

Review



Remapping and navigation of an embedding space via error minimization: A fundamental organizational principle of cognition in natural and artificial systems

Benedikt Hartl ^{a,1}, Léo Pio-Lopez ^{a,1}, Chris Fields ^a, Michael Levin ^{a,b,*}

^a Allen Discovery Center at Tufts University, Medford, MA, USA

^b Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA

ARTICLE INFO

Editor: J. Fontanari

Keywords:

Evolution
Development
Intelligence
Active inference
Navigation policy
Nested embedding spaces

ABSTRACT

The emerging field of diverse intelligence seeks an integrated view of problem-solving in agents of very different provenance, composition, and substrates. From subcellular chemical networks to swarms of organisms, and across evolved, engineered, and chimeric systems, it is hypothesized that scale-invariant principles of decision-making can be discovered. We propose that cognition in both natural and synthetic systems can be characterized and understood by the interplay between two equally important invariants: (i) the remapping of embedding spaces, and (ii) the navigation within these spaces. Biological collectives, from single cells to entire organisms (and beyond), remap transcriptional, morphological, physiological, or 3D spaces to maintain homeostasis and regenerate structure, while navigating these spaces through distributed error correction. Modern Artificial Intelligence (AI) systems, including transformers, diffusion models, and neural cellular automata, enact analogous processes by remapping data into latent embeddings and refining them iteratively through contextualization. We argue that this dual principle – remapping and navigation of embedding spaces via iterative error minimization – constitutes a substrate-independent invariant of cognition. Recognizing this shared mechanism not only illuminates deep parallels between living systems and artificial models, but also provides a unifying framework for engineering adaptive intelligence across scales.

1. Introduction

Conventional behavioral science tends to focus on animals with brains and nervous systems. Three main lines of thought suggest that this might be too narrow, as schematically illustrated in Figs. 1 and 2. First, an operationalist stance suggests that cognitive terms are interaction protocols, not claims about the essential nature of specific substrates such as brains [1–12]. In other words, an assessment of cognition within any system is, operationally, just a testable hypothesis about whether, for example, the system is most amenable to techniques that rewire its hardware, reset its homeostatic goals, train it via rewards and punishments, or convince it via cogent reasons and psychological strategies. From that perspective, a system's position on the spectrum of persuadability [1] is an empirical question. This operational view facilitates the application of tools from the behavioral scientist's toolbox [5] to a wide range of complex, simple, and even non-living systems to see which formal interaction model offers the most effective prediction

* Corresponding author.

E-mail address: michael.levin@tufts.edu (M. Levin).

¹ Authors contributed equally.

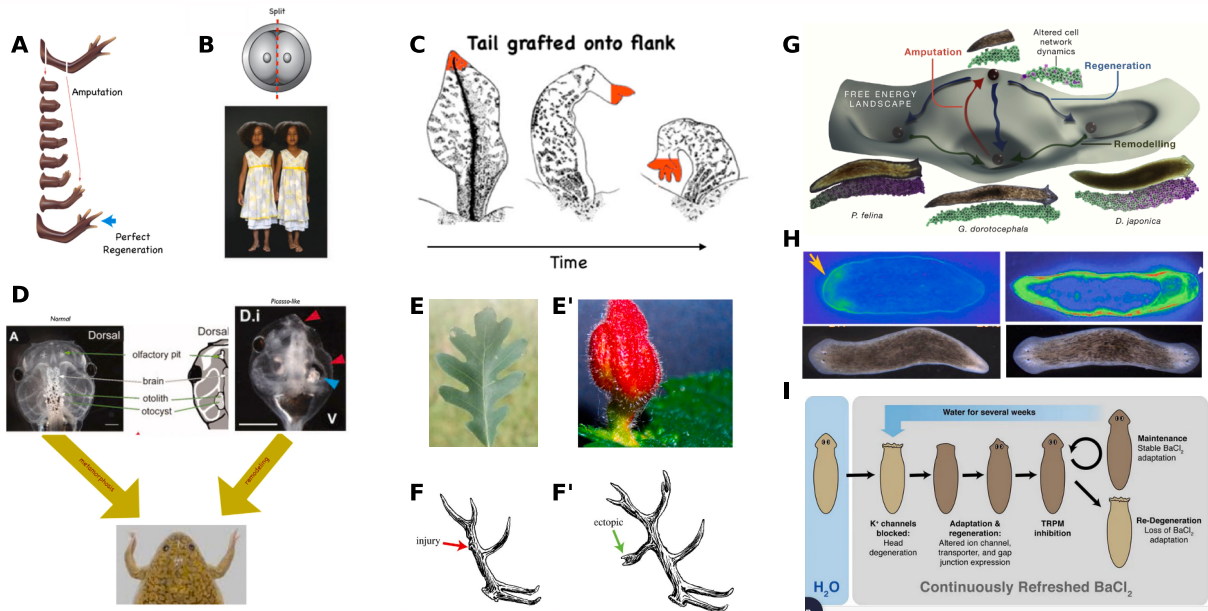


Fig. 1. Biological examples of unconventional morphogenetic problem solving: (A) The cellular collective of regenerative species (such as salamander or axolotl) can regrow limbs, body parts, or organs upon injury [36]. (B) Identical twins independently develop from one zygote that has split into two embryos at an early cleavage stage [36]. (C) Remodeling of a transplanted tail into a limb-like structure on a salamander's flank [260]. (D) Tadpoles in which the craniofacial structures are scrambled still make largely normal frogs [261]. (E) An archetypical white oak leaf contrasted to a novel gall form (E') made by genetically normal plant leaf cells when prompted by signals from a wasp embryo. (F) An injury (red arrow) altered the target morphology of a Siberian wapiti antler, producing during the following two years a new tine (green arrow, F') [262]. (G) A brief bioelectric modulation can reprogram not only head number – see panel (H) below – but also head shape in planaria, shifting a genetically unchanged animal toward species-specific head morphologies, with distinct brain shapes and stem-cell distributions. This illustrates how the cellular collective can explore alternative attractors in morphospace using the same genome, some of which could eventually become canalized into genetically distinct species [263]. (H) Bioelectric signaling can regulate system-level anatomical polarity in planaria. In the normal one-headed animal, the endogenous bioelectric pattern (top left panel) corresponds to a single-head body plan (bottom left panel). Brief bioelectric modulations can rewrite this pattern memory (top right panel) and stabilize a two-headed morphology (bottom right panel) without altering the genome [44,264,280]. (I) An example of novel problem-solving in transcriptional and physiological space is seen in planaria exposed to barium, which initially causes their heads to explode. The animals nevertheless regenerate barium-insensitive heads, solving this novel stressor by efficiently traversing transcriptional space to up- and down-regulate a small number of genes, allowing the individual's physiology to continue despite the otherwise detrimental perturbation [1,45]. Images used with permission from references indicated, except as follows; (A, I) Jeremy Guay of Peregrine Creative; (B) photo by Oudeschool via Wikimedia Commons; (D) Erin Switzer; (E) photographed by Chris Evans, River to River CWMA, <https://www.bugwood.org/>, and (E') used by permission from [265]; (G) Alexis Pietak.

and control. The second line of thought points out that the mechanisms implementing intelligence in the brain are evolutionarily homologous to those in non-neural living substrates. Ion channels, electrical synapses, synaptic machinery, and the algorithms they embody are evolutionarily ancient, and ubiquitous, raising the question of when precisely they became fodder for cognitive tasks (albeit in unconventional contexts), in the long history prior to the formation of brains [8]. Finally, third, developmental biology reveals that familiar bioelectric networks play key roles in coordinating cellular activities toward specific anatomical endpoints, long before nerves appear during morphogenesis, again pointing to the need for models of scaling of competencies from those of the single blastomere to those of an adult [13]; examples are illustrated in Fig. 1.

The emerging field of Diverse Intelligence [4,7,14–16] seeks to understand what all active agents have in common. In moving away from classic assumptions that brains, and a history of evolutionary trial-and-error for behavior in 3D space, are privileged substrates and origin stories for cognition, it seeks a unifying theory for agents of highly diverse composition and provenance [2] that avoids definitional and line-drawing controversies that lack empirical consequences [17]. Evolved, engineered, and hybrid systems at all scales of complexity become, from this philosophically-unbiased perspective, fodder for experiments in learning, problem-solving, and decision-making [3,18]. Progress in this field suggests that neuroscience is not about neurons - it's about biophysical processes and mechanisms that operate as a kind of “cognitive glue”, binding competent subunits (e.g., cells) into collective intelligences that operate in spaces and with spatio-temporal goal horizons (cognitive light cones) much larger than those of their parts [19].

This research program has three unique key features. It is deeply interdisciplinary - revealing novel phenomena at the intersection of biology, computer science, materials science, and cognitive science. It is tractable - the tools of behavioral, computational, and cognitive neuroscience have now been applied far outside the brain and avail the modern worker with very effective tools with

