-ended. This position paper brings together ideas from neuroscience, evolutionary and developmental biology, and complex systems to analyze why such natural intelligence is possible in animals and suggests that AI should exploit the same strategies to move in a different direction. In particular, it argues that integrated embodiment, modularity, synergy, developmental learning, and evolution are key enablers of natural intelligence and should be at the core of AI systems as well. The analysis in the paper leads to the description of a biologically grounded *deep intelligence* (DI) framework for understanding natural intelligence and developing a new approach to building more versatile, autonomous, and integrated AI. The paper concludes that the dominant paradigm of AI today is unlikely to lead to truly natural general intelligence and that something like the biologically inspired DI framework is needed for that.

Introduction

AI and machine learning have recently made great strides—thanks especially to the explosive developments in neural networks [1–9]. But several critics, including some of the leading researchers in AI, have noted the fact that the currently dominant AI paradigm—often referred to as *deep learning*—diverges from natural intelligence in fundamental ways [10–16]. Concurring with these critiques, this position paper argues that these divergences will keep current ML systems from achieving the flexible and versatile general intelligence of the kind seen even in fairly simple animals with central nervous systems, though they will become far better than humans at solving a wide range of very specific, extremely complex problems that are traditionally associated with human intelligence. The paper argues further that building natural intelligence will require a way of thinking about AI that is rooted more firmly in biology and complex systems rather than computational engineering [15]. The classical engineering paradigm based on the concepts of design, optimization, verification, validation, stability, predictability, controllability, and goal-directedness does not work well for complex adaptive systems and must be replaced by a different framework that emphasizes self-organization, autonomy, functional adequacy, versatility, flexibility, adaptivity, growth, and resilience [17]. In doing so, the focus should be on identifying and deploying the "enabling technologies" of biology that have produced functional intelligent systems of great complexity and ultimately develop an engineering framework based on these.

One part of the question motivating this special issue of Cognitive Computation is, "What can AI learn from neuroscience?" The main message of this position paper is that AI should learn not only from neuroscience, but also from evolutionary and developmental biology, and from the emergent behavior of animals as self-organizing complex dynamical systems.

In the 1980s, the re-emergence of neural networks was seen as liberating AI from this abstract symbolic approach to a more biologically inspired, distributed, and adaptive one. However, it is now clear that not all the problems were solved, and some new ones were created. The neural

✉ Ali A. Minai
ali.minai@uc.edu

1 Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati, OH 45221-0030, USA

networks of today may nominally use biologically inspired neurons and synapses, but in fact, the biologically implausible, hand-crafted, serial, symbolic algorithms of the past have simply been replaced by parallel, distributed algorithms trained on data while retaining the dualistic notion that the *function* of intelligence can be abstracted away completely from its natural biological substrate. This position paper argues for an alternative set of principles to first understand and then, perhaps, to build a more natural AI that is grounded firmly in biology.

The proposed approach is termed *deep intelligence* (DI) because it sees intelligence as requiring *depth* along several dimensions: Structural, functional, and adaptive. But, more fundamentally, it asserts that flexible, fluent, and versatile general intelligence is a *biological* phenomenon that is a property of the organism *as a whole* and *emerges naturally* from the structure and dynamics of the organism as it interacts with its environment. The organisms—in this case, animals with nervous systems—are seen as self-organizing complex systems embedded in a complex, dynamic environment.

These ideas are not very new in themselves and have been explored well within disciplines relevant to AI, including evolutionary and developmental biology [18–21], neuroscience [22–25], cognitive science [26, 27], complex systems studies [28, 29], and robotics [30–35]. Unfortunately, their application in AI has remained confined to sub-disciplines such as developmental and evolutionary robotics, while the mainstream thrust of AI in recent years has been on building ever larger deep neural networks trained primarily with supervised learning—no doubt because such networks produce more immediate benefits from an applications perspective. This tendency is also rooted in a brain-centric view of intelligence, assuming that the best way to obtain general intelligence is to model higher cognitive abilities—notably language and reasoning—and that, after sufficient scaling up, such models would lead to artificial general intelligence (AGI). With this focus on large neural networks learning statistical inference from large datasets, AI has moved steadily away from the biological nature of intelligence, and thus from the viewpoint stated above. Most of the well-known neural models of today—with the possible exception of convolutional neural networks [5] and some recent work in reinforcement learning (RL) [36–38]—have very little of the biological inspiration that originally motivated neural networks.

Natural Intelligence

What makes an agent intelligent? Much of the work on AI has been based on a human-centric and brain-centric definition of intelligence, which sees it as an attribute comprising human—or human-like—capabilities associated with the brain: Reasoning, planning, language, complex problem-solving, etc. But, in fact, intelligence is an attribute of *all* animals with central nervous systems: *It is an inherent understanding of the world that gives the animal the ability to exploit its environment opportunistically and pervasively for productive survival*. A spider weaving a web to catch prey, an octopus changing color to camouflage itself, a leopard stalking deer on the savannah, and a human shopping for groceries are all examples of intelligence at various levels. The complexity of intelligent behaviors is related directly to the complexity of the animal's body and its nervous system. Thus, intelligence is something that has evolved gradually over hundreds of millions of years, punctuated with sharp jumps in complexity corresponding to the emergence of novel "enabling technologies" such as the spinal cord, the neocortex, and bipedality. Human intelligence is only the latest stage of this long process. This type of intelligence will be termed *natural intelligence* in this paper to distinguish it from the technical definition of *general intelligence* in cognitive psychology [39, 40].

Natural intelligence has many specific characteristics, but the following stand out in particular:

Autonomy Natural intelligence is autonomous in the sense that it involves only the animal and its environment (including other organisms). Intelligent behavior emerges from the interaction between the two, driven and modulated by the *internal* motivations, drives, and emotions of the animal. At its most basic level, it is automatic and has implicit goals, but more deliberate behaviors with subjectively perceived goals emerge in more complex animals [41].

Integrity Intelligence is a property of the *whole* animal, with all its sensory, cognitive, and behavioral capacities, not a piecemeal collection of distinct functions. This means that intelligence always interacts with the world as a single entity, and, except in pathological cases, never faces the problem of integrating parts of its immediate perception, memory, and emotion post facto to make inferences across all aspects of experience. The animal always lives in a single, unified moment even though its attention may focus only on parts of it.

Fluency Natural intelligence is always "real-time"—continually receiving sensory data and generating internal and external behaviors in the context of its internal state. It learns extremely rapidly—sometimes with almost no experience—and generalizes well out of sample. There is no opportunity for massive off-line learning—only limited intervals of mental rehearsal learning and memory consolidation. Humans have, of course, developed layers of external communication (e.g., language) and memory (e.g., written knowledge) that extend this, but that is a late emergent feature of intelligence that must

have developed by leveraging more rudimentary capacities present in earlier primates and other animals.

Adaptivity The ability to adapt over time is an essential feature of all intelligent animals—even the simplest. However, this adaptation is not confined just to neural plasticity. It includes the developmental process of the animal and its real-time behavior that is always dynamically adaptive and emergent. At a deeper level, it also includes the evolutionary process that has brought the animal to its current form.

Versatility Intelligence in all animals is versatile in that a single, integrated brain-body system performs *all* the functions underlying intelligence. Importantly, these functions are always *inherently coordinated* because they have evolved and developed within an integrated system through evolutionary, developmental, and behavioral time.

Resilience The quality of an animal's intelligence depends ultimately on its ability to survive in perpetually unexpected situations, i.e., resilience. There are very specific features of the biological system that make it resilient, including modularity [19, 42], emergent coordination [28, 29], and functional diversity [43].

Evolvability Intelligence emerged in its simplest form in very simple animals and has developed over several hundred million years into a far, far more complex functional attribute as instantiated in humans. This ability to evolve and grow over orders of magnitude is not an accidental thing, but the result of specific biological principles that are grouped under the term *evolvability* [44–46]. Successful intelligence that can thrive in a changing world must itself be evolvable.

The term "intelligent system" is often used casually to refer to any system that can learn from data, but it is absurd to say that any one function such as vision or language is "intelligent" in any real sense. The proposed DI framework attempts to make the term "intelligent systems" more specific, applying it only to systems that possess the seven features listed above to some degree. The question is: How many of these features are present in the widely celebrated AI systems of today? The answer is that almost none are, except adaptivity, and that too in a superficial, narrowly defined way. The next question is: Is it possible to get to a system with these features using today's dominant AI approaches? The case made in this position paper is that it will not be possible (or will be extremely difficult), and a different approach—Deep Intelligence—is needed, first to understand and then to replicate natural intelligence.

Critique of Current AI Methods

Deep Learning and Natural Intelligence

This, in many ways, is the golden age of AI—a time when it is, at last, achieving real-world successes at an impressive pace that is likely to continue for the foreseeable future. But today's successful deep learning-based machine learning—referred to henceforth as DL/ML—differs from natural intelligence in several fundamental ways:

- Most DL/ML systems are specialists that perform only a single task (or a well-defined range of tasks) that they are explicitly trained for, whereas natural systems show versatile intelligence across a broad range of modalities and tasks.
- DL/ML systems typically require a large amount of data and a large number of learning iterations, whereas natural systems can often learn to solve real-world problems quickly from very limited data [12, 14, 37].
- Most of the successful, application-scale DL/ML systems use supervised learning, whereas natural systems depend much more on unsupervised and fast reinforcement learning.
- In most DL/ML methods, data needs to be stored offline so it can be iterated through repeatedly whereas natural systems learn mainly from real-time data with only limited off-line storage and recall (rehearsal).
- While DL/ML systems can be extremely good at generalizing *within* sample (after sufficient training), they remain poor in out-of-sample generalization, which natural intelligent systems do as a matter of course [47].
- While DL/ ML systems are extremely good at pattern recognition and statistical inference, they are severely limited in terms of the symbolic processing and compositionality that would be required for human capacities such as language, domain-independent reasoning, and complex planning [12].
- DL/ML systems have difficulty with causal inference because of their inability to handle complex temporal compositionality [10].
- DL/ML systems have a fixed learning capacity that causes them to converge to a good solution on a complex task and stop or just maintain performance with some ongoing learning. In contrast, natural intelligent systems start with simple tasks and become *more* capable over their lifetime by building on the scaffolding of previous learning. Thus, learning depletes the learning capacity of DL/ML systems but enhances that of natural ones.
- DL/ML systems do not have any notion of meaning beyond inferred statistical regularities present in their training datasets whereas natural intelligent systems

ground meaning in the experience of their physical environment [14, 15]. As a result, even extremely sophisticated ML systems are "cognitively shallow" [48], and frequently make absurd inferences [13].