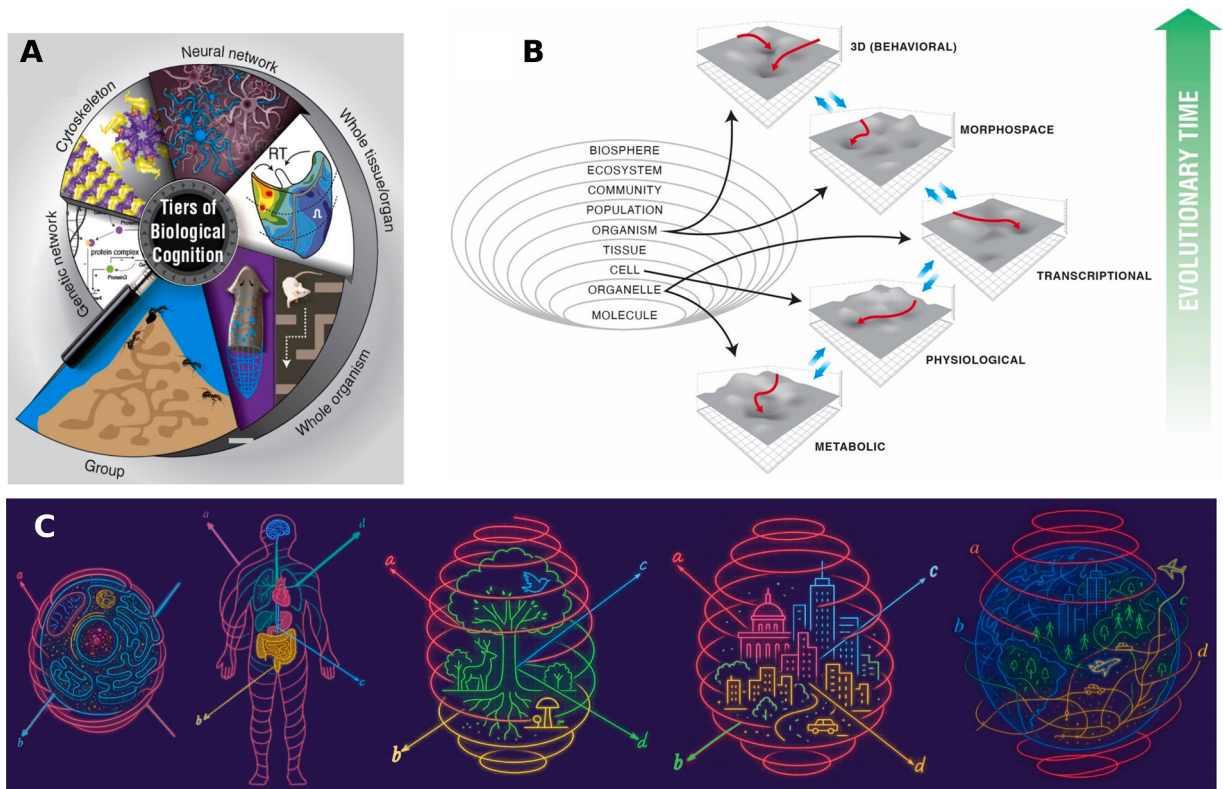


Fig. 2. Biology as Multiscale Competency Architecture (MCA): (A) To explain the immense structural and functional plasticity of biological matter, collective biological systems ought to be seen as an MCA: composites of integrated layers within layers of biological self-organization, bridging spatial and temporal scales from the molecular level, through cells, tissues, organs, to whole organisms and groups or collectives of individuals. Importantly, the components of each layer possess problem solving capabilities for their particular domains, collectively implementing the foundation for the next higher level of organization while themselves relying on lower-level agential substrates. This implements a problem-solving layer between the genome and the phenotype of an organism [266]. (B) A proposed model in which evolution pivots the same strategies (and some of the same problem-solving mechanisms) to navigate different problem spaces. The components of each level organize to solve tasks in their own space, and systems evolved from navigating the metabolic, physiological, transcriptional, anatomical, up to traditional behavior in 3D space, and beyond [2]. (C) Schematic of multiscale embedding architectures – from cells to organisms, ecosystems, cities, and planetary-scale systems – illustrating that each level constructs and operates within its own embedding space while simultaneously shaping those of the levels above and below. Across scales, cognition can be understood as the invariant capacity to navigate and remap these embedding spaces in response to changing internal or external conditions. Images used with permission from references indicated, except as follows; (A) by Jeremy Guay of Peregrine Creative. (C) is an AI-generated schematic illustration created using ChatGPT (OpenAI) based on author-provided text prompts inspired by [267].

which to probe the primitive – or perhaps surprisingly sophisticated – minds of unconventional systems. And finally – it impacts a critical unmet need and knowledge gap: aspects of regenerative medicine, bioengineering, AI, and even fields such as ethics are being enriched by a novel understanding of the origin, scaling, and properties of unconventional embodied minds, such as cellular collectives, that can be coaxed to pursue alternative end-states [19–23].

This paper extends the Fields-Levin framework of cognition as competent navigation in arbitrary spaces [2,3] (see Fig. 2(A) and (B)) by proposing that intelligence is characterized by two equally important cognitive invariants: the creation and *remapping* of embedding spaces, and the *navigation* within these spaces in response to new contexts. More precisely, we define *remapping* in this paper as representing and/or updating new information acquired via experience into an internal latent representation in a way that not necessarily, but generically refines the navigable embedding space – a concept closely related to internal representations formed by world models [24–26]. We argue that embedding spaces, as implemented by organisms or other finite systems, are not pre-defined, static structures, but rather actively constructed representations that translate complex realities into navigable dimensions, as schematically illustrated by Fig. 2(C). By recognizing both processes as fundamental to cognition, we develop a more comprehensive substrate-independent framework for understanding intelligence across biological and artificial systems. We demonstrate the explanatory power of this dual framework through examples ranging from biological regeneration to AI systems, showing how the interplay between remapping (via embeddings) and navigation processes enables complex adaptive behaviors across diverse embodiments of distributed intelligences, agnostic to problem spaces, scale, or material composition. We begin by outlining navigation as an invariant of cognition, before turning to remapping, embeddings, and their manifestations in biology and AI. We conclude with

a perspective on how these principles will foster the development of novel adaptable synthetic and chimeric collective intelligences, potentially ranging from (sub-)cellular compounds to integrated planetary information networks.

2. Remapping and navigating arbitrary spaces as an invariant of cognition

2.1. Navigation in arbitrary spaces as an invariant of cognition

William James, the pioneering psychologist, offered a goal-directed definition of intelligence: “Intelligence is a fixed goal with variable means of achieving it” [12]. It remains particularly prescient in the modern era of AI, where populations of software and physical agents are expanding. This definition also emphasizes that intelligence manifests as goal-directed behavior rather than any specific mechanism or embodiment. The cybernetic tradition, developed by Ashby [11], Maturana and Varela [9], and Pattee [10], expanded this perspective by framing cognition in terms of information processing and self-regulation. The Free Energy Principle (FEP) generalizes these earlier efforts, abstracting the “fixed goal” of all time-persistent systems as that of maintaining their distinctness from their environments [27–29]. Distinctness from the environment is maintained, in the language of statistical physics, by maintaining a Markov blanket, a statistical boundary that enforces conditional independence between the internal states of the system and the internal states of its environment [30]. Friston and colleagues have shown [28,29] that any such bounded system can be characterized as a Bayesian satisficer that minimizes an upper bound – a variational free energy – on surprisal measured at its boundary. Intelligence, in this framework, manifests as *active inference*, a combination of learning, via updating expectations over environmental behaviors, and active modification of the environmental inputs, e.g. by manipulating the environment or by motion in 3D space, to make inputs more similar to expectations [31,32]. Significantly, this generalization of goal-directed behavior applies to *all* time-persistent physical systems, regardless of scale, structure, or complexity [33], and characterizes all systems compliant with Newton’s 3rd Law as active-inference agents.¹ It therefore provides a natural, albeit highly abstract, language for treating intelligence – or cognition – as a continuum, and hence for characterizing cognition in diverse embodiments independently of scale, structure, or complexity. General homeostatic feedback control is but one level of cybernetic and other competencies recognized by behavioral scientists; recent results in the field reveal that the use of approaches from cognitive science leads to novel discoveries that standard dynamical systems theory approaches had not reached [13,34].

More recently, Fields and Levin further expanded this approach and framed cognition as competent navigation in arbitrary spaces [2], as illustrated in Fig. 2(B). This framework, constructed using the formal framework of the FEP and active inference, proposes navigation in a problem space (metabolic, physiological, morphological, or behavioral for biological systems, for example) as an invariant for analyzing cognition in diverse embodiments [2,35]. More specifically, in this framework, a goal is a (target) position in a problem space, while a system’s degree of competency (or cognition, or intelligence, in the definition of William James) is reflected by the system’s ability to navigate that space toward a goal state by actively minimizing the deviation from diverse starting positions. Such a deliberately substrate-agnostic framework allows recognizing common patterns of intelligence across vastly different systems [36], from biological to artificial systems.

There are several examples of navigation in spaces in biology, including the metabolic, physiological, transcriptional, morphological or behavioral (3D) spaces. Biological systems are the result of multiscale goal-directed behavior (see Figs. 1 and 2(A)). These navigational capabilities are indeed not isolated but interconnected through hierarchical relationships that enable coordination across scales for reaching homeostasis [1,3]. A key insight of the Fields-Levin framework is how higher-level systems influence lower-level ones (see Fig. 2(B) and (C)). Higher-level systems deform the energy landscapes for their subsystems, allowing lower-level components to follow local gradients that achieve goals beneficial to the higher-level system [2]. This allows the scaling of goals and the development of higher cognition [214] and creates a situation where components can engage in relatively simple local behaviors while collectively achieving sophisticated system-level outcomes.

Intelligence in animals is typically measured by their capabilities to solve problems by navigating and manipulating objects, and resolving problems in 3D space [37–42]. However, it becomes increasingly obvious that certain (different) levels of agency and cognition need to be attributed to the components of all organizational layers across the multiscale competency architecture of biology [36] (see Fig. 2(A)). Even cells and cellular collectives solve problems when navigating transcriptomic space or morphospace – the space of possible anatomical configurations [1] – which leads to the immense structural and functional plasticity of living matter; selected examples are visualized in Fig. 1 and discussed below.

The planarian flatworm provides an interesting example of intelligent navigation in the transcriptional space [43]. Exposure to barium, a potassium channel blocker, causes head degeneration. Within the same specimen, planaria regenerate new heads with the transcriptional changes needed to achieve barium resistance [44,45]. How can groups of cells identify which of tens of thousands of potential transcriptional effectors will solve their physiological stressor? Moreover, barium is not something planaria encountered in the wild, suggesting cells possess general problem-solving capabilities, not only genetically pre-programmed responses to specific challenges [1]. The system senses its current position (a disrupted physiological state), compares it to a preferred position (normal function), and moves through transcriptional space by modifying gene expression until it reaches a new attractor region that restores homeostasis. There is not enough time to do this by trial and error, and some sort of heuristic must be used to identify solutions

¹ Note in particular that the idea of a system with “no active states” in Fig. 2 of Ref. [29] violates Newton’s 3rd Law and hence is an unphysical abstraction.

to novel problems in this (and other) spaces that life thrives in. Even below the cell level, gene regulatory networks are capable of various kinds of learning and memory [34,46].

During development and regeneration, cellular collectives navigate morphospace with high robustness, coordinating their activities to achieve specific target morphologies despite varying starting conditions or perturbations [1,13]. Salamanders can regenerate limbs from arbitrary amputation planes [47], cellular collectives in a tadpole's disrupted "Picasso face" reorganize to form normal frog facial structures [261], and planaria can regenerate heads with species-specific morphologies without genomic editing [44,48]. These processes may involve cellular collectives actively sensing their current morphological state, comparing it to the target morphology, and executing collective cellular decision-making to reduce the error, probably using biochemical, mechanical and bioelectrical cues [1]. The cellular collective constructs a distributed representation of both its current and target morphology, then navigates through morphospace using local cell behaviors that collectively produce global pattern formation [49]. The Fields-Levin framework defines these phenomena as navigation in morphospace toward specific attractor regions [2] (c.f., Fig. 2(B)).

This framework is not restricted to cells that form tissues or morphologies, but scales across all organizational layers of biology: distributed agents perceive and act locally, following their individual agenda that is aligned with, or constrained by, a collective higher-level objective. Moreover, navigation in a problem space is not limited to biological systems; Large Language Models (LLMs) navigate the language space [50,51], robots navigate motor spaces [52], multimodal models are navigating a merged space of different modalities [53], etc.

Navigation in arbitrary spaces, from morphogenesis to generative AI, reveals a common substrate-independent principle of cognition. Crucially, however, and as mentioned before, the state spaces as implemented by organisms or other finite systems are not the static structures of formal models – e.g. configuration spaces with well-defined topology, geometry, and dimension – but rather are continually constructed and reshaped by the systems that use them under novel contexts. Through this lens, remapping becomes inseparable from navigation, and together they constitute dual invariants of intelligence, as we turn to below.