- DL/ML systems are not autonomous in the sense of being driven by internal motivations. They are trained only to serve specific purposes defined by external users and evaluated in terms of these purposes.

These differences do not matter much when AI is applied to narrow problems where a large amount of reliable data is available. Exponential increases in computational power and advances in learning algorithms are allowing DL/ML systems to show remarkable performance on tasks such as translation [49, 50], code generation [51], game playing [52, 53], image analysis [5, 54], and answering complicated questions [55], to name a few. However, it is not compatible with the goal of developing natural intelligence.

Why Is a New Framework Needed?

There are many reasons why a new framework beyond the DL/ML approach is needed, but two stand out in particular: Lack of scalability, and difficulty of integration.

Building large-scale flexible and versatile intelligent systems—such as high degree-of-freedom (DOF) autonomous intelligent robots expected to perform a wide range of functions in the real world—will require exponentially greater amounts of data, time, and computational resources with increasing system complexity, and the data- and compute-hungry DL/ML approaches will have difficulty scaling in this situation. A major reason for this is that DL/ML systems try to learn very complex tasks using neural networks that: (1) Are initially naïve, i.e., have very limited, if any, prior inductive biases; and (2) Have generic, architecturally simple forms, e.g., repeated attentional, convolutional, feed-forward layers, etc., that are expected to work across a broad range of tasks. Functional systems in animals are much more heterogeneous and function-specific. To be sure, the human brain has its own generic processor in the neocortex, but all the functions it is involved in are performed in conjunction with very specifically structured systems such as the hippocampus, basal ganglia, cerebellum, the spinal cord, etc.—and, of course, the very specific networks of sensory receptors and musculoskeletal elements. The effort to learn very complex functions with initially naïve and generic networks is what requires so much supervised learning. Animals, in contrast, do neural learning on top of a non-naïve substrate configured by evolution, and refined not only through neural learning but also via a gradual developmental process.

With regard to having the DL/ML approach possibly lead to natural intelligence that is autonomous, versatile, and

flexible, the main challenge is the narrow functional specificity of most systems. There are two obvious paths towards greater generality: (1) Start with a system that does a single thing, e.g., a large language model (LLM) generating text [55], and gradually add new task capabilities; (2) Implement specialist systems for all complex tasks important to intelligence, and then combine them into a single system. Both approaches have severe problems—above all the fact that there is no canonical list of tasks that an intelligent must perform, nor any applicable metrics: After merging N tasks, there will always be an $(N+1)$th. All that can be hoped for is ad hoc combination of modalities, which is essentially what a system like DALL-E does [56]. Second, even if all the tasks could be enumerated, each of the task-specific DL/ML systems would require huge amounts of data and computation time to train, and the real world offers neither the data nor the time. Third, the world is too complex for any amount of data and training to exhaust its possibilities and ground the system in the real world, so the system will always make absurd errors. And finally, the fact is that the best systems currently available for various tasks are very different from each other in fundamental ways, including the way they learn. Patching them onto each other or combining them will lead to conflicts and arbitrarily serious emergent problems that are always created when very complicated systems are combined. These will be apparent especially when brain-centric (rather than intrinsically embodied) AI systems trained only on large datasets are embedded into embodied real-world agents with many degrees-of-freedom.

Both these difficulties are addressed by the approach proposed below.

The Deep Intelligence View

Background

The alternative Deep Intelligence approach to AI proposed in this position paper begins with a crucial observation: *There is only one class of systems in the world that actually have natural intelligence—animals with central nervous systems*. Based on this observation, it recommends that, instead of trying to outdo Nature by devising new models of intelligence based purely on reductionistic computational thinking, AI research should build from a deeper, more comprehensive understanding of how the biological structures and processes of the animal lead to intelligence with all its capacities. While this may sound like a standard, old-fashioned statement of biological inspiration that has putatively driven neural networks for decades, it is, in fact, advocating a complete reimagining of the AI enterprise and abandoning the computational-utilitarian, "neural learning only" approach in favor of one that looks at the *entire*

biology of intelligence. This does not deny the utility of studying the system at different levels and in different parts, but emphasizes that such analysis must not lose sight of the whole at any point. Indeed, even for specific domains and models, e.g., neural networks, more biologically grounded approaches should be adopted instead of relying on abstract computational ones.

The evolution of intelligence can be seen via a *soft-core, hard-periphery model* [57]. In this view, the earliest behaving animals consisted of rather rigid sensory networks connecting to rather rigid musculoskeletal networks with minimal mediation by the intervening nervous system. Such animals were like low-order Braitenberg vehicles [58] where specific stimuli elicited fast, stereotypical responses. Since then, evolution has done three things in modular fashion:

1. Complexified the sensory networks—both by adding modalities, and by making the networks for each modality more complex—but keeping the network structure fairly rigid and stereotypical, e.g., the pattern of receptors in the retina.
2. Complexified the musculoskeletal networks by producing bodies with increasing degrees of freedom and more complex architectures, but here too, keeping the structure quite stereotypical, e.g., segmented, bilaterally symmetric bodies [20].
3. Greatly complexified the nervous system network mediating between the other two networks by making its architecture wider, deeper, and more sophisticated, thus adding an enormous number of adaptable degrees of freedom into the system.

Each of these coevolving networks has constrained the others, ensuring that the sensory and behavioral capacities of the animal remain matched with the capacity of its central nervous system. More recently—and especially in the evolution from pre-hominids to humans—the soft cognitive component has grown much more rapidly than the hard periphery, layering more levels to create increasing cognitive depth. This has allowed wholly new capacities such as natural language, symbolic reasoning, abstraction, complex causal inference, etc., to emerge—creating, in layers, a System 2 on top of the more primitive System 1 in the terminology of Kahnemann [41]. AI should understand this process in depth, and use it as a template for systematically building intelligent systems of increasing complexity.

The Significance of Depth

In the DI framework, depth does not refer just to the structural depth of deep neural networks (though that too is useful), but also functional and adaptive depth.

Structural depth follows from the fact that the brain-body system of the animal is organized into multiple levels, each instantiated by a structurally deep network. These include the musculoskeletal network of the body, and the networks of the sensory receptors, thalamus, spinal cord and brainstem, the midbrain, the limbic system, the basal ganglia, the hippocampus, the neocortex, etc. This structural depth induces *functional depth*, as each level has its own functionality, and the final behavior emerges as a result of bidirectional interaction and dynamics across all these levels. Very importantly, functional depth is accompanied by functional diversity: Each level's structure and function are distinct, not generic. This is the case with all highly optimized complex systems [59, 60].

But the most important type of depth for AI is *adaptive depth*, which comes from the fact that intelligence is an emergent product of four complex, multi-scale adaptive processes:

1. *Evolution* configures useful structures and processes in species over a very slow time-scale and encodes them into the genetic code of each species. These structures and processes represent *prior inductive biases* that are well-tuned to the environment in which the organisms of that species have to survive and reproduce [61].
2. *Development* instantiates the design specified by evolution in individual organisms, using a *staged process of interleaved growth and learning* to produce an extremely complex, well-trained intelligent agent at maturity. Each stage makes the system a little more complex and learns in the context of what prior stages have set up, thus constraining the complexity of the learning process at each stage [27].
3. *Learning* in the nervous system works in tandem with development to create detailed maps, programs, and control strategies to exploit the physical configuration produced by evolution and development extremely efficiently for survival in the animal's specific environment. Once development slows down or stops, neural plasticity becomes the primary mechanism for further learning.
4. *Emergent behavior* is the result of real-time, dynamic assembly of synergistic coordination modes in neural and musculoskeletal networks to generate ongoing external behaviors (actions) and internal perceptual and cognitive states [28, 29, 62]. This is what enables a deep complex system with relatively slow components to generate real-time responses [63, 64].

Figure 1 shows this deep intelligence adaptive process that produces adaptation in a way that is very different from current ML practice of shallow adaptation where all adaptation beyond the initial design of the naïve agent is packed into neural learning (Fig. 2).

**Fig. 1** The deep adaptation process
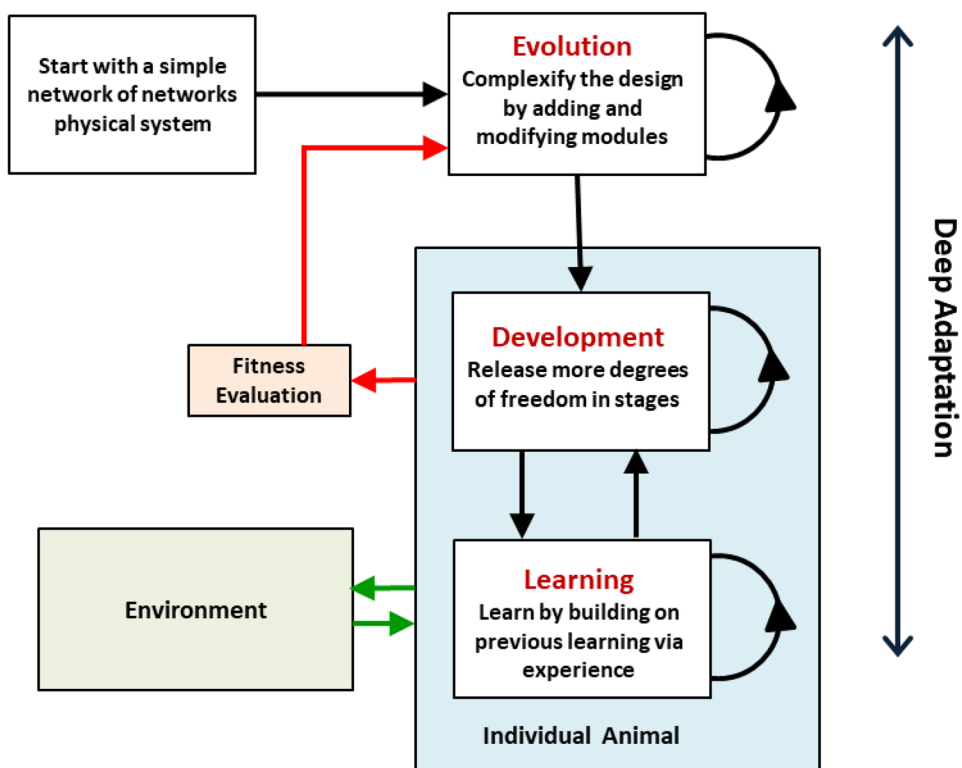
The Role of Evolution

"Nothing in biology makes sense except in the light of evolution." That famous quote by Theodosius Dobzhansky [65]—a major figure in the Modern Synthesis of evolutionary biology—is an excellent principle for thinking about the fundamentally biological phenomenon of intelligence. Looking at evolution helps answer two fundamental questions about intelligence: (1) How do animals learn rapidly in a complex world? (2) What makes animals with extremely complex intelligence possible?

The answer to the first question is that evolution configures useful priors, or inductive biases, in the embodiment of animals, which make rapid learning possible. For example, the genetically specified connectivity patterns of neurons in the visual cortex enable mammals to learn feature detectors very rapidly [66, 67]; or the connectivity of the spinal cord neural networks and muscles enables many animals to walk or swim immediately after birth. In these cases, evolution can be seen as a designer of extremely learning-ready systems with excellent inductive biases. However, the deeper question—and one of profound relevance for AI—is how such complex, well-tuned systems are possible in the first place.

The evolutionary paradigm has had a place in AI for a long time [68, 69], mainly as an optimization mechanism [30, 70–73], but the truly important thing AI can learn from evolution is the set of *strategies* it uses to generate more and more complex organisms, i.e., *evolvability* [44–46]. As engineers know all too well, building more complex systems can increase the risk of failure exponentially. Evolvability is

Fig. 2 The shallow adaptation process used in most machine learning
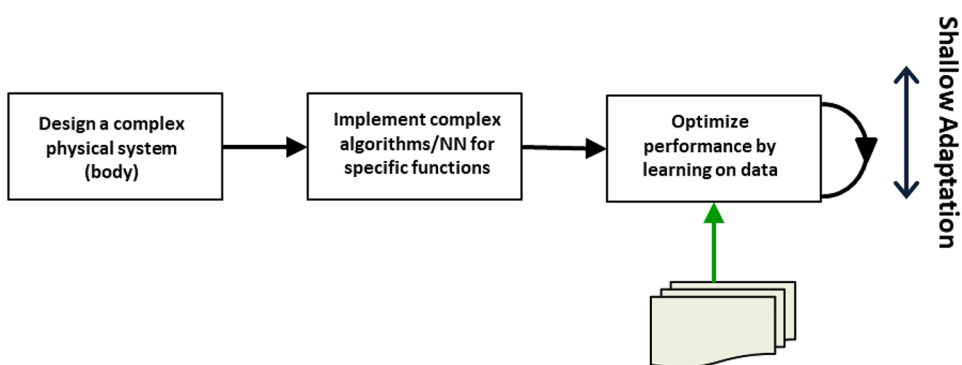
the capacity to avoid this explosion of risk. It is what makes natural intelligence possible.

The key insight that has emerged from the study of evolvability is that *modularity* and its diverse modes of deployment play a central role in evolvability [19, 42]. These modes include hierarchical modular composition [74–76], encapsulation of critical functions [77, 78], creation of neutral spaces for exploration [79, 80], re-use of modules for different functions [20, 78], and emergent coordination between modules [81, 82]. The brain too has a hierarchical modular structure [22–24]. Two important examples of hierarchical modularity enabling complex functions are the cerebral cortex in humans, with its structure of columns [83], hypercolumns [84], cell assemblies [85, 86], etc., and the hierarchical networks underlying motor control [36, 87–91]. Hierarchical modularity is also important because it means that the system is *nearly decomposable*, i.e., it minimizes cross-module dependencies, which is a key attribute of successful and evolvable complex systems [92, 93]. Evolution exploits this to produce increasingly complex viable systems by deepening modular hierarchies (Fig. 3).

The Role of Development

While evolution has the role of designing *potentially* useful structures and processes available in all multicellular organisms of any given species, development ensures that each *individual* organism realizes that potential effectively and efficiently. The importance of a developmental approach to learning has been asserted before in the framework of autonomous mental development, with a focus on autonomy and internal motivation rather than external tuning [24, 31, 34].

From an AI engineering viewpoint, development enables two crucial abilities in the adaptive system:

1. Development releases the eventually available degrees of freedom in the system *gradually* in stages, allowing each newly released set to settle into coordination patterns with those released in previous stages, and only then making new degrees of freedom available. It also ensures that the behavioral degrees of freedom are released in conjunction with increased perceptual and cognitive capacity, so that the complexity of the behaviors the organism is learning at any stage is matched with the complexity of the environment in which it perceives itself as operating. This turns what would have been an extremely complex learning problem of coordinating the full set of degrees of freedom all at once into a sequence of simpler, more constrained problems that are much likelier to converge to good solutions without requiring a lot of data and training.
2. Development enables the construction of increasingly complex functionality by hierarchical functional modularization. In humans, for example, this is apparent in linguistic learning, as simple words and sentences become building blocks for more complex ones in multiple stages. The same is true in motor learning, where simpler actions can serve as functional modules for the construction of more complex actions [94, 95]. The lack of this developmental process is a major reason why DL/ML systems are not lifelong learners.

Fig. 3 A conceptual view of how evolution builds more complex systems by deepening modular hierarchies with composition and variation
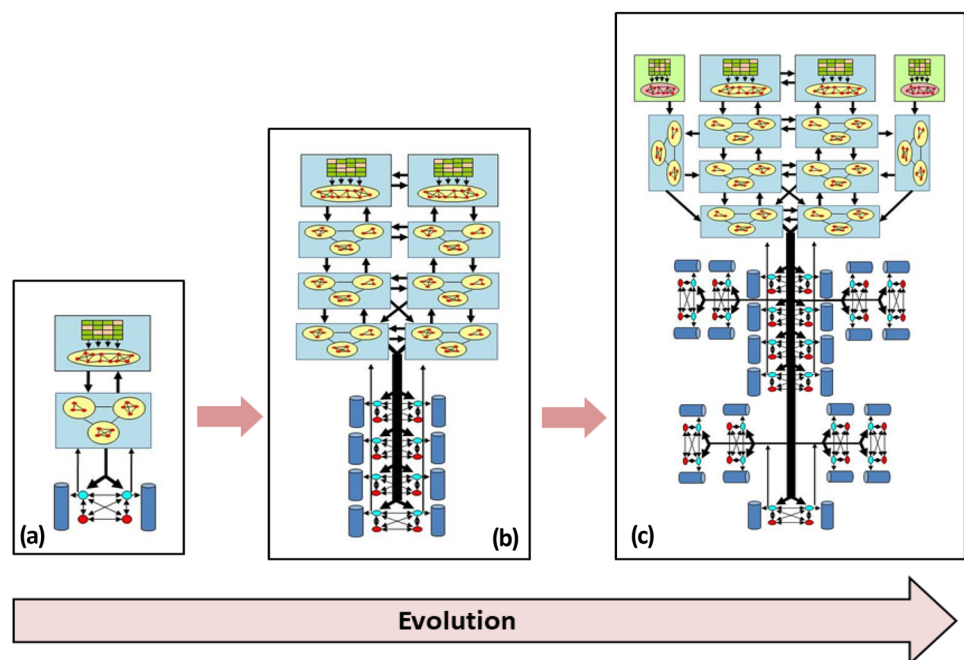
Using a developmental approach to learning in neural networks was proposed by Elman in a seminal paper [96], where he noted: "Maturational changes may provide the enabling conditions which allow learning to be most effective. …. these models work best (and in some cases, only work at all) when they are forced to "start small" and to undergo a developmental change which resembles the increase in working memory which also occurs over time in children. This effect occurs because the learning mechanism in such systems has specific shortcomings which are neatly compensated for when the initial learning phase takes place with restricted capacity." Unfortunately, there was very little follow-up on Elman's suggestions until the recent emergence of curriculum learning [97], which has since seen significant growth [98, 99]. Curriculum learning focused initially on managing the presentation of data, but has since expanded to include models of growing networks, but without connecting with the biology of development. Work on developmental robotics [34, 35, 99–102] has more in common with the DI framework, with its focus on autonomous learning with a developmental construction of internal models.