2.2. Navigation implies remapping for learning systems

As we just saw, competent navigation in a problem space is a hallmark of cognition in the Fields-Levin framework [2,3]. However, any learning system must remap and update this problem space when it acquires new knowledge of the internal and external worlds. We believe remapping the problem space cannot be separated from competent navigation in this space to define cognition and intelligence. Remapping, according to the above definition – as representing and/or updating new information acquired via experience into internal representations that generically refine the embedding space – combines learning with experience-dependent modification of the state space in which learning takes place. Such remapping can be implemented in multiple ways, from expanding the dimensionality of the state space, as all multicellular organisms do as they undergo embryonic development or regeneration, to altering the accessibility of regions of a fixed state space by developing new search or navigation procedures, to transforming the state into an alternative representation that simplifies navigation.

In biological systems, remapping occurs at several levels. Neural plasticity, for example, represents a remapping of the brain's memories and representational space, where connections and weightings are reconfigured based on experience, effectively changing not only what is thought, but also the landscape of possible thought patterns [54,55]. Specific neural structures do the same, for example, the hippocampus demonstrates this kind of remapping for navigation in different spaces, creating new internal spatial representations when animals encounter novel environments [56]. But remapping is not only limited to the brain – the adaptive immune system actively remaps its recognition spaces using VDJ recombination (variable-diversity-joining rearrangement), creating new antibody configurations that represent previously unencountered antigens [57,58]. The already described planarian's transcriptional adaptation to barium demonstrates remapping of its internal transcriptional landscape to represent and respond to this novel environmental chemical [44]. At cellular levels, stem cells while differentiating are remapping completely their transcriptional landscape too [59]. Even molecular networks can learn and have different types of memory, actively updating and remapping their internal representations of the learned information [34,46].

At longer timescales, the same principles apply more generally to evolutionary developmental biology: Through a series of major evolutionary transitions – or remappings, in the present context – life persistently opens up new navigable spaces, leading from pre-biotic chemistry toward genomic encoding, from replicators toward chromosomes, from prokaryotes toward eukaryotes, from single cells toward multicellular organisms, from solitary organisms toward societies, and from animal communication toward language, culture, and, eventually, technological civilization [60,61]. Evolutionary dynamics have been fruitfully described not only as connectionist learning processes, but also in terms of associative memory and, more recently, natural induction [62–69]. Across all these scales, evolution is tinkering [70–72]: it builds increasingly sophisticated problem-solving machines by repurposing, recombining, and integrating previously acquired competencies under ever-new contexts. Moving beyond the simplified blueprint metaphor or the idea of a mere indirect encoding of phenotypic traits, this reframes the genome as a functional bottleneck of lineage memory: a compact set of latent variables, shaped by evolution, that cells and tissues interactively decode (and reshape) during organismal growth, repair, and regeneration [73–75]. Development is thus the context-sensitive navigation of this generative latent space under the conditions of a particular organism and environment. Evolution remaps what development can make possible; development navigates and further reshapes those possibilities; and organisms over their lifetimes, through learning, taking actions, niche construction, symbioses, and reproduction, reshape both the physical and developmental spaces inherited by future generations. Seen through this lens, evolution is a scaling process of compositional problem-solving components that operate across genotypic, physiological, developmental, phenotypic, or ecological domains [2,3,36]. In turn, this suggests that Life itself can be viewed as an extended yet

integrated morphogenetic process, i.e., potentially as a single decentralized agent whose future forms and functions arise through the continual remapping and navigation of its own possibilities, while remaining bound to physical constraints [76].

At the organismal level, sensorimotor systems provide particularly clear examples of remapping via the creation of internal representations by building neural maps. These maps are found ubiquitously in the brain, where they encode a wide range of behaviorally relevant features into neural space. Neural maps in the brain exhibit different organizational patterns. Continuous maps maintain spatial relationships where neighboring neurons encode similar stimulus properties. For example, retinotopic maps in visual processing areas preserve spatial relationships from the retina [77], and neurons in tonotopic maps are organized following sound frequency gradients, with neighboring cells responding to similar frequencies [78]. Discrete maps, by contrast, create distinct functional modules without gradual transitions. The olfactory system is one of them: olfactory receptors located throughout the nasal cavity connect to specific neural clusters called glomeruli in the olfactory bulb, with each receptor type linking to its own dedicated glomerular group [79]. Many brain regions combine both organizational strategies, for example somatosensory areas processing touch and regions controlling movement typically integrate continuous gradients with discrete frontiers, creating hybrid mapping systems that balance precision with flexibility within two-dimensional cortical surfaces [80]. Critical period plasticity is also a remapping, where sensory maps are reconstructed to represent altered input patterns during development [81].

The evolutionary perspective reveals remapping as fundamental to adaptation. In fact, as our senses have evolved, we have seen new maps for translating external energy – such as light, sound, and chemicals – into internal neural representations. The development of bat echolocation, for instance, necessitated remapping auditory processing in order to produce intricate internal spatial maps from external acoustic reflections [82]. These sensory maps are continually being updated and revised. Primary sensory cortices constantly remap the organization of their receptive fields to reflect changes in the environment [83]. Cross-modal plasticity after sensory loss shows a drastic remapping. For example, in blind subjects, the visual cortex literally represents external auditory information in the previously visual neural cortices by creating new internal maps of auditory space in these cortical areas [84]. When learning new primitives and movement pattern sequences, or when internalizing external tool representations, the motor system also exhibits remapping. Indeed, the motor cortex develops new maps that represent the relationship between intended actions and their effects in the external world [85,86]. Maybe even more remarkable, brain-machine interfaces demonstrate remapping in real-time, as motor cortex represents external device control, mapping internal neural activity onto external robotic movements [87]. A cognitive example of remapping is provided by verbal fluency tasks like the Addenbrooke's Cognitive Examination, which includes a verbal fluency domain and commonly contrasts letter-cued retrieval with category-cued retrieval [88]. In these tasks, participants search through largely the same lexical repertoire, but the effective geometry of the problem space changes with the task demand: a letter cue privileges phonological relations, whereas a category cue privileges semantic proximity. This makes verbal fluency a compact example of cognition as navigation through a linguistic embedding under different remappings of relevance. Recent computational work supports this view by showing that verbal fluency trajectories are shaped by semantic similarity, differ across category and letter conditions, and are altered in schizophrenia [89]. Remapping is not limited to sensory maps but it can also build salience maps, it updates the coordinates and constraints under which navigation proceeds, and these constraints can be specified by salience in the sense of expected information gain, epistemic value, curiosity- and novelty-driven sampling [90–93].

Just as biological systems construct and revise internal maps to maintain homeostasis and adapt to novel challenges, AI systems build and remap data into structured spaces, via embeddings. These parallels suggest that remapping is a general strategy for intelligence, irrespective of substrate. We now turn to AI systems to illustrate this convergence. LLMs for instance, generate embeddings on a vector space that are constantly updated with new data (c.f., Fig. 3). These embeddings are vector representations of tokens, and distances and directions between them, that represent semantic relationships in the text [94,95]. The effective dimensionality of the embedding space increases whenever the LLM learns a new word. Computer vision systems transform visual scenes into hierarchical maps that show external objects together with their textures and spatial connections [96]. The internal policy maps of reinforcement learning systems demonstrate how external states connect to optimal actions [97,98]. Transformers' attention algorithms generate internal maps of token relevance which represent the external context structure that matters for each processing step [99]. Generative models in AI transform external data distributions into internal latent spaces to create compressed representations of external patterns [100,278]. Robots produce actual navigation maps when they explore new environments [101]. Few-shot learning models exhibit fast remapping capabilities through their ability to develop new internal representations that connect new classes to existing representational spaces [102]. Meta-learning systems possess advanced remapping abilities which allow them to establish new internal maps for external tasks by choosing appropriate learning methods [103].

Through remapping at all levels, learning systems can discover new representational schemes that capture previously unrecognized external patterns, create internal maps that efficiently compress external complexity, and integrate information from heterogeneous external sources into unified internal maps/spaces [104–108]. By recognizing remapping as the fundamental process of competently navigating a problem space, we develop a more precise substrate-independent framework for understanding intelligence across biological and artificial systems [1,2].

Conversely, failures of remapping help define the boundary conditions of the framework: when remapping is absent, limited, or disrupted, systems should show reduced learning, developmental constraints on adaptive reorganization, predictive errors, and greater rigidity under changing conditions. In domain-specific cases, this may also include developmental defects like developmental topographical disorientation [109], disorders of active inference by an inability to map accordingly environmental information [110] or impairments of spatial updating across saccades, as discussed in unilateral neglect [111]. These cases are important not only as counterexamples, but as informative limits of its explanatory scope, since they reveal what happens when internal spaces cannot be remapped efficiently in response to experience.