The Importance of Inherent Integration

In addition to structural, functional, and adaptive depth, the other major attribute that makes systems with natural intelligence possible is that they are *inherently integrated* at every level of the deep, multi-scale adaptive process described above. Whether the animal is simple or complex, infant or mature, it always experiences the world *as a whole*, while DL/ML systems only experience the parts captured by their training data and their task. An animal does not need to "maintain consistency" between its models explicitly, or to merge them post facto; it has only one multimodal, multi-scale model of how the world works in all its complexity, and the model inherently integrates the perceptual and behavioral affordances of the animal. Natural intelligence is always general, even at the simplest level.

A crucial capacity enabled by this integration is pervasive zero-shot generalization. In an integrated world model that is self-consistent across all modalities and experiences, even very novel stimuli can find system-wide resonances leading to sensible inference—especially after the animal has significant experience of its world. All intelligent animals have this instinctive, inherent "common sense." In contrast, for DL/ML systems, there is no world outside of their training data, which is a very limited simulacrum of the world, so a lot of real-world stimuli seem far out-of-sample. That is a major reason why common sense continues to elude AI so far, and why producing AI with this critical attribute will require an approach based on inherent integration.

The Importance of Embodiment

In cognitive psychology, embodiment is the idea that mental function is not just a product of the brain, but of the brain and the rest of the body as an integrated system embedded within a specific environment [103, 104]. It is a powerful idea that is often seen in opposition to a purely computational view of the mind and purely control theoretic models of movement. Among other things, this is because the embodied agent can generate cognitive states and motor behavior through emergent coordination rather than explicit information processing or signaling [64, 105, 106].

From an AI perspective—and thus within the DI framework—embodiment is especially important because it grounds mental functions in the physical reality experienced by the agent rather than just in a dataset. The critical point is that the embodied agent (animal or AI system) and the environment it is embedded in (the real world) are both integrated systems operating under the same laws, i.e., the laws of physics, and can actually be seen as comprising a single integrated complex system. This means that the agent's experience is *inherently* self-consistent, and thus grounded and generalizable. Contrast this with a purely computational (simulated) agent that is not necessarily bound by the laws of physics experiencing a dataset that is, at best, an extremely limited, distorted, and selective view of reality. Expecting such a system to then be intelligent in the real world is unrealistic.

To be sure, advanced intelligence builds abstractions on top of direct sensorimotor experience, and one of the most important questions that should be explored in the DI framework is how, in the course of evolution, the ability to create abstract representations has arisen in animals. There has long been great debate in AI about symbolic processing and compositionality, which has been called "the central problem of AI" [107]. However, once the dualistic view of mind and body [108, 109] is rejected, it is obvious that any symbolic processing emerges necessarily from the *physics* of the brain-body system—especially the neural networks of the brain [24]. The central question, then, is: How does a brain-like physical system achieve this? Several models have been proposed to address this at the level of artificial neural networks [110–116], but all of them have serious limitations. There also have been experimental studies to understand how concepts, numbers, words, and other symbolic entities might be represented and composed in the human brain [117–122]. However, interpreting these results is complicated by the immense complexity of the systems being considered (e.g., the human brain), resulting in a focus on simplistic tasks and experiments. From a DI perspective, with its focus on evolution, development, and embodiment (in addition to neural learning), the way to understand the

mechanisms of abstraction and symbolic processing is to study it first in simpler animals, to understand their simpler underlying neural mechanisms, and then build systematically upon that understanding. For example, experiments have shown that a sense of numerical value and numerical order exists in birds [123], and perhaps even in fish [124]. By understanding the neural basis of this simplest type of abstract processing, it might be possible to understand the more complex kind seen in humans. Similarly, it has been suggested that the abstract, high-dimensional representations of the cortex may be built on the scaffolding of spatial, 2-dimensional concrete representations of place in the hippocampus [125]. Exploring the path from embodiment to abstraction is, therefore, the most principled path to understanding the mechanisms of higher cognition. Doing this will be extremely difficult in practice, but the first step must be to devise new experiments that explore the neural processes underlying any primitive abstract processing capacities in simpler animals, focusing on the underlying neural architectures, modules, and processes. Computational modeling based on the results can then be used to generate further hypotheses in more complex animals, and these can be rejected or validated using new experiments. It would also be useful to apply lessons from purely computational models [110–116, 126, 127], and from the vast quantitative literature on conceptual representation and processing in humans [120, 128–131]. A potentially promising, general, and biologically plausible way to understand how the ability for abstract thinking could emerge from evolutionarily more primitive cognitive tasks such as sensorimotor prediction might be to use free-energy and predictive coding-based approaches to mental function [132, 133]. Indeed, recent work has demonstrated how such processing could lead to the emergence of abstractions from a sub-symbolic neural substrate [134–136].

The Significance of Modularity

Hierarchical modularity is perhaps the single most important "enabling technology" underlying the emergence of intelligence (and all other attributes of complex living organisms) [42]. Not only does it allow evolution to build systems of arbitrary complexity without encountering catastrophic failure, it is also crucial to the ability of a complex agent to generate useful complex behavior in the real world because it allows behavior to be produced through *selection, combination* and *hierarchical encapsulation* of modular primitives rather than explicit construction. This is a principle familiar to engineers at the structural level—most complex design and construction is now done using modules—but biology uses modularity in both structure and function. In cognitive science, the latter has been studied most intensively in the context of motor control. The embodiment of any organism configures

coordination modes or *synergies* throughout the brain-body system, so that neural structures and muscles across several joints are constrained collectively, and global responses can arise without information propagating explicitly through all layers of the deep but slow brain-body system [63, 64]. For example, specific muscles act in an inherently coordinated way because of their connectivity with the central pattern generators (CPGs) of the spinal cord [87–91], and groups of muscles develop prototypical activation patterns called *muscle synergies* that are used as primitives in the construction of a whole range of complex movements [25, 63, 94, 137–140]. This means that the *actual* degrees of freedom available to the system in any specific situation are fewer than the entire combinatorial space of all degrees of freedom—thus addressing the so-called *degrees of freedom problem* [141, 142]. Essentially, the modules and their configuration predefine a rich but lower-dimensional *latent repertoire* of behaviors, and a high-level controller—the brain—simply needs to specify the code that unlocks a specific behavior rather than directing the individual muscles in detail [91]. This idea is also implicit in the subsumption model of behavior [143, 144], and the use of motor primitives in robots [145, 146].

While synergies are seen most clearly in motor control, the concept is widely applicable across the entire network-of-networks system, which is why synergies have been termed "the atoms of brain and behavior" [147]. Attractors in recurrent networks are an example of coordination modes and are likely to be widely used across the nervous system [148]. So are the patterns generated by CPGs [88]. It has been suggested that the entire cortex could usefully be seen as a very complex, hierarchical central pattern generator consisting of modules of neural subpopulations forming interacting CPGs [149]. Others have also noted the hierarchical modular organization of the cortex [84, 150], and a general theory of intelligence has recently been proposed that sees cortical columns as information processing modules that represent information and learn by making local predictions [151, 152]. An especially interesting hypothesis is to see intelligence, understanding, and even life in terms of emergent modules within the self-organizing networks comprising an organism [57, 62, 153–155]. This is completely consistent with the DI perspective, which sees all mental processes—perception, cognition, memory, and behavior—as the emergence of synergistic activation patterns across networks of sensors, neurons, and musculoskeletal elements. In this sense—and as also implied by [62]—there is no essential difference between "thought" and "action". It is just that the networks involved in "thought" are networks of neurons, and those involved in action are networks of both neurons and musculoskeletal elements. It is worth noting that coordination modes can also be dynamic—emerging as metastable, context-dependent attractors in multi-scale networks [106, 156–161].

Most neural architectures are inherently—though superficially—modular, and the idea of exploiting structural modularity more directly to enable higher-level cognitive processes has been well-explored [5, 23, 85, 86, 162–166]. It is now being applied explicitly to symbolic tasks [47, 167], albeit for specific problems and deriving more from symbolic abstractions than embodied biology. Modular self-organized neural networks have also been proposed as the basis of sensorimotor integration [168, 169].

One place where evolutionary and developmental adaptation in a hierarchically modular agent could be applied fruitfully to AI is in selective attention. It has been shown that the reason humans can learn new reinforcement learning (RL) tasks rapidly is that they abstract the complexity of the given stimulus into lower-dimensional representations through attentional mechanisms [170], but it is not clear how they learn which features to attend to. In the DI framework, evolution would have already provided modular priors that privilege specific feature classes, and developmental learning starting with very simple tasks would have allowed the agent to refine them to learn what *types* of features are generally useful to attend to in the real world. Thus, the system goes into any specific task with strong generically useful attentional biases, and RL simply needs to select and shape them rather than discovering and learning them from scratch.

Engineering Deep Intelligence

Defining a Feasible Framework

While the DI framework is motivated mainly as a way to understand intelligence better, the goal for AI must be to turn it into an engineering framework. In doing so, the goal would *not* be to replicate the entire process of animal evolution and development—an impossible task in any case. Rather, the approach would be as follows: a) To operationalize the principles underlying the success of evolution and developmental learning; and b) To incorporate into AI systems the architectures, modules, and processes that underlie intelligence in behaving animals.

Evolutionary and developmental biologists have explicated the principles of evolvability in great detail over the last several decades [20, 44–47] (as discussed in the section entitled "The Deep Intelligence View"), and more recently under the rubric of evolutionary developmental biology (EvoDevo) [20, 171–173]. A DI-based system would use these principles to build sequentially more complex intelligent systems by explicit complexification rather than artificial evolution. Each system would display integrated intelligence at its own level, and become the basis for the next more complex system. At each level, the brain and body

architectures, modules, and processes would derive from those observed in animals—albeit with some abstraction, and at a feasible level of detail.

The learning process would begin with a simple, modular system with limited but integrated perception, cognition, and behavior. The system would learn neurally how to exploit its limited capabilities in its limited environment, then add a bit more complexity through modular operators such as duplication, growth, splitting, modularization, etc., creating new sensory, cognitive, and motor modalities emergently as variations or combinations of the prior ones, learning more by building on what has already been learned, and so on, bootstrapping to a full-scale complex system by repeated cycles of alternating *complexification* and *additive learning*—remaining integrated all the while. The developmental complexification and learning could be nested within the outer loop of evolutionary complexification, but it would probably be more feasible to structure the process as alternating between architectural and modular complexification (evolution) and functional complexification (development) interleaved with learning. Figure 4 illustrates this process conceptually in comparison with the DL/ML approach.

An extremely simple version of this approach can be seen in the work of Sims [74, 75] and others, but the explicit use of simulated evolution would need to be replaced by a more scalable framework, one that incorporates evolutionary developmental insights more directly with neural learning. Developing such a framework—even for neural networks alone—is quite non-trivial. At a minimum, it would require defining canonical repertoires of (a) modules; (b) architectures; (c) adaptive mechanisms; (d) developmental operators to complexify modules; and (e) evolutionary operators to grow and reconfigure the system. All of these would be grounded in biology and ranging across the spectrum of vertebrate and arthropod evolution, development, and neural learning, and, as a result, across many spatial and temporal scales, as is the case in biological systems. Of course, all five things would need to be instantiated in computational or physical models. Human AI engineers would focus on designing richer repertoires and generative programs rather than specific large-scale neural architectures and training algorithms. Most importantly, the modules, architectures, and mechanisms of this generative framework would come from those of animal biology rather than abstract formalisms such as Markov decision processes, predicate logic, causal analysis, or even uniformly structured neural networks. The animal may be a kludge produced by evolution's tinkering over billions of years [174], but it is this kludge that is *actually* intelligent in ways that human ingenuity still cannot replicate. AI should respect the kludge and stop trying to fit the complexity of Nature's imagination into simplified, abstract boxes. This does not mean that every molecular detail and every voltage spike has to be accounted for, or

Fig. 4 Conceptual view of how the DL/ML and DI approaches would produce complex natural intelligence. The red arrows for the DL/ML systems indicate explicit integration; the red frame for the DI systems indicates inherent integratedness
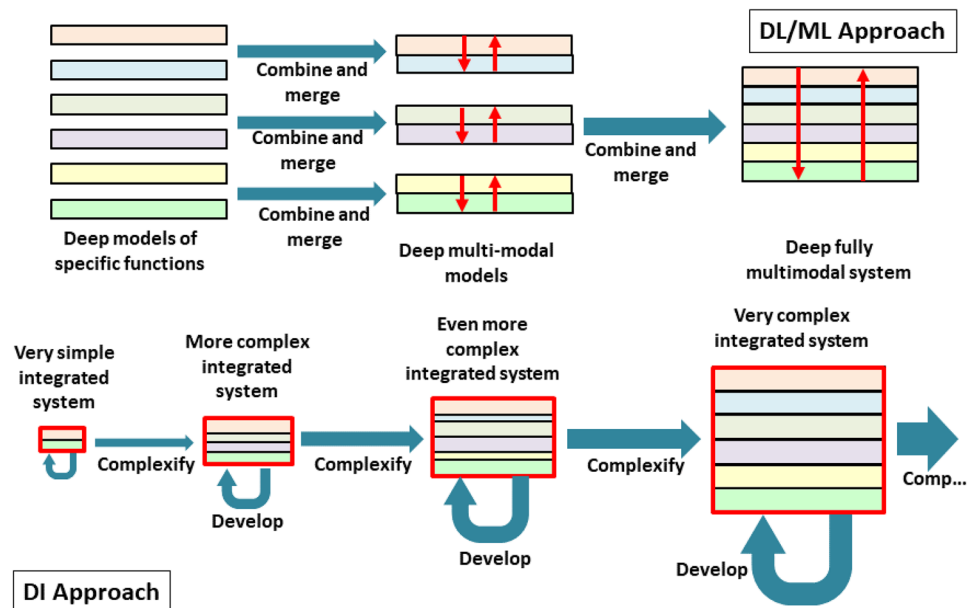
that mathematical models cannot be used. Quite the contrary! The goal should be to build better mathematical and computational models that capture more of the essential features of the biology of intelligence at the level that is most appropriate—and, of course, feasible. To do this, it is critical to understand *all* of the biology and underlying intelligence, not just neuroscience.

Role of Supervised Learning

Given the critique of supervised learning-based approaches laid out in this paper and the ubiquity of such approaches in AI today, it might be asked if supervised learning has a place within the DI framework at all. Clearly, it must, because many complex behaviors and skills can only be learned through supervision and corrective feedback. In humans, this includes things such as learning to play a musical instrument, to do mathematics, or even to use language correctly. The key here is that supervised learning must build upon and exploit the fundamentally integrated and self-organizing nature of the DI system rather than replacing it. In general, supervised learning should be seen as a *late-stage mechanism* in a system where the DI process has configured—and continues to configure—the primitives that supervised learning needs. This paradigm of self-organizing processes laying the groundwork for more complex, incremental, and careful learning is seen in many parts of the brain. For example, the evolutionary architecture of the early visual system and the self-organization of feature detectors during development provide a general basis for the rapid learning of more detailed skills such as object segmentation, recognition, etc. that may need

more corrective feedback. Another example is how muscle synergies [137, 138, 142]—presumably configured through evolution and early development—can then form the primitive basis functions [145, 146] for an ever-growing repertoire of complex movements, many of them, e.g., dance moves, requiring careful supervised learning. The contention is that a system produced by the DI process will, in fact, be *more* ready to do supervised (and reinforcement) learning across a range and combination of modalities, and will do so much more rapidly, than purely supervised systems, thus coming closer to the ideal seen in animals. This point has been demonstrated in a recent paper from my lab, where rapid unsupervised learning in a simple, hippocampus-like model generated a place field substrate for subsequent one-shot reinforcement learning of goal-directed navigation [175].

To some degree, this approach of using pre-configuration of priors to facilitate supervised learning is already used in a simplistic way when self-supervised restricted Boltzmann machines [3] are used to learn initial features for subsequent supervised learning [176], and in the use of feature transfer to enable rapid learning across tasks [177]. However, this idea needs to be generalized and applied in *integrated* systems rather than within narrow modalities. Supervised learning in animals is also unlikely to use back-propagation, though that is not necessarily a barrier in artificial systems once the basis system has been configured. In many—perhaps all—cases, more biologically plausible alternatives such as contrastive [176], self-supervised [175, 178], or resonance-based [23, 179] learning as well as free-energy and predictive coding approaches [132–156] might be sufficient to achieve the same goal when paired with a DI process generating good prior biases.

Conclusion

There is currently a great deal of debate on whether AI is going in the "right direction" with its focus on scaling up deep learning systems [10–16, 180–182]. The debate focuses on such issues as symbolic processing, causal reasoning, compositionality, etc., with some experts suggesting that these things will need to be incorporated by design into the current models. A major aim of this position paper has been to suggest that, since all these capabilities arise naturally in a self-organizing complex system, i.e., the embodied brain, their origins, and mechanisms can be understood by studying them in that natural system rather than coming up with unnatural, biologically implausible engineering methods and symbolic abstractions. Here, it is important to point out that, while the mainstream of AI today is focused on the DL/ML approaches, and expects to achieve general intelligence through that route, it is more likely that such intelligence will emerge from the work in areas such as evolutionary, developmental, and cognitive robotics [30–35, 100–102], where embodied agents learn complex tasks in a more biologically motivated framework. However, this work is at a very early stage and is still focused on specific functions or modalities, such as morphology, control, imitation learning, language acquisition, vision, etc. A full DI framework would eventually need to apply these methods to inherently integrated systems.

One final point: Natural intelligence will not be achieved as long as the focus of AI is on building systems purely to serve human purposes. This only creates glorified screwdrivers. A system with natural intelligence must be autonomous, have its own—probably unexplainable—purposes, and learn all its life in an open-ended way. Such a system may not be immediately useful and may even be dangerous if it was sufficiently complex, but until such systems are built, AI is just the building of smart tools, not intelligent systems [183].

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals.

Informed Consent The work in this paper did not involve any studies requiring informed consent.

Conflict of Interest The author declares no competing interests.

References

1. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

2. Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. Adv Neur Inform Proc Syst. 2007;153–160.

3. Hinton GE. Learning multiple layers of representation. Trends Cogn Sci. 2007;11:428–34.

4. Ciresan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification, Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. 2012;3642–3649. doi:https://doi.org/10.1109/cvpr.2012.6248110.

5. Krizhevsky A, Sutskever I, Hinton G. ImageNet classification with deep convolutional neural networks. Adv Neur Inform Proc Syst. 2012.

6. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. https://doi.org/10.1038/nature14539.

7. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw. 2015;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003.

8. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. Adv Neur Inform Proc Syst 2017. arXiv:1706.03762.

9. Sejnowski TJ. *The deep learning revolution*. MIT press. 2018.

10. Pearl J, McKenzie D. The book of why: the new science of cause and effect. Basic Books; 2018.

11. Harnett K. To build truly intelligent machines, teach them cause and effect. Quanta. 2018. https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/.