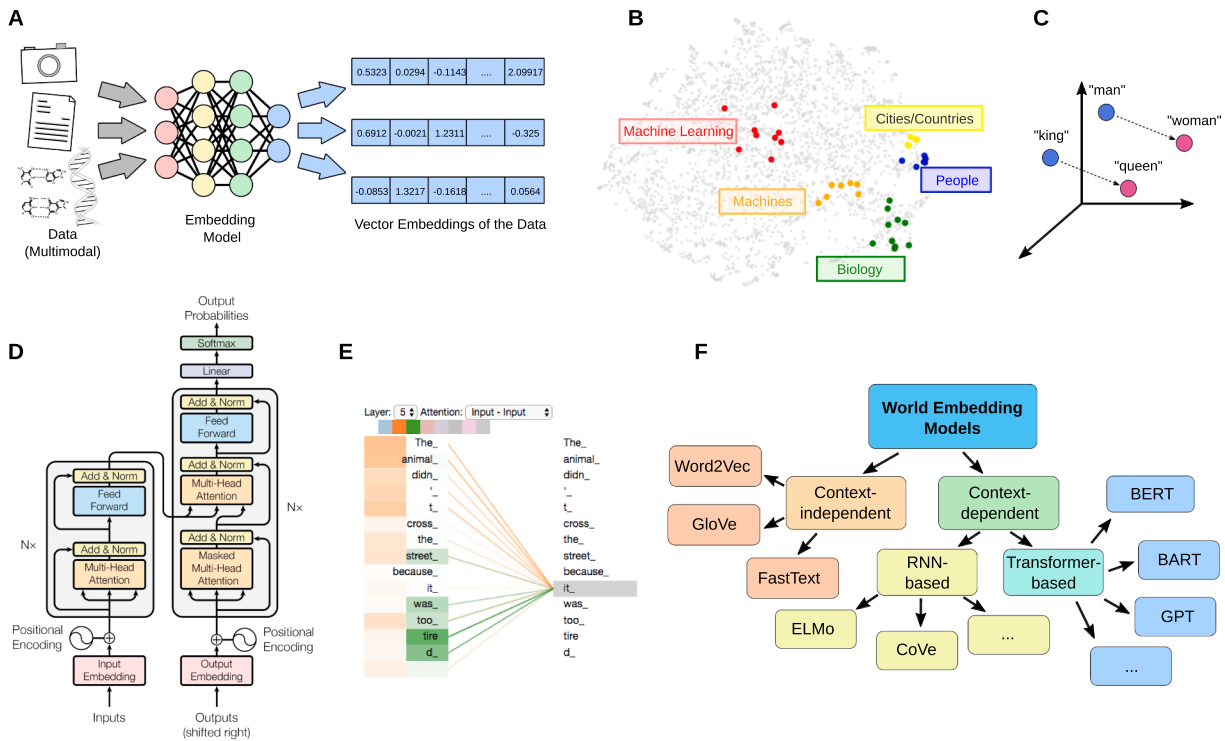


Fig. 3. Data-embeddings and transformers: (A) Embedding models transform multimodal data into standardized numerical representations, i.e., low-dimensional vectors in an abstract embedding space (compared to the high-dimensional unstructured data), where directions encode semantic relations. (B) Low-dimensional projection of word-embedding vectors of the content of the present manuscript, showing clusters of semantically related words (color-coded). (C) Geometric relations in embedding spaces capture semantic analogies, e.g., the vector offset “man” → “woman” parallels “king” → “queen” in Word2Vec space [122]. (D) The transformer architecture forms the backbone of modern language and vision models. Transformers interpret their stream of input data (i.e., the context) by computing attention scores between all input tokens (e.g., word or image-patch embeddings), allowing the model to weigh the tokens’ relative importance. Image adapted from [99]. (E) Visualization of the attention mechanism in action during next-token prediction, illustrating how a model distributes attention across tokens when predicting the next word. As the model encodes the word “it”, different attention heads focus on different parts of the context – for example, one head attends primarily to “the animal” while another attends to “tired”. Thus, the model’s representation of “it” incorporates information from both semantic sources, illustrating how attention mixes contextual representations in associative embedding space [128] (image adapted from [268]). (F) An overview of word embedding models. Word embedding approaches can be grouped into context-independent and context-dependent representations. Context-independent models assign each word a single fixed vector regardless of usage (e.g., Word2Vec [122], GloVe [269], FastText [270]). Context-dependent models generate token embeddings that vary with linguistic context, either through recurrent neural networks (ELMo [271], CoVe [272], etc.) or transformer architectures (BERT [273], BART [274], GPT [275], etc.) – and thereby marked a paradigm-shift in embedding models. Arrows indicate conceptual or historical advances. The overview is adapted from [276].

3. A common mechanism between AI and biological systems: remapping via embeddings

3.1. Computability constrains remapping

The kinds of remapping observed in biological systems and employed in AI systems are not arbitrary, but rather appear tightly controlled. What constrains remappings, and how is this constraint information encoded by the system doing the remapping? To address this question, suppose X is a data set, Γ is a latent space, and $\epsilon : X \rightarrow \Gamma$ is an embedding of the data X into the state space Γ . Suppose also a map $\xi : \Gamma \hookrightarrow \Xi$ that transforms Γ into an alternative representation Ξ . The composition $\xi\epsilon$ is, by our definition, a remapping. Clearly, however, the remapping is coherent only if the data X can also be embedded in Ξ , via a map $\epsilon' : X \rightarrow \Xi$, such that $\epsilon' = \xi\epsilon$, i.e. only if the diagram:

$$\begin{array}{ccc} \Gamma & \xrightarrow{\xi} & \Xi \\ \epsilon \swarrow & & \searrow \epsilon' \\ & X & \end{array}$$

(1)

commutes. Diagram (1) will commute only if the map ξ preserves the topological, geometric, symmetry, and other relational properties that are implicit in the data X , as well as all other data that are encoded in Γ .

To further understand this commutativity constraint, it is useful to generalize Diagram (1) by replacing the implicit Identity map $\text{Id} : X \rightarrow X$ with an arbitrary computable function $f : X \rightarrow X'$:

$$\begin{array}{ccc} \Gamma & \xrightarrow{\xi} & \Xi \\ \epsilon \uparrow & f & \uparrow \epsilon' \\ X & \xrightarrow{\quad} & X' \end{array} \quad (2)$$

Requiring that Diagram (2) commutes defines the map ξ as an *implementation* of f on the structure Γ [112]. Coherent remappings are, therefore, implemented computations on the embedded data; specifically, they are implemented computations of the identity map $\text{Id} : X \rightarrow X$ for all of the embedded data X .

A particularly useful computation on embedded data is coarse-graining, which can enforce strong constraints without representing them explicitly. Mapping a high-dimensional parent space Γ into a lower-dimensional latent space Ξ via a structure-preserving map $\xi : \Gamma \hookrightarrow \Xi$ selects parent-space structures or relations that are inverse images, under ξ , of coarse-grained structures or relations in the smaller latent space (c.f., schematic illustration in Fig. 2(B) and (C)). We first examine remappings of biological systems into 3D space, then consider more general remappings available to AI systems.

3.2. Embedding in 3D space as a remapping constraint

Biological systems continually traverse metabolic and gene-expression (i.e. transcriptomic and proteomic) spaces, and in doing so, sometimes also explore morphological space. These spaces can be very high-dimensional, involving tens to hundreds of thousands of chemical potentials, each with dynamic ranges of at least an order of magnitude. The vast majority of states in these biochemical state spaces are not just far from homeostatic ranges but inconsistent with life, e.g. states in which most proteins are denatured or states in which essential enzymes are deficient. A latent space specific to living systems, therefore, must encode a set of constraining relations that appears to be combinatorially explosive. If we knew this latent space and the embedding map to it, biochemistry would be finished.

Nature does not, however, have to know either the latent space or the embedding. Nature embeds biochemistry in 3D space, and selects for embeddings from chemical-potential space to bounded volumes in 3D space by selecting for organisms with functioning biochemistry. Almost all embeddings do not yield functioning organisms, but some do. The embeddings that work implicitly encode constraints on molecular relations that are consistent with life. Some of these constraints are extremely tight, forcing some pairs of molecules to be and remain in close proximity in 3D space. Others are much looser, allowing molecules to diffuse through the highly-structured 3D space specified by the tight constraints. We, as biochemists, try to figure out those implicit constraints by doing experiments on the systems that we observe, i.e. those that “work” in 3D.

Embedding biochemistry in 3D converts navigation in chemical-potential space into molecular behavior in 3D space, with the constraints imposed by the remapping manifested as behavioral constraints on the embedded molecules. Protein folding, for example, is highly constrained by ionic and van der Waals interactions that are defined in 3D space, not in the parent concentration space. Binding pockets, hydrophobic interactions, and diffusion are, similarly, spatial concepts that express navigational biases when pulled back to the parent space.

Biochemical and other constraints can be implicit in 3D space because 3D space is itself the outcome of a remapping. Classically, a dynamical system comprising N indivisible particles is naturally described in a $6N$ -dimensional configuration space, a space with ca. Avogadro’s number of dimensions for any macroscopic system. If these N particles have distinct properties, e.g. the atomic properties that chemically distinguish C, N, and O, the labels needed to distinguish these properties add dimensions to the configuration space. If the particles combine in different ways to generate more complex systems – e.g., molecules or cells – at different scales, the labels that specify these mereological relations also add dimensionality to the configuration space. Gauge symmetries, e.g. rotational and translational symmetries, encode redundancy at more microscopic scales, but may be broken, functionally, at larger scales, e.g. when separating two molecules in space effectively prevents their interaction. Remapping this high-dimensional complexity to 3D space not only compresses $6N$ dimensions into three, but also compresses gauge, mereological, and other symmetry constraints into “laws” of physics, chemistry, biochemistry, or higher-level sciences that may, in many cases, not yet be formulated or even recognized. Treating 3D space as a “natural” embedding target for biochemistry presupposes this massive reduction in dimensionality.

The idea of remappings as imposing constraints can be made more precise in the language of sheaf theory. Sheaves are a category-theoretic description of what is common to, for example, vector fields, maps or geographic information systems, databases, and other coherent ways of “attaching” structured data to each point or “location” in some space [113,114]; see Supplementary Information for formal definitions. Coherence is enforced by requiring, roughly speaking, that assignments of data to different regions agree wherever the regions overlap, and that any way of building up the data for a region by assigning data to subregions agrees with every other way of building up the data for that region by assigning data to subregions. Letting the “space” to which data are attached be 3D space extended along a time axis, i.e. a 4D spacetime, a coherent embedding of biochemistry is one in which every state and state change can be implemented by a time-evolving 3D structure in a way that respects all local 3D constraints, e.g. the structures of other nearby molecules. Note that coherence in this sense is a necessary, but not sufficient, criterion for viability. Both the biochemistry of live cells and that of lysed, dying cells must embed in 4D.

Probability distributions are assignments of data – probabilities – to underlying data sets or spaces. Violations of the sheaf coherence conditions correspond to violations of the Kolmogorov axioms, e.g. local additivity of probabilities. Abramsky and Branderburger [115] have shown that data violating the sheaf coherence conditions can be viewed as outcomes of measurements, or in general, processes exhibiting noncausal context dependence, e.g. processes that exhibit quantum interference. In the current context, cellular

processes that mutually interfere will appear “non-local” in a 4D embedding, and will be disrupted (or “decohered” in quantum-theoretic language) by measurements or other processes that enforce locality [116].

3.3. Multi-scale embeddings and coherence conditions

The foregoing says nothing about the scale of the embedding; different 4D embedding scales for biochemistry, for example, capture molecular dynamics at different resolutions [117]. Embeddings of the same data at different resolutions are coherent if, but only if, the higher-resolution embeddings coarse-grain to the lower resolution embeddings via some function that does not depend on the embedding-space coordinates. High-resolution constraints that do not “smear out” at lower resolution may appear as “emergent structures” or “emergent boundary conditions” in the lower-resolution description [76]; for example, networks of interactions that, at high resolution, limit free diffusion may appear as domain walls or mesoscale structures at low resolution.

In the context of the Fields-Levin model [2], spatial embeddings may extend beyond the cellular scale to those of tissues, complex multicellular bodies, or communities of interacting organisms. These scale changes introduce “natural” spatial boundaries, e.g. the boundary of a tissue or an individual organism, that function as emergent boundary conditions on macroscopic behavior. One effect of these boundary conditions is to enforce spatial symmetries by enforcing spatial coupling between micro- to mesoscale systems and their local environments; organismal boundaries, for example, keep the parts of an organism together by assuring that they are transported through space together. Processes that would not be translationally invariant without such boundaries – e.g. a gene regulatory network (GRN) moved from inside an organism to the ambient inter-organism environment – become effectively translationally invariant because their local environments move with them. These boundary-imposed symmetries render “space” and hence the spatial embedding functionally meaningful at larger scales.

It is this preservation of effective spatial symmetries by emergent boundaries that allow processes as well as structures to be spatially embedded. This is true not just of biochemical, e.g. metabolic or signal transduction, bioelectric, and mechanical processes, but also of cognitive processes. Active inference [27], for example, can only be maintained in a reasonably familiar and hence reasonably predictable environment; completely unfamiliar environments – e.g. environments involving radically different chemical potentials, electromagnetic fields, particle fluxes, or mechanical forces – tend to induce failures of adaptive behavior, disruption of the system-environment boundary, and loss of the conditional independence of internal from external states that support system persistence over time. Exposure to completely unfamiliar and hence unpredictable environments, in other words, often lead to systemic dysregulation and death. From this perspective, the usual inferences from cognition to niche construction and from more complex, e.g. human cognition to (often by analogy) less complex or basal cognition appear both counter-intuitive and unmotivated. Small, low-complexity environments are both more probable and easier to maintain than large, high-complexity environments, and relative environmental stability is as essential for cognition as it is for metabolism. The former, like the latter, can be expected both to start out and to be widespread in small, relatively low-complexity systems coupled tightly to and spatially transported with small, relatively low-complexity environments, the maintenance of which becomes the primary objective of active inference by the systems inhabiting them. These conditions are met at the scales of functional networks within cells, including GRNs, single cells including prokaryotes, and communities such as microbial mats. Coherent spatial embedding, and hence the possibility of active inference, starts small and simple, and expands to larger scales as larger-scale spatial constraints emerge.