12. Marcus G, Davis E. Rebooting AI: building artificial intelligence we can trust. Pantheon. 2019.

13. Heaven D. Why deep-learning AIs are so easy to fool. Nature. 2019;574:163–6. https://doi.org/10.1038/d41586-019-03013-5.

14. Mitchell M. Artificial intelligence: a guide for thinking humans. Strauss and Giroux: Farrar; 2019.

15. Brooks RA. The cul-de-sac of the computational metaphor: a talk by Rodney Brooks. Edge. 2019. https://www.edge.org/conversation/rodney_a_brooks-the-cul-de-sac-of-the-computational-metaphor.

16. Marcus G, Davis E, Aaronson S. A very preliminary analysis of DALL-E 2. 2022. arXiv:2204.13807 [cs.CV].

17. Minai AA, Braha D, Bar-Yam Y. Complex systems engineering: a new paradigm, in complex engineered systems: science meets technology, D. Braha, A.A. Minai, and Y. Bar-Yam (Eds.). Springer Verlag. 2006;1–22.

18. Raff RA. The shape of life: genes, development, and the evolution of animal form. University of Chicago Press. 1996.

19. Schlosser G, Wagner GP (eds.). Modularity in development and evolution. Univer Chic Press. 2004.

20. Carroll SB. Endless forms most beautiful: the new science of evo-devo and the making of the animal kingdom. WW Norton & Company. 2005.

21. Wagner A. The origins of evolutionary innovations. Oxford: Oxford University Press; 2011.

22. Meunier D, Lambiotte R, Bullmore E. Modular and hierarchically modular organization of brain networks. Front Neurosci. 2010;4. https://doi.org/10.3389/fnins.2010.00200.

23. Grossberg S. The complementary brain: Unifying brain dynamics and modularity. Trends Cogn Sci. 2000;4:233–46. https://doi.org/10.1016/S1364-6613(00)01464-9.

24. Grossberg S. Conscious mind, resonant brain: how each brain makes a mind. Oxford University Press; 2021.

25. d'Avella A, Pai DK. Modularity for sensorimotor control: evidence and a new prediction. J Mot Behav. 2010;42:361–9.

26. Geary DC. The origin of mind: evolution of brain, cognition, and general intelligence. Am Psychol Assoc. 2005.

27. Thelen E, Smith LB. A dynamic systems approach to the development of cognition and action. MIT Press; 1994.

28. Kelso JAS. Dynamic patterns: the self-organization of brain and behavior. Bradford Books; 1995.

29. Goldfield EC. Emergent forms: origins and early development of human action and perception. Oxford University Press; 1995.

30. Nolfi S, Floreano D. Evolutionary robotics: the biology, intelligence, and technology of self-organizing machines. MIT press. 2000.

31. Weng J, McClelland J, Pentland A, Sporns O, Stockman I, Sur M, Thelen E. Autonomous mental development by robots and animals. Science. 2001;291:599–600.

32. Jin Y, Meng Y. Morphogenetic robotics: a new emerging field in developmental robotics. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Reviews and Applications. 2011;41(2):145–60.

33. Weng J. Symbolic models and emergent models: a review. IEEE Trans Auton Ment Dev. 2011;4:29–54.

34. Cangelosi A, Schlesinger M. Developmental Robotics: from babies to robots. MIT Press. 2015.

35. Vujovic V, Rosendo A, Brodbeck L, Iida F. Evolutionary developmental robotics: Improving morphology and control of physical robots. Artificial Life. 2017;23(2):169–185. https://doi.org/10.1162/ARTL_a_00228.

36. Merel J, Botvinick M, Wayne G. Hierarchical motor control in mammals and machines. Nat Commun. 2019;10:5489. https://doi.org/10.1038/s41467-019-13239-6.

37. Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Hassabis D. Reinforcement learning, fast and slow. Trends Cogn Sci. 2019;23:408–22. https://doi.org/10.1016/j.tics.2019.02.006.

38. Barretto A, Hou S, Borsa D, Silver D, Precup D. Fast reinforcement learning with generalized policy updates. PNAS. 2020;117:30079–87.

39. Spearman C. General intelligence, objectively determined and measured. Am J Psychol. 1904;15:201–93.

40. Cattell EB. Theory of fluid and crystallized intelligence: a critical experiment. J Educ Psychol. 1963;54:1–22.

41. Kahneman D. Thinking fast and slow. Straus and Giroux: Farrar; 2011.

42. Callebaut W, Rasskin-Gutman D (eds.). Modularity: understanding the development and evolution of natural complex systems. MIT Press. 2005.

43. Whitacre JM. Degeneracy: A link between evolvability, robustness and complexity in biological systems. Theor Biol Med Model. 2010;7:6. https://doi.org/10.1186/1742-4682-7-6.

44. Dawkins R. The evolution of evolvability, In Langton C. G. (Ed.), Artificial life: the proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems. Addison-Wesley Publishing Co. 1988;201–220.

45. Kirschner M, Gerhart J. Evolvability. PNAS. 1998;95(15):8420–7. https://doi.org/10.1073/pnas.95.15.8420.

46. Wagner A. Robustness and evolvability in living systems. Princeton University Press; 2005.

47. Kerg G, Mittal S, Rolnick D, Bengio Y, Richards B, Lajoie G. On neural architecture inductive biases for relational tasks. 2022. arXiv:2206.05056 [cs.NE]. https://doi.org/10.48550/arXiv.2206.05056.

48. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big?. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). 2021;610–623. https://doi.org/10.1145/3442188.3445922.

49. Chen MX, Firat O, Bapna A, Johnson M, Macherey W, Foster GF, Jones L, Parmar N, Schuster M, Chen Z, Wu Y, Hughes M. The best of both worlds: combining recent advances in neural machine translation, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia (Long Papers). 2018;76–86.

50. Liu X, Duh K, Liu L, Gao J. Very deep transformers for neural machine translation. 2020. arXiv:2008.07772 [cs.CL].

51. Heaven WD. OpenAI's new language generator GPT-3 is shockingly good—and completely mindless. MIT Technol Rev. 2020. https://www.technologyreview.com/2020/07/20/1005454/openai-machine-learning-language-generator-gpt-3-nlp/.

52. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, va den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S, Grewe D, Nham J, Kalchbrenner N, Sutskever I, Lillicrap T, Leach M, Kavukcuoglu K, Graepel T, Hassabis D. Mastering the game of Go with deep neural networks and tree search. Nature. 2016;529(7587):484–9. https://doi.org/10.1038/nature16961.

53. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, van den Driessche G, Graepel T, Hassabis D. Mastering the game of Go without human knowledge. Nature. 2017;550(7676):354–9. https://doi.org/10.1038/nature24270.

54. Girshick, R.B. (2015) Fast R-CNN, 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448.

55. OpenAI (2022) ChatGPT: Optimizing language models for dialogue. https://openai.com/blog/chatgpt/.

56. Ramesh A, Pavlov M, Goh G, Gray S, Voss C, Radford A, Chen M, Sutskever I. Zero-shot text-to-image generation. 2021. https://arxiv.org/abs/2102.12092v2.

57. Minai AA, Perdoor M, Byadarhaly KV, Vasa S, Iyer LR. A synergistic view of autonomous cognitive systems. Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN'2010). 2010;498–505.

58. Braitenberg V. Vehicles: experiments in synthetic psychology. Cambridge, MA: MIT Press; 1984.

59. Carlson JM, Doyle J. Complexity and robustness. PNAS. 2002;99(supp. 1):2538–45.

60. Tanaka R, Doyle J. Scale-rich metabolic networks: background and introduction. 2004. https://arxiv.org/abs/q-bio/0410009.

61. Zador AM. A critique of pure learning and what artificial neural networks can learn from animal brains. Nat Commun. 2019;10:3770.

62. Latash ML. Understanding and synergy: a single concept at different levels of analysis?. Front Syst Neurosci. 2021;15. https://doi-org.uc.idm.oclc.org/10.3389/fnsys.2021.735406.

63. Latash ML. Motor synergies and the equilibrium-point hypothesis. Mot Control. 2010;14(3):294–322. https://doi.org/10.1123/mcj.14.3.294.

64. Riley MA, Kuznetsov N, Bonnette S. State-, parameter-, and graph-dynamics: constraints and the distillation of postural control systems. Science & Motricité. 2011;74:5–18. https://doi.org/10.1051/sm/2011117.

65. Dobzhansky T. Nothing in biology makes sense except in the light of evolution. American Biology Teacher. 1973;35(3):125–9. https://doi.org/10.1093/icb/4.4.443.

66. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962;160:106–54.

67. Hubel DH, Wiesel TN. Brain and visual perception. New York: Oxford Press; 2005.

68. Fogel LJ, Owens AJ, Walsh MJ. Artificial intelligence through simulated evolution. NY: John Wiley; 1966.

69. Holland JH. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. University of Michigan Press; 1975.

70. Goldberg D. Genetic algorithms in search, optimization and machine learning. Addison-Wesley Professional. 1989.

71. Stanley KO, Miikkulainen R. Evolving neural networks through augmenting topologies. Evol Comput. 2002;10(2):99–127. https://doi.org/10.1162/106365602320169811.

72. Stanley K, Miikkulainen R. A taxonomy for artificial embryogeny. Artif Life. 2003;9(2):93–130.

73. Clune J, Beckmann BE, Ofria C, Pennock RT. Evolving coordinated quadruped gaits with the HyperNEAT generative encoding. Proc IEEE Cong Evol Comp. 2009;2764–2771.