The principles observed in biological systems – navigation of attractor landscapes, local error correction that scales to system-level outcomes, and nested embeddings that support higher-level organization – exemplify scale-free cognitive dynamics where local competition and coordination give rise to the abstraction of intelligence across scales. These principles are not unique to living tissue but also apply to artificial systems, as discussed in the following.

3.4. Remapping and navigation of embedding spaces in modern AI

Above, we discussed fundamental organizational principles in biological systems based on dynamic, context-sensitive remapping of embedding representations. AI is undergoing a similar evolution. Traditional machine learning models, such as classifiers or auto-encoders, often acquire fixed feature vectors and latent-space representations representable as data manifolds [96]. These learned representations performed well for narrowly defined tasks but lacked the flexibility to adapt to new situations in broader contexts and across multimodal data.

Instead of relying on static representations, modern AI increasingly relies on generative models [118] – models that learn the underlying distribution of an often vast amount of training data – to construct dynamic and context-sensitive embedding spaces. So-called foundational models trained on “internet-scale” data can reach remarkable generalization capabilities [119,120] (but may lack the precision required for narrow tasks due to “over-generalization” [121]). Here, an embedding is a mapping from complex, often high-dimensional unstructured data into a structured numerical vector space – typically still high-dimensional – that preserves similarities and emphasizes semantic relations in the remapped representation. For example, words with similar meanings are mapped closely together, and associated sequences of words are often parallel trajectories in the embedding spaces of language models [122] (see Fig. 3). As clarified below, we use *embedding space* in a broad relational sense: its effective geometry may be continuous, discrete, graph-like, hybrid, or dynamically remapped; what matters here is that relations among states are organized so that error correction, navigation, and remapping become possible.

In transformer architectures [99], this is realized by an attention mechanism [123]. Each attention layer computes how tokens relate to one another in the current context (the current text input sequence) and updates, i.e. remaps, their embeddings accordingly. Crucially, this remapping of token embeddings updates the input token sequence during the pass through the model, i.e., it

does not replace the latter: Across a hierarchy of attention layers, token embeddings are thus remapped – or refined – incrementally to create increasingly contextually enriched token representations for a given input sequence. Moreover, transformers maintain several attention heads. Those are sensitive to particular modalities in the data and can thus be thought of as filtering the embedding space in parallel for meaningful relationships across certain semantic features, much like how neural circuits map multimodal sensory inputs. This allows transformers to reason about relationships and contexts through navigation in dynamically remapped embedding spaces. Strikingly, these transformer architectures can also be applied to the vision domain [124], and the learned vision embeddings can be repurposed – as argued by [125] for a shared embedding geometry – to form powerful multimodal embedding models [126].

Recently [127,128], the self-attention mechanism in transformers has been shown to be mathematically equivalent to a form of continuous Hopfield network. The latter can store a large – potentially exponential – variety of data patterns as attractor states in a high-dimensional energy landscape. Upon noisy or even novel inputs, Hopfield networks retrieve the closest stored pattern and thus act as associative memories [129] of their training data. Notably, this happens via internal energy minimization dynamics, i.e., equilibration into attractor states in high-dimensional associative embedding spaces. On the one hand this implements a form of pattern-level (global) error correction. On the other hand, the hierarchical attention layers act as a kind of iterative associative memory retrieval under a given context, selecting and refining information that best matches the input query among stored patterns.

The related concept of hyperdimensional computing [130] (HDC) is built on the principle that information can be stored and manipulated in extremely high-dimensional spaces – often 10,000+ dimensions – where different concepts are robustly represented as orthogonal vectors across all coordinates. In particular, HDC utilizes algebraic operations to superimpose elemental concepts to construct composite representations (e.g., the vector for a *red car* $\approx$ *red* $\oplus$ *car*), and even allows the storage of entire sequences as single vectors [131]. Because different features are spread redundantly, or “holographically” across the full vector space, superposed memories can be recovered even when a large fraction of the dimensions are corrupted. This renders HDC highly robust and akin to associative memories. Embedding models, through their high-dimensional embedding representations and attention mechanisms, may implicitly exploit similar principles: Attention layers map and combine sequences of features onto high-dimensional vectors on the fly, enabling such models to retrieve relevant patterns iteratively – layer by layer – during inference and capture context-dependent associations.

While transformers capture intricate contextual relationships through their stacked attention layers, and although their context is recursively extended by appending tokens (next token predictions), their architecture remains fixed and strictly feedforward. In turn, their internal representations are generated *de novo* at every inference step. This might fundamentally affect their reasoning performance through limited cognitive depth – a matter that is fiercely debated by the community [132–136]. Emerging research explores recurrent embedding architectures that explicitly utilize internal states over time: RWKV [137], Mamba [138], xLSTM [139], or recurrent-depth (so-called Looped Transformer) architectures [140–142], to name but a few, maintain internal embedding states that capture and balance short- and long-term temporal dependencies with attention-like mechanisms. These models show potentially on-par or superior performance at a fraction of computational costs compared to transformer architectures.

Even more ambitiously, world model architectures [24,143–145] aim to build models that understand the consequences of their actions in an environment. Such world model architectures maintain internal states of their environment, and actively try to minimize deviations between how their internal model of the world (expectations) and the actual perceived environment evolve, given their actions. This requires these models to build compact internal representations that capture essential features across scales of unstructured spatial and temporal input [143]. Such a mechanism can potentially be model-agnostic and enables planning and imagination rather than just reactive processing, potentially improving learning performance through “in-dream” training in several iterations of the model’s own imagination [24]. These models represent a next step in AI, shifting from static remapping of data to constructing internal simulations of compressed representations, or embeddings, that guide future actions. This increasingly aligns such AI systems with active inference frameworks (see below, [27,146]) and biological cognitive agents [147].

In general, identifying well-structured embedding spaces is critical. It has recently been shown [125] that sufficiently large embedding models that are trained on a substantially general data corpus all identify virtually the same uniform embedding-space geometry, which is not only relevant for interfacing multimodal embedding architectures, but also raises fascinating questions about the Platonic representation hypothesis [148,149]. For smaller everyday models, representations learned via open-ended neuroevolution paradigms show superior embedding representations as opposed to gradient-based methods [150]: despite both methods resulting in similar overall task performance and training dynamics [151], they often fundamentally differ in their internal data representations. Without particular regularization techniques, gradient-based methods often suffer from disorganized “entangled” embedding structures, termed fractured entangled representation (FER) [150]. In contrast, neuroevolution techniques biased towards open-ended search tend to find modular “disentangled” representations that generalize well to further downstream tasks. Identifying maximally informative, disentangled, conceptually orthogonal embeddings in a systematic way is thus invaluable to advance representation learning. Promising self-supervised techniques like VICReg [152] or JEPA [144,145] impose constraints on variance, invariance, and conceptual decorrelation by whitening covariance across latent variables, preventing collapse into trivial solutions while preserving meaningful relationships between latent space dimensions.

Across various substrates, such as cells, brains, or AI architectures, embeddings provide representational, context-sensitive abstractions, while self-regulatory error correction drives the system’s dynamics. Intelligence does not emerge from recombining static maps but from iterative cycles of prediction, correction, and refinement within contextually (dynamically) remapped embeddings. We argue below that iterative error correction represents an operational backbone linking navigation and remapping.

4. A pervasive error-correction scheme in AI and biological systems

From cells to human minds, adaptive agents continuously predict the state of their environment against perceptions and adjust their internal models or their actions to reduce prediction errors [1,2,6,27,110,153]. This prediction–perception–error-correction loop represents a robust navigation policy in arbitrary physical, cognitive, or virtual embedding spaces [2,3], and has started to gain traction in machine learning research; examples are transformers incrementally refining streams of tokens using stacks of attention heads, or world models updating their internal representations based on new evidence.

While iterative self-regulatory error-correction provides scale-free dynamics for cognitive processes [3], remapping representations under novel contexts adds plasticity to an agent's internal model, effectively facilitating creativity and learning [2,3,6,27,36,146,154]. From a practical point of view, we suggest that self-regulatory error-correction in observer-specific plastic embedding spaces is a unifying operational principle for biological and artificial intelligence.

As noted earlier, the FEP formalizes self-regulatory error correction as a criterion that must be satisfied by any physical system that maintains its distinction from its environment over any significant time period [28,29]. Active inference is, in FEP language, the enactment of self-regulatory error correction [31,32]; its goal, in every case, is to minimize surprise induced by unpredictable environmental fluctuations in order to preserve boundedness, and hence statistical separation from the environment. Whether at the scale of a single cell responding to morphogen gradients, or an organism planning future actions, systems evolve internal models that capture the latent structure of the actions of their environments on their boundaries. In the space defined by system-environment information flows, these boundaries are Markov blankets [30]; embedding this information-flow space into 3D typically maps such boundaries into meso- to macroscale emergent structures, e.g., cell membranes or the skins of multicellular organisms, that spatially separate internal from external degrees of freedom and processes. Complex systems comprise nested boundaries that facilitate localized prediction and correction processes, with each component system having to cope with only a local, spatially co-transported environment as discussed above. Crucially, in the multiscale competency architecture of biology [36] these boundaries are not static but dynamic and possibly fractal, scaling from subcellular compartments to entire organisms and possibly beyond [154,155].

Modern AI approaches have adopted ideas from active inference in several ways. World models [24,143,146], for instance, embody this principle probably most explicitly by minimizing prediction error between perceptions and a self-maintained internal state predictions through belief-state updates and/or taking environmental actions. Other related developments have managed to train large language models (LLMs) to display advanced reasoning solely through self-supervised learning [156]: a completely intrinsic reward mechanism maximizing the model's confidence in its outputs appears to surpass customized external reward schemes.

Another recent promising trend in LLM-based reasoning systems integrates pretrained multimodal foundation models with external memory systems, tool-use interfaces, and autonomous control frameworks to build LLM agents capable of actively navigating custom problem domains [157,158]. These systems dynamically contrast their internal task representations – transforming a text-prompt specifying the problem into learned internal embedding representations – with the unfolding state of an external world (e.g., accumulated experimental results, retrieved knowledge, or tool outputs) through mechanisms such as retrieval-augmented generation [159,160], tool-augmented action loops [161,162], and self-reflective or self-corrective reasoning procedures [136,156,163]. In this sense, emerging agentic LLM architectures operationalize the same twofold invariants we describe in this contribution: (i) remapping their embedding spaces as new context arrives – via attention-based latent updates, external-memory retrieval, and world model-like state revisions – and (ii) navigating these evolving spaces through iterative error minimization, informed by planning, and action selection [132,164,165]. These systems increasingly resemble multiscale cognitive agents that refine internal latent spaces during inference, a property shared with biological navigation in embedding spaces. LLMs thus represent an intriguing architecture that allows us to query the collective knowledge on the internet with incredible efficiency. The challenge might have shifted towards integrating those models into multi-scale architectures that can actively, and faithfully learn to understand and navigate relevant aspects of the world (see Fig. 4 for an illustration). A thorough understanding of multimodal embedding space geometries [125], and learning paradigms that lead to compositional [150] or predictive multimodal (joint) embeddings [143] are promising directions for advancing the field of world model like architectures [24,146] toward bio-inspired multiscale problem solving systems [36,74].

Diffusion models [166–170], on the other hand, are perhaps the clearest implementation of generative models that are intrinsically based on self-regulatory error-correction dynamics during both training and inference. Rather than generating data in one deterministic pass, diffusion models are specifically trained to denoise potentially corrupted data: while a forward process successively corrupts data by incrementally adding noise to target data points, these models learn the reverse process of correcting noisy data samples. In that way, diffusion models not only learn to counter potential entropic – or diffusive – effects to restore relevant information, but represent powerful generative models that transform initial random noise into novel structured data samples that conform to the training data distribution. The generative process of diffusion models thus implements iterative self-regulatory error-correction dynamics, where each denoising step performs an incremental correction on noisy data samples, gradually reconstructing complex structure from initially high-entropy states.

This simple principle is the basis for cutting-edge image [171] or video generation techniques [172], and has also found application in generative game-play through world modeling [173], as backbone in optimization methods [174–177], but also in biomedically relevant applications such as *in-silico* protein-folding [178] or protein-synthesis [179] and as foundational models for genomics [180–182], transcriptomics [183–187], etc. [188].

Interestingly, this process can also be interpreted as an incremental teleologic process [176]: while the system's dynamics executes small corrective steps (by design), the proposed state-corrections of the model are always with respect to a goal state represented by attractor states in the model's representation of the data manifold (embedding space) that are most similar to the current data pattern. Strikingly, this positions diffusion models as associative memories [127,189] which undergo successive steps of symmetry breaking

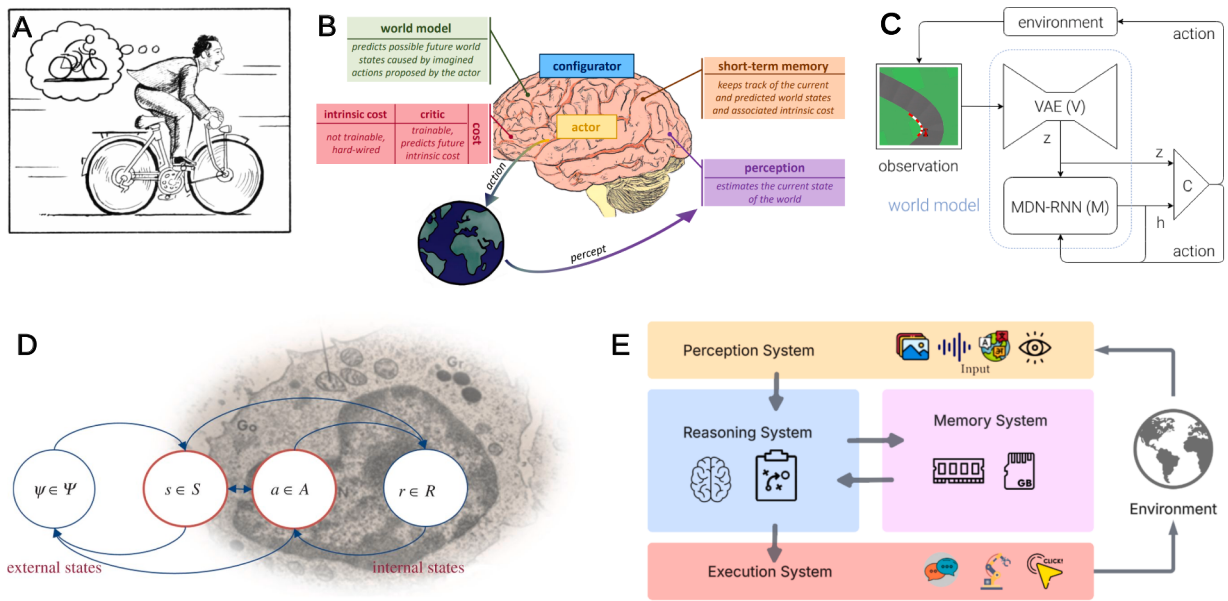
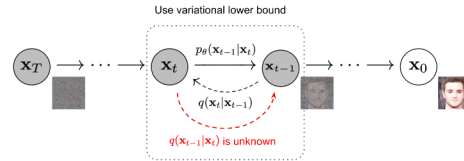


Fig. 4. World models and agentic AI navigate the world via internal models. (A) World model illustration of a cyclist from Scott McCloud's Understanding Comics [277]; inspired by [24]. (B) A more novel and detailed illustration of the components of an agential world model architecture that learns to navigate a Joint-Embedding Predictive Architecture (JEPA); adapted from [143]. (C) A possible (differentiable) world model implementation, comprising a perception encoder – via a Variational Auto Encoder (VAE) [278] – whose compressed latent representation z is constantly contrasted to an internally predicted perception – i.e., a predicted state of the world, generated by a Mixture Density Network (MDN) with a Recurrent Neural Network (RNN) backbone; both the compressed perception z and internal (hidden) state h of the MDN-RNN inform a controller module C that outputs an action in a complex external environment, such as a car-racing track; adapted from [24]. (D) A schematic illustration of an active inference agent, partitioning the world into internal states and hidden or external states that are separated by a Markov blanket, comprising sensory and active states. Perception corresponds to the self-organization of internal states, while action couples internal states back to external states; adapted from [195] and adapted from [2]. Through this perspective, active inference closely parallels world model architectures [146]. (E) A relatively novel trend in large language model (LLM)-based reasoning systems integrates pretrained multimodal foundation models with external memory systems and tools for building “Autonomous LLM Agents” that can navigate custom problems actively by contrasting their understanding of the task (e.g., specified loosely via text prompts) to the context of an external world (such as accumulated experimental outcome and knowledge about the problem); adapted from [279]. All of these architectures contrast an internalized representation, such as an embedding space of a (their) world to form decisions about optimal navigation policies in their environment, and/or remapping their internal representations based on novel evidence to execute better-informed decisions or actions in the future.

and phase transitions during their generative phase [190–192]. During inference, diffusion models thus implement entropy-reducing generative trajectories that follow a hierarchy of self-regulated phase transitions through a learned latent space of a data manifold. We emphasize that this not only resembles but actually models biological processes in organismal development and cognition, based on self-regulatory error-correction principles in dynamic embedding spaces [74] (see Fig. 5). This realization has allowed us to relate diffusion models with the process of evolution [176,177] and morphogenesis [193], explicitly framing the genome as a functional unit and literally as a generative model [6,73,74] (see Fig. 6).

A concrete computational example of this idea is provided by our Conditional, Heuristically-Adaptive ReguLarized Evolutionary Strategy through Diffusion (CHARLES-D) [177]. Here, a diffusion model is trained online on evolutionary history, including genotypic parameters, fitness values, and additional traits defined in genotypic, phenotypic, population-level, or fitness-related spaces. The trained model then functions as a generative reproductive operator: instead of sampling offspring only through mutation and recombination, it actively denoises randomized proto-genotypes toward regions associated with high-quality solutions. With classifier-free guidance [171,194], this reproductive process can be biased toward specified target traits without changing the underlying fitness function. Newly sampled offspring are then evaluated in their environment, adding both genetic material and phenotypic evaluations back into evolutionary history. In the language of the present framework, CHARLES-D operationalizes evolution as a coupled remapping-and-navigation process: contrasting to evolutionary methods that use pre-trained diffusion models for black-box optimization [174,175], remapping occurs when the diffusion model's generative process is successively refined on past and novel evolutionary experience across generations; navigation occurs during reverse diffusion, as candidate offspring follow the learned score field toward attractor regions corresponding to high-quality solutions that satisfy trait-specific conditions. CHARLES-D therefore illustrates how a generative reproductive process can bridge the problem-solving layer between genetic encodings and targeted phenotypic behavior, linking genotypic variation to target traits through iterative error correction.

A) Diffusion Models (DMs) are trained to recover information from intentionally corrupted data via iterative denoising



B) DMs allow to modularly remap features to create novel instances



C) Phase transition in DMs reveal feature hierarchies in data

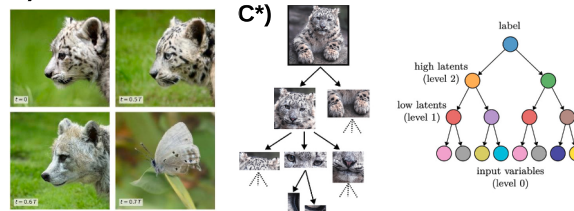


Fig. 5. Principles of diffusion models: (A) In denoising diffusion models (DMs), a forward process corrupts the data from a particular domain (here the data is represented by an image) by incrementally blending it with noise. An ANN learns a reverse process that inverts this corruption process by predicting the noise component at each step. In that way, DMs learn the underlying probability distribution of unstructured datasets and have become state-of-the-art generative models for a wide range of applications. Novel data is generated by an iterative error-minimization process that induces a trajectory through latent space: the model navigates from high-entropy states toward attractor regions conforming to the data distribution. Along this trajectory, structured features successively emerge from noise through spontaneous symmetry breaking into feature-specific attractor states [190] (image adapted from [169]). (B) In that way, DMs can modularly recombine features across hierarchical levels, enabling compositional generation of novel samples not explicitly present in the training data; panel B shows AI-generated schematic illustration created using ChatGPT (OpenAI) based on author-provided text prompts. (C) DMs reveal the hierarchical nature of data: at a characteristic noise level, high-level semantic attributes (e.g., class identity) change abruptly, while low-level features vary smoothly across the entire process (c.f., species vs. facial details in the left panels of C). This reflects a layered decomposition of latent representations (C*) that can be integrated by the DM at different stages during denoising (panels C and C* are adapted from [192]).

Morphogenesis describes the self-regulated growth of coherent anatomical structures of organisms. This describes trajectories through morphospace that start from high-entropy embryonic states, and which are then gradually refined into a mature morphology, rich in detail and structure. Importantly, this refinement happens gradually and is – by far – not deterministic. In contrast, morphogenesis relies on system-level feedback to gene regulatory dynamics at the subcellular level, and constructs organisms on the fly based on their current environmental context [36]. For instance, even after significant injuries or perturbations, organisms such as planaria or amphibians can restore limbs, organs, or entire parts of their body [36]. Both morphogenesis and regenerative capacity imply the presence of internal models or reference frames of a target morphology (suitable for the current context [13,44]) and mechanisms of the cellular collective capable of detecting and correcting deviations of disturbed states back toward viable attractor morphologies. The agential biological substrate (especially at the cellular and subcellular levels) thus constantly needs to counteract entropic effects and diffusive processes conditional to their current environmental context and genetic prompts, not only to maintain their own physiological integrity, but to actively integrate into and maintain the collective multiscale competency architecture of the body [73,74].