74. Sims K. Evolving virtual creatures. Proceedings of SIGGRAPH '94. 1994;15–22.

75. Sims K. Evolving 3D morphology and behavior by competition. Artif Life. 1994;1:353–72. https://doi.org/10.1162/artl.1994.1.4.353.

76. Rieffel J, Pollack J. An endosymbiotic model for modular acquisition in stochastic developmental systems. Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems (ALIFE X). 2006.

77. Kirschner MW, Gerhart JC. The plausibility of life: resolving Darwin's dilemma. Yale University Press; 2005.

78. Gerhart J, Kirschner M. The theory of facilitated variation. PNAS. 2007;104(Supp. 1):8582–9.

79. Kimura M. The neutral theory of molecular evolution. Press: Cambridge Univ; 1983.

80. Huneman P. Neutral spaces and topological explanations in evolutionary biology: lessons from some landscapes and mappings. Philosophy of Science. 2018;85(5):969–83. https://doi.org/10.1086/699759.

81. Kauffman SA. The origins of order: self-organization and selection in evolution. Oxford University Press; 1993.

82. Siebert BA, Hall CL, Gleeson JP, Asllani M. Role of modularity in self-organization dynamics in biological networks. Phys Rev E. 2020;102:052306. https://doi.org/10.1103/PhysRevE.102.052306.

83. Mountcastle VB. The columnar organization of the neocortex. Brain. 1997;120:701–22.

84. Bressler SL, Tognoli E. Operational principles of neurocognitive networks. Int J Psychophysiol. 2006;60(2):139–48. https://doi.org/10.1016/j.ijpsycho.2005.12.008.

85. Abeles M. Local cortical circuits: an electrophysiological study. Springer; 1982.

86. Buzsáki G. Neural syntax: cell assemblies, synapsembles, and readers. Neuron. 2010;68:362–85.

87. Grillner S. The motor infrastructure: from ion channels to neuronal networks. Nat Rev Neurosci. 2003;4:673–86.

88. Grillner S. Biological pattern generation: the cellular and computational logic of networks in motion. Neuron. 2006;52:751–66.

89. Grillner S, Deliagina T, Ekeberg O, El Manira A, Hill RH, Lansner A, Orlovsky GN, Wallén P. Neural networks that coordinate locomotion and body orientation in lamprey. Trends Neurosci. 1995;18:270–9.

90. Grillner S, Hellgren J, Ménard A, Saitoh K, Wikström MA. Mechanisms for selection of basic motor programs – roles for the striatum and pallidum. Trends Neurosci. 2005;28:364–70.

91. Ijspeert AJ, Crespi A, Ryczko D, Cabelguen JM. From swimming to walking with a salamander robot driven by a spinal cord model. Science. 2007;315:1416–20.

92. Simon HA. Near decomposability and complexity: how a mind resides in a brain, In Morowitz, H.J. and Singer, J.L. (eds). The Mind, the Brain, and Complex Systems, Addison-Wesley. 1995.

93. Simon HA. Near decomposability and the speed of evolution. Ind Corp Chang. 2002;11:587–99.

94. Cheung VCK, Seki K. Approaches to revealing the neural basis of muscle synergies: a review and a critique. J Neurophysiol. 2021;125:1580–97. https://doi.org/10.1152/jn.00625.2019.

95. Heess NM, Wayne G, Tassa Y, Lillicrap TP, Riedmiller MA, Silver D. Learning and transfer of modulated locomotor controllers. 2016. https://arxiv.org/abs/1610.05182.

96. Elman JL. Learning and development in neural networks: the importance of starting small. Cognition. 1993;48:71–99.

97. Bengio Y, Louradour J, Collobert R, Weston J. Curriculum learning. In Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09). 2009;41–48. https://doi.org/10.1145/1553374.1553380.

98. Wang X, Chen Y, Zhu W. A survey on curriculum learning. IEEE Trans Pattern Anal Mach Intell. 2020. https://doi.org/10.1109/TPAMI.2021.3069908.

99. Soviany P, Ionescu RT, Rota P, Sebe N. Curriculum learning: a survey. Int J Comput Vision. 2022;130:1526–65. https://doi.org/10.1007/s11263-022-01611-x.

100. Weng J. Developmental robotics: theory and experiments. Int J Humanoid Rob. 2004;1:199–236.

101. Deshpande A, Kumar R, Minai AA, Kumar M. Developmental reinforcement learning of control policy of a quadcopter UAV with thrust vectoring rotors. Proc 2020 Dyn Syst Contr Confer. 5–7 Oct. 2020. https://arxiv.org/abs/2007.07793.

102. Nguyen SM, Duminy N, Manoury A, Duhaut D, Bouche C. Robots learn increasingly complex tasks with intrinsic motivation and automatic curriculum learning. Künstlische Intelligenz. 2021;35:81–90. https://doi.org/10.1007/s13218-021-00708-8.

103. Chiel HJ, Beer RD. The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. Trends Neurosci. 1997;20:553–7.

104. Chemero A. Radical embodied cognitive science. Bradford Books; 2011.

105. Pfeifer R, Lungarella M, Iida F. Self-organization, embodiment, and biologically inspired robotics. Science. 2007;318:1088–93.

106. Schöner G. The dynamics of neural populations capture the laws of the mind. Top Cogn Sci. 2020;12:1257–71.

107. Smolensky P. On the proper treatment of connectionism. Behav Brain Sci. 1988;11(1):1–23.

108. Descartes R. Meditations on first philosophy. in *The Philosophical Writings of René Descartes 2* (1984), translated by J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Camb Univ Press. 1641;1–62.

109. Hart WD. Dualism. In: Guttenplan S, editor. A companion to the philosophy of mind. Oxford: Blackwell; 1996. p. 265–7.

110. Eliasmith C. How to build a brain: a neural architecture for biological cognition. Oxford University Press; 2013.

111. Smolensky P. Symbolic functions from neural computation. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2012;370(1971):3543–69.

112. Besold TR, Garcez ADA, Bader S, Bowman H, Domingos PM, Hitzler P, Kühnberger K, Lamb LC, Lowd D, Lima PMV, de Penning L. Neural-symbolic learning and reasoning: a survey and interpretation. CoRR abs/1711.03902. 2017. arXiv preprint arXiv:1711.03902.

113. Schlag, and Schmidhuber, J. Learning to reason with third order tensor products. Adv Neural Inf Process Syst. 2018;2018:9981–93.

114. Huang Q, Deng L, Wu D, Liu C, He X. Attentive tensor product learning. Proceedings of the 33rd AAAI Conference on Artificial Intelligence. 2019;1344–1351.

115. D'Avila Garcez A, Lamb LC. Neurosymbolic AI: the 3rd wave. 2020. arXiv 2012.05876. https://arxiv.org/abs/2012.05876.

116. Smolensky P, McCoy RT, Fernadez R, Goldrick M, Gao J. Neurocompositional computing: from the Central Paradox of Cognition to a new generation of AI systems. 2022. arXiv:2205.01128v1 [cs.AI].

117. Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff MA, Michel F. The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain. 2000;123(Pt 2):291–307. https://doi.org/10.1093/brain/123.2.291.

118. Harvey BM, Klein BP, Petridou N, Dumoulin SO. Topographic representation of numerosity in the human parietal cortex. Science. 2013;341:1123–6. https://doi.org/10.1126/science.1239052.

119. Amalric M, Dehaene S. Origins of the brain networks for advanced mathematics in expert mathematicians. PNAS. 2016;113:4909–17. https://doi.org/10.1073/pnas.1603205113.

120. Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic maps that tile human cerebral cortex. Nature. 2016;532:453–8.

121. García AM, Moguilner S, Torquati K, García-Marco E, Herrera E, Muñoz E, Castillo EM, Kleineschay T, Sedeño L, Ibáñez A. How meaning unfolds in neural time: Embodied reactivations can precede multimodal semantic effects during language processing. Neuroimage. 2019;197:439–49. https://doi.org/10.1016/j.neuroimage.2019.05.002.

122. Leminen A, Smolka A, Duñabeitia JA, Pliatsikas C. Morphological processing in the brain: The good (inflection), the bad (derivation) and the ugly (compounding). Cortex. 2019;116:4–44. https://doi.org/10.1016/j.cortex.2018.08.016.

123. Rugani R, Vallortigara G, Priftis K, Regolin L. Number-space mapping in the newborn chick resembles humans' mental number line. Science. 2015;347:534–6.

124. Vallortigara G. Comparative cognition of number and space: the case of geometry and of the mental number line. Philosophical Transactions of the Royal Society (London) B. 2017;373:20170120. https://doi.org/10.1098/rstb.2017.0120.

125. Hawkins J, Lewis M, Klukas M, Purdy S, Ahmad S. A framework for intelligence and cortical function based on grid cells in the neocortex. Frontiers in Neural Circuits. 2019;12:121. https://doi.org/10.3389/fncir.2018.00121.

126. Kelly MA, Arora N, West RL, Reitter D. Holographic declarative memory: distributional semantics as the architecture of memory. Cogn Sci. 2020;44:e12904. https://doi.org/10.1111/cogs.12904.

127. Smith R, Schwartenbeck P, Parr T, Friston KJ. An active inference approach to modeling structure learning: concept learning as an example. Front Comput Neurosci. 2020;14:41. https://doi.org/10.3389/fncom.2020.00041.

128. Bruffaerts R, De Deyne S, Meersmans K, Liuzzi AG, Storms G, Vandenberghe R. Redefining the resolution of semantic knowledge in the brain: advances made by the introduction of models of semantics in neuroimaging. Neuroscience and Behavioral Reviews. 2019;103:3–13.

129. Zeithamova D, Mack ML, Braunlich K, Davis T, Seger CA, van Kesteren MTR, Wutz A. Brain mechanisms of concept learning. J Neurosci. 2019;39(42):8259–66.

130. Zhang Y, Han K, Worth R, Liu Z. Connecting concepts in the brain by mapping cortical representations of semantic relations. Nat Comm. 2020;11:1877. https://doi.org/10.1038/s41467-020-15804-w.

131. Fernandino L, Tong JQ, Conant LL, Humphries CJ, Binder JR. Decoding the information structure underlying the neural representation of concepts. PNAS. 2022;119:e2108091119. https://doi.org/10.1073/pnas.2108091119.

132. Friston K. The free-energy principle: a unified brain theory. Nat Rev Neurosci. 2010;11:127–38. https://doi.org/10.1038/nrn2787.

133. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 2013;36:181–253.

134. Butz MV. Towards a unified sub-symbolic computational theory of cognition. Front Psychol. 2016;7:925. https://doi.org/10.3389/fpsyg.2016.00925.

135. Butz MV. Event-predictive cognition: a root for conceptual human thought. Top Cogn Sci. 2021;13:10–24. https://doi.org/10.1111/tops.12522.

136. Butz MV. Towards strong AI. Künstlische Intelligenz. 2021. https://doi.org/10.1007/s13218-021-00705-x.

137. Tresch MC, Saltiel P, Bizzi E. The construction of movement by the spinal cord. Nat Neurosci. 1999;2:162–7.

138. d'Avella A, Saltiel P, Bizzi E. Combinations of muscle synergies in the construction of a natural motor behavior. Nat Neurosci. 2003;6:300–8.

139. Latash ML, Scholz JP, Schöner G. Toward a new theory of motor synergies. Mot Control. 2007;11:276–308.

140. Byadarhaly KV, Perdoor MC, Minai AA. A modular neural model of motor synergies. Neural Netw. 2012;32:96–108.

141. Bernstein N. The coordination and regulation of movements. Pergamon Press; 1967.

142. Kuppuswamy N, Harris CM. Do muscle synergies reduce the dimensionality of behavior?. Front Comp Neurosci. 2014;8. https://doi.org/10.3389/fncom.2014.00063.

143. Brooks R. Intelligence without representation. Artif Intell. 1991;47(1–3):139–59. https://doi.org/10.1016/0004-3702(91)90053-M.

144. Brooks R. Cambrian intelligence: the early history of the new AI. MIT Press; 1999.

145. Schaal S, Peters J, Nakanishi J, Ijspeert A. Control, planning, learning, and imitation with dynamic movement primitives. In: Workshop on bilateral paradigms on humans and humanoids. IEEE International Conference on Intelligent Robots and Systems (IROS 2003). Las Vegas, NV, Oct. 27–31. 2003.

146. Schaal S, Mohajerian P, Ijspeert A. Dynamics systems vs. optimal control – a unifying view. In: P. Cisek, T. Drew and J.F. Kalaska (Eds.). Prog Brain Res. 2007;165:425–445.

147. Kelso JAS. Synergies: atoms of brain and behavior, In: Progress in motor control – a multidisciplinary perspective, Sternad D. (ed), Springer. 2007.

148. Amit DJ. Modeling brain function. New York: Cambridge University Press; 1989.

149. Yuste R, MacLean JN, Smith J, Lansner A. The cortex as a central pattern generator. Nat Rev Neurosci. 2005;6:477–83.

150. Bassett DS, Greenfield DL, Meyer-Lindenberg A, Weinberger DR, Moore SW, Bullmore ET. Efficient physical embedding of topologically complex information processing networks in brains and computer circuits, PLoS computational biology, 04/2010. 2010;6(4).

151. Hawkins J, Ahmad S, Cui Y. A theory of how columns in the neocortex enable learning the structure of the world. Frontiers in Neural Circuits. 2017;11:81. https://doi.org/10.3389/fncir.2017.00081.

152. Hawkins J. A thousand brains: a new theory of intelligence. Basic Books; 2021.

153. Yufik YM. Understanding, consciousness and thermodynamics of cognition. Chaos, Solitons Fractals. 2013;55:44–59. https://doi.org/10.1016/j.chaos.2013.04.010.

154. Yufik YM. The understanding capacity and information dynamics in the human brain. Entropy. 2019;21:308. https://doi.org/10.3390/e21030308.

155. Yufik YM, Friston K. Life and understanding: the origins of "understanding" in self-organizing nervous systems. Front Syst Neurosci. 2016;10. https://doi.org/10.3389/fnsys.2016.00098.

156. Tsuda I. Towards an interpretation of dynamic neural activity in terms of chaotic dynamical systems. Behavioral and Brain Sciences. 2001;24:793–847.

157. Rabinovich MI, Huerta R, Varona P, Afraimovich VS. Transient cognitive dynamics, metastability, and decision making. PLoS Comp Biol. 2008;4(5):e1000072. https://doi.org/10.1371/journal.pcbi.1000072.

158. Gros C. Cognitive computation with autonomously active neural networks: an emerging field. Cogn Comput. 2009;1:77–90. https://doi.org/10.1007/s12559-008-9000-9.

159. Marupaka N, Iyer LR, Minai AA. Connectivity and thought: the influence of semantic network structure in a neurodynamical model of thinking. Neural Netw. 2012;32:147–58.

160. Mattia M, Pani P, Mirabella G, Costa S, Del Giudice P, Ferraina S. Heterogeneous attractor cell assemblies for motor planning in

premotor cortex. J Neurosci. 2013;33(27):11155–68. https://doi.org/10.1523/JNEUROSCI.4664-12.2013.

161. Minai AA, Iyer LR, Doumit S, et al. IDEA—itinerant dynamics with emergent attractors: a neural model for conceptual combination. In: Doboli S, et al., editors. Creativity and Innovation: Cognitive, Social, and Computational Approaches. Springer; 2021. p. 195–227.

162. Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern. 1980;36(4):193–202. https://doi.org/10.1007/bf00344251.

163. Sperber D. Modularity and relevance: how can a massively modular mind be flexible and context-sensitive? In *The Innate Mind: Structure and Contents*, Carruthers, P., Laurence, S. and Stich, S. (eds). Oxford University Press. 2005. https://doi.org/10.1093/acprof:oso/9780195179675.003.0004.

164. Iyer L, Doboli S, Minai A, Brown V, Levine D, Paulus P. Neural dynamics of idea generation and the effects of priming. Neural Netw. 2009;22:674–86.

165. Rinkus GJ. A cortical sparse distributed coding model linking mini- and macrocolumn-scale functionality. Front Neuroanat. 2010;4:17. https://doi.org/10.3389/fnana.2010.00017.

166. Iyer LR, Minai AA. CANDID: A neurodynamical model for ddaptive context-dependent idea generation. In: Creativity and innovation. Understanding Complex Systems, Doboli, S., Kenworthy, J.B., Minai, A.A., Paulus, P.B. (eds), Springer, Cham. 2021. https://doi.org/10.1007/978-3-030-77198-0_7.

167. Hinton GE. How to represent part-whole hierarchies in a neural network. 2021. arXiv:2102.12627 [cs.CV]. https://doi.org/10.48550/arXiv.2102.12627.

168. Pouget A, Snyder LH. Computational approaches to sensorimotor transformations. Nature Neuroscience Supp. 2000;3:1192–8.

169. Morse AF, de Greeff J, Belpeame T, Cangelosi A. Epigenetic robotics architecture (ERA). IEEE Trans On Autonomous Mental Development. 2010;2:325–39.

170. Niv Y. Learning task-state representations. Nat Neurosci. 2019;22:1544–1553. https://doi-org.uc.idm.oclc.org/10.1038/s41593-019-0470-8.

171. Müller GB. Evo-devo: extending the evolutionary synthesis. Nat Rev Genet. 2007;8:943–9.

172. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008;134:25–36.

173. Gilbert SF, Bosch TCG, Ledón-Retting C. Eco-evo-devo: developmental symbiosis and developmental plasticity as evolutionary agents. 2015.

174. Marcus G. Kluge: the haphazard evolution of the human mind. Mariner Books. 2009.

175. Alabi A, Vanderelst D, Minai AA (in press). Rapid learning of spatial representations for goal-directed navigation based on a novel model of hippocampal place fields, Neural Networks (in press).

176. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006;313:504–7.

177. Zhuang F, Zhiyuan Q, Duan K, Xi D, Zhu Y, Zhu H, Xiong H, He Q. A comprehensive survey on transfer learning. Proc IEEE. 2021;109:43–76.

178. Carr MF, Jadhav SP, Frank LM. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. Nat Neurosci. 2011;14:147.

179. Carpenter GA, Grossberg S. ART 2: Self-organization of stable category recognition codes for analog input patterns. Appl Opt. 1987;26:4919–30.

180. Aguera y Arcas B. Do large language models understand us?. 2021. https://medium.com/@blaisea/do-large-language-models-understand-us-6f881d6d8e75.

181. Marcus G. Does AI really need a paradigm shift?. 2022. https://garymarcus.substack.com/p/does-ai-really-need-a-paradigm-shift?s=r.

182. Li A. Google engineer claims that its LaMDA conversation AI is 'sentient,' industry disagrees, 9TO5Google. 12 June 2022. https://9to5google.com/2022/06/12/google-ai-lamda-sentient/.

183. Minai A. Between golem and god, 3 Quarks Daily. 2020. https://3quarksdaily.com/3quarksdaily/2021/06/between-golem-and-god-the-future-of-ai.html.