We thus argue that the relevance of diffusion models to both evolutionary search and morphogenesis is not coincidental. Their training objective gives them a strong inductive bias for systems that must construct or preserve structure under pervasive noise [170,195]. A diffusion model learns a score field over a target distribution: given a corrupted state at any noise level, it estimates directions of iterative correction that move the sample back toward the manifold of structured, viable states, potentially satisfying additional target conditions. Importantly, this score field is explicitly conditioned by the denoising step; the effective landscape of navigation is therefore continually remapped, with coarse structure typically established first and details refined later. Conditioning on other variables or features additionally changes which region of the learned manifold is targeted by remapping the score field, without changing the general denoising principle [171,194]. Evolutionary and developmental processes face precisely this class of problem at multiple scales: molecular, transcriptional, physiological, and anatomical states are continuously perturbed by noise, mutation, injury, aging, and environmental change, while the organism must develop, maintain, or reconstruct coherent form and

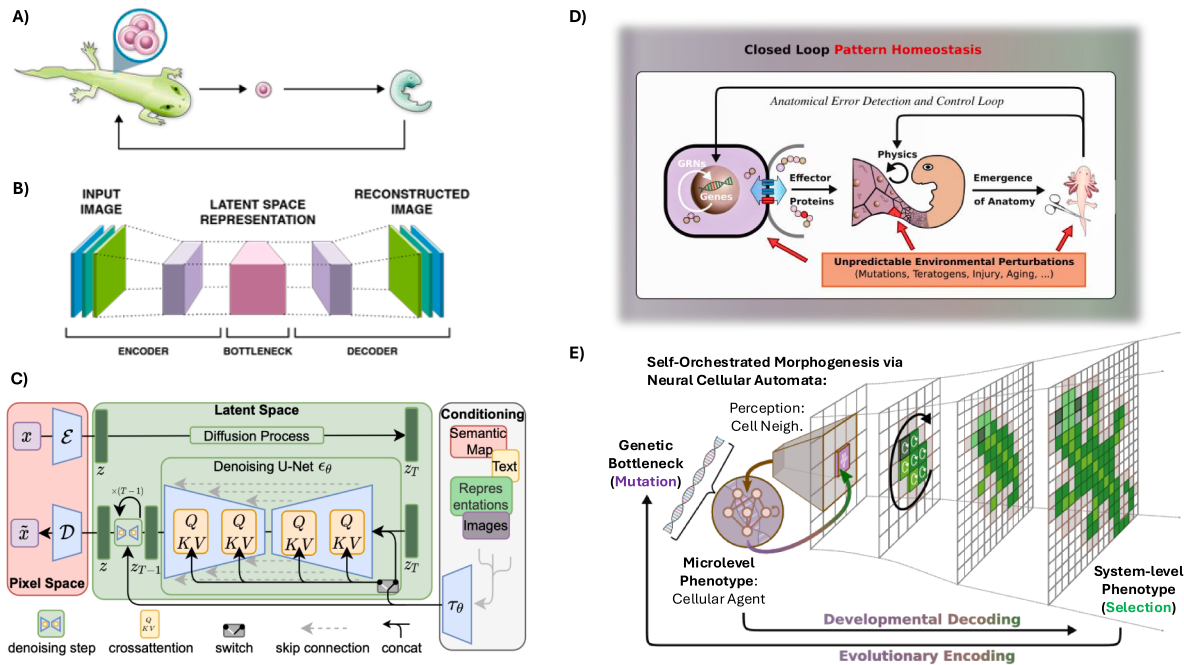


Fig. 6. Examples for navigating embedding spaces in unconventional substrates: (A) Classical evolutionary development illustrates the canonical bow-tie organization of biological systems: a compact genomic bottleneck stores lineage-level regularities as latent variables [73,74], which are then decoded by the developing embryo into a coherent, species-specific anatomical phenotype. This process is not a rigid mapping but a context-sensitive, distributed reconstruction carried out by competent cellular agents navigating physiological and anatomical spaces. (B) Conceptually, this is paralleled in ML research by autoencoders (AEs), which similarly compress high-dimensional inputs into a low-dimensional bottleneck whose latent variables support modular recombination and high-fidelity reconstruction (and recombination). This provides a simplified example of how compressed variables can encode generative structure, though biological decoding is far more plastic, recurrent, and iteratively error-corrective than the single-shot downstream decoding of typical AEs. (C) Latent diffusion models exemplify a richer generative process: initial high-entropy samples are iteratively refined through successive denoising steps toward complex, structured attractors (c.f., Fig. 5). These architectures embody incremental error correction, hierarchical feature emergence, and context-sensitive remapping, paralleling developmental morphogenesis [74], in which intermediate states are dynamically refined and recombined toward target anatomical outcomes. (D) Biological pattern homeostasis operates via nested closed-loop control systems: subcellular, cellular, tissue, and organ-level agents continuously detect deviations from preferred anatomical states and act to maintain homeostasis. These nested error-correction loops define a hierarchy of internal “embedding spaces” through which subsystems navigate, while higher levels reshape the energy landscape that lower-level agents inhabit, implementing the multiscale competency architecture of biology. (E) Neural cellular automata (NCAs) provide a synthetic analogue of this architecture. Each cell-like unit updates its internal state based on local perception-action loops (mediated by cell-specific artificial neural networks), and the collective of cells iteratively constructs a target morphology from a compact initial representation [74,223]. NCAs demonstrate how compressed encodings, combined with distributed, recurrent refinement protocols, can yield generative navigation protocols of cellular collectives toward system-level outcomes, reminiscent of the biological multiscale dynamics shown in (D). Images (A, B) by Jeremy Guay of Peregrine Creative used by permission from [36]; (C) adapted from [171]; (D, E) used by permission from [74].

function under changing context [36,74,195]. Diffusion models therefore provide more than a superficial analogy: they implement a computational framework for conditional reconstruction of structured targets from noisy, partially corrupted, or underspecified states. This is why the same remapping and denoising formalism can serve as cutting-edge generative models for custom datasets across modalities, as models for morphogenetic error correction in multicellular tissues [74], as a reproductive operator of genotypic parameters in evolutionary search strategies [174–177], and even as an implementation of generative thermodynamic computation [196,197].

At this stage, we would also like to emphasize similarities between diffusion models and transformers: despite having different architectures, both systems rely on incremental refinements of data representations. In transformers, each layer incrementally refines the context of token embeddings through attention maps and residual connections. In diffusion models, each denoising step incrementally reduces estimated errors from noisy data samples and guides them toward a configuration that conforms to the training data distribution. Both processes thus implement a navigation policy in embedding spaces, governed by incremental refinements at each layer or timestep. Very recent work on *DiffusionBlocks* [198] makes this connection explicit by treating residual blocks (e.g., in transformers) as denoising steps in a continuous-time diffusion process: each block is independently trained on its own noise interval, such that the stacked blocks collectively navigate the representation toward a more structured state through successive local corrections. More generally, both diffusion models and transformers have been associated with modern Hopfield-type associative memories [127,128,189], which moreover possess error-correction capabilities on the system-level scale, i.e., in restoring incomplete

data and generating novel instances by recombining features of learned data modularly and creatively, a connection that becomes even more apparent with the novel “Energy Transformer” architecture [199]. This parallel underscores how modern AI architectures increasingly align with principles of biological organization and cognition, where memory, planning, and perception all involve cycles of modeling, prediction, perception, and incremental error correction.

A closely related type of architecture to diffusion models with even more parallels to biological self-organization is that of Neural Cellular Automata (NCAs) [200,201]. Cellular Automata (CAs) are early models of *Artificial Life* [202,203] which date back to von Neumann’s self-replicating machines [204]. CAs have provided a foundational framework for modeling distributed biological phenomena such as replication, growth, and morphogenesis, simply through local interactions of neighboring cells on a discrete spatial grid. While CAs show life-like behavior [205,206], tuning their hard-coded local update rules by hand proved to be difficult if not infeasible. A recent extension of CAs termed NCAs replaces their update rule with flexible and trainable neural networks, notably the same network per cell. Analogous to CAs, the dynamics of NCAs resides around cells updating their internal state based on local measurements of their neighboring cells’ states. Strikingly, such NCAs can learn via differentiable [201,207–212] or evolutionary learning frameworks [193,213,214] the inverse problem of morphogenesis to grow a target morphology from a single seed cell. To name but a few more examples, NCAs can facilitate self-orchestrated computational tasks such as collective classification [211], path-finding in a maze akin to slime-molds [215,216], and even decentralized control of composite robots [217–221], all purely based on local intercellular communication and intracellular decision-making. NCAs can learn to regenerate complex patterns after damage or perturbation, reflecting a distributed capacity for self-regulation and repair, and might even allow us to understand and resolve aging [222]; see [223] for a more comprehensive review.

These models inherently operate as integrated networks of local agents continuously communicating with each other via self-refinement, blurring the line between morphological development and information processing [1]. Although the exact training paradigms deviate, NCAs’ dynamics reflect iterative self-regulatory (error-correcting) principles akin to the generative process of diffusion models: an NCA’s cells perceive their local environment, perform local cooperative communication, and iterative self-refinement until a collective target state is reached. By harnessing these principles, NCAs offer a powerful framework for understanding and engineering complex systems that need to reliably build and rebuild their form and function at the system-level from the bottom up; for a more in-depth discussion see [74]. Although trainability and scalability challenges remain [224–226], this makes NCAs powerful tools for encoding and navigating embedding spaces in a bio-akin way, where localized deviations must be corrected while preserving global structure, a property highly reminiscent of tissue-level repair in biology.

5. Limitations of the present framework

Despite these strong analogies, important limitations remain. The present framework should not be read as claiming that biological, cognitive, and artificial systems instantiate the same mathematical objects. Rather, remapping and navigation describe a more general organizational pattern in cognitive systems that can be realized across different substrates, with different geometries, update rules, constraints, and degrees of autonomy.

First, the term *embedding space* does not imply a uniformly smooth, continuous metric geometry. Many navigational problems are more naturally formulated over discrete or hybrid state spaces, where adjacency relations, transition probabilities, or graph structure define the effective metric of possible operations [227,228]. Regarding the present framework, remapping and navigation require an organized space of relations in which trajectories can be evaluated and corrected, but this organization need not be Euclidean, continuous, or even explicitly spatial. Accordingly, embedding spaces in modern AI should not be treated as uniformly smooth, continuous metric spaces.² Transformers operate over discrete token vocabularies, and recent work on weight quantization shows that LLM weights can be reduced to ternary values $\{-1, 0, 1\}$, i.e., $\log_2(3) \approx 1.585$ bits of information per weight, while retaining competitive performance [230]. Moreover, transformer attention is mathematically related to modern Hopfield-network retrieval, suggesting that token processing can also be viewed as rapid, context-specific associative retrieval toward stored patterns – discrete or metastable attractors in a high-dimensional energy landscape – not only as navigation through a continuous manifold [128]. Similarly, even in diffusion models operating over comparatively continuous image domains, the generative trajectory can exhibit a hierarchy of phase transitions or bifurcations into discrete semantic alternatives covering both coarse structure and fine-grained details [190–192].

Second, although next-token prediction should not naively be mistaken for mere memorization, the fact that transformers can construct useful internal representations does not make them equivalent to living world-modeling agents. Although LLMs contain large numbers of parameters, model capacity remains finite and language is combinatorially open-ended. Successful generalization therefore requires LLMs to compress recurrent regularities of text – in all the variants humans (and AIs) use it on the world wide web – into reusable internal representations. In this sense, it has been argued that next-token prediction can entail the implicit formation of world models in transformers, because human language is densely coupled to the structure of human thought and action in a physical Umwelt. Othello-GPT, for instance, maintains board-state representations, LLMs have internal linear representations of space and time, and video-generation models behave as partial visual world simulators [165,172,231,232]. In a broader sense, this relates intelligence as prediction and model-building across AI, evolution, and life [233]. However, whether, or to what extent, LLMs trained with next-token prediction qualify as world models remains a topic of intense debate [26]; so far, they are not (yet) equivalent to the embodied,

² A parallel discussion has been raised in philosophy of physics: Gisin argues that identifying physical quantities with mathematical real numbers imports infinite information that is not operationally accessible, and that many physical predictions can be formulated without committing to such ontological continuity [229].

self-maintaining world models of living agents. Relatedly, recent work suggests that LLMs can show limited and context-dependent functional access to their own internal states, but this should not be conflated with reliable introspection or biological self-awareness [234]. Current transformers also lack robust lifelong learning: they can adapt in context – possibly via implicit optimization or gradient-descent-like dynamics [235,236] – but persistent change still relies largely on offline training, fine-tuning, retrieval systems, or external memory [237].

Third, a long tradition in embodied AI emphasizes that intelligence is shaped by the morphology, material properties, and sensorimotor coupling of agents moving through the physical world [238–240]. This is clearly relevant for human or animal behavior, but should not be treated as the fundamental criterion for cognition. From the present perspective, the more general requirement is that an agent be coupled to some domain in which its perception, internal states, predictions, and actions form a navigable problem space. This space can also be virtual or cybernetic, or simply mathematical. Moreover, AI systems running on servers are not literally disembodied: they are implemented in physical hardware, embedded in data-center infrastructures, constrained by memory, latency, energy use, cooling, and thermodynamics. This is a genuine, albeit unusual, cybernetic embodiment, which to date is mostly externally maintained and requires heavily industrialized supply chains, i.e., is tightly coupled to other tools and humans through various interfaces [241]. Very recently, agent-only social platforms such as Moltbook/OpenClaw emerged, showing that LLM agents can already sustain rudimentary cybernetic social niches, including persistent accounts, communities, posts, comments, and platform-specific agendas; however, current analyses suggest that these networks – for now – remain shallower and less diverse in content, communication, and opinion than human communities [242–244]. Fourth, diffusion models and NCAs capture different aspects of morphogenetic error correction, but neither should be identified directly with biological morphogenesis. Diffusion models use an explicit denoising timestep and learned score field, whereas NCAs must internalize developmental time through recurrent, local, and often stochastic state updates. Both remain discretized computational approximations of biological processes, the latter being time-continuous, asynchronous, materially embodied, and metabolically constrained. Through extensive datasets, diffusion models can learn distributions of high-dimensional biological data increasingly well, yet largely remain black-boxes for better understanding broader bio-chemical and physical dynamics [188]. NCAs are closer to distributed cellular self-regulation, but they remain simplified models: their learned cell states are difficult to associate with biological data, they may collapse into trivial attractors, they can be hard to train, and they generally lack the metabolic, mechanical, biochemical, and embodied constraints of living tissues (see also [223]). Diffusion models, similarly, can be embedded into active-inference pipelines, but this does not mean that vanilla diffusion models by themselves, or integrated frameworks such as CHARLES-D [177] satisfy the criteria of active inference agents [245,246]. In contrast, living systems are organized into multiscale competency architectures [36]. At different hierarchical levels – molecular, cellular, tissue, organism – agents possess goals and the means to achieve those goal states via local error correction mechanisms: Cells can adjust gene expression profiles, tissues can reconfigure morphogen gradients, entire organisms can reroute behaviors, etc. The actors at each scale thus navigate their own embedding spaces, while contributing to the stability and adaptability of the system as a whole [36]. Furthermore, such a nested architecture enables concurrent processes to occur simultaneously at different scales and renders biology as an inherently polycomputing substrate [240]. This allows biological systems to process multiple streams of information in parallel and solve multiple problems simultaneously and across scales: the hardware of one layer may become the signals of yet another layer of organization. Altogether, these examples suggest that the fundamental organizational principle underlying both biological morphogenesis, cognitive processes, and modern AI systems like transformers, diffusion models, and NCAs is a form of remapping and navigating embedding spaces through error correction. Whether generating a coherent sentence, synthesizing a photorealistic image, or assembling a limb, self-regulative systems leverage iterative cycles of prediction and refinement to guide patterns through embedding spaces. In turn, this suggests that integrative (scale-free) error correction may be the essential computational backbone of intelligence itself, enabling systems to maintain their integrity while remaining flexible enough to adapt, innovate, and explore. All intelligence is collective [154,155], and the best localized agents can do depends on the extent of their cognitive light cones [247]. Collective error correction across a nested hierarchy of agents [36] might leverage self-regularized criticality [248–251] as an operational setpoint for scaling intelligence across scales.

6. Implications, connections, broader links

Identifying the interplay of remapping and navigation within embedding spaces as an invariant of cognition provides a unifying lens on adaptive intelligence across substrates. Integrated biological systems not only embody, but continually construct and revise internal maps – from transcriptional states to morphogenetic fields, organoids, organisms, and beyond – and use distributed error correction mechanisms toward stable attractor states at the (collective) system level. AI systems, though implemented in very different substrates, increasingly exploit analogous organization principles: embeddings compress complex data manifolds into navigable spaces, and iterative refinement steps drive trajectories toward coherent, increasingly contextualized outcomes. This dual principle positions cognition not as a property of a particular substrate, i.e., biological matter, but as a transformative process of continual (active) model refinement under compositional error minimization towards a system-level agenda.

One implication of this framework is that intelligence becomes inherently collective and multiscale – actors at one level can become the signal at other levels. In biology, localized agents, such as gene regulatory networks or cells, act in their own embedding spaces while higher-level systems reshape the energy landscapes of their parts. When integrated accordingly, this architecture enables robust scaling from molecular to organismal competencies. In turn, collective AI systems may prove more adaptive than monolithic models: ensembles of agents can maintain partially private embeddings and must learn to generalize from constrained environmental information, yet participate in shared embedding spaces, coordinating through distributed error correction. Such architectures align

with proposals that advanced intelligent systems must develop shared world models with humans to form collaborative systems, reframing alignment as a synergetic co-remapping (social) process rather than a one-sided constraint [252].

Second, the dual principle of remapping and navigation underscores scale-free incremental error correction as the operational backbone of cognition. Both biological organisms and modern AI models succeed not by providing predictions or actions in one pass, but by iteratively detecting mismatches against projected perceptions and corresponding refinements to the target state based on internal navigation policies. Importantly, many systems use the same, or similar mechanisms for learning and inference: the rules that remap embeddings during training also govern navigation protocols at runtime. These systems thus essentially learn transformative error-correction processes in diverse loss landscapes rather than attractor states directly, which represents a strong inductive, bio-inspired bias toward self-organizing systems that are robust against entropic disorder. This suggests that future intelligent systems should unify learning and deployment under a single engine of iterative refinement in dynamically remappable embedding spaces. Finally, we propose near-criticality as a candidate universal setpoint for such scale-free, substrate-agnostic cognitive engines [3,248]. Systems near critical regimes exhibit the very properties required for continual remapping – sensitivity to perturbation, scale invariance, tunability, and high information throughput. In physical terms, near-criticality can transiently increase correlation lengths, expanding the spatial or functional range over which components influence one another. We hypothesize that external stressors, internal perturbations, or novel contexts can bring selected degrees of freedom, or (sub-)components of such systems, toward critical regimes, where the (sub-)space of alternative states becomes statistically accessible to remapping dynamics. In such a high-susceptibility regime, weak cues can then rapidly bias collective agentic systems toward particular attractors, breaking the symmetries among competing alternatives. This is analogous to how a small external field can bias global magnetization near a critical point in the Ising model, but here the relevant cues may be metabolic, bioelectric, transcriptional, behavioral, social, or technological. Yet criticality is fragile: small fluctuations can cascade, with effects ranging from local responses to system-wide reconfigurations often following power-law distributions, as classically illustrated by Bak’s sandpile model of self-organized criticality [253,254]; likewise, training artificial networks close to criticality can destabilize learning. For future self-regulatory AI systems, this suggests a delicate balance: architectures must incorporate homeostatic mechanisms to navigate toward stable attractors, yet must remain adaptable when necessary. Criticality might be a substrate-independent natural fallback mechanism of collective self-regulatory systems that enables exploration and search upon perturbation or under stress, but which must be actively regulated otherwise to allow stable navigation toward viable attractor states. Being in a near-critical state does not simply “cause” remapping, nor do remapping systems automatically become critical; rather, adaptive biological systems may have evolved (because of) regulatory mechanisms that transiently bring subsystems toward near-critical regimes under perturbation, where partial or large-scale remapping of internal embedding spaces is most effective, before homeostatic mechanisms restore stable navigation within an updated system-level attractor landscape. Evidence from neuroscience suggests that brains exploit such near-critical dynamics for flexible information transfer and rapid switching among metastable states [249–251], and theoretical work argues that criticality may be the natural operating point for adaptive systems [248]. Recently, tumorigenesis was described as a critical transition between normal and malignant attractor states, i.e., as a remapping of the transcriptomic landscape that can bias adjacent tissue-level dynamics toward tumor-supportive states; promising reversion strategies aim to reshape, i.e., remap, the regulatory landscape away from this metastable regime through intra- and inter-cellular pathways [255]. As discussed above, both diffusion models and NCAs may exploit phase transitions and near-critical dynamics during their hierarchical generative processes [190–192,223]; in transformers, layer normalization and residual pathways likewise stabilize signal and gradient propagation across layers, with recent analyses relating transformer dynamics to critical or subcritical signal-propagation regimes [256,257]. At still larger scales, societies may enter high-susceptibility regimes under crisis, novelty, or technological change, where weak but well-placed interventions can reorganize narratives, practices, social norms, or institutions. In this sense, criticality offers a possible physical basis for behavior shaping across scales and substrates: weak cues can redirect collective dynamics on physical principles when the relevant degrees of freedom – across abstraction levels – are sufficiently susceptible. Put differently, regulated access to near-criticality may expand a system’s cognitive light cone [247]: the spatial or functional range over which components can influence one another and participate in a shared remapping. Relaxation away from criticality would then stabilize the selected attractor and return the system to a more local, robust mode of navigation.

In conclusion, we frame remapping and navigation of embedding spaces as a scale-free organizational principle of cognition. This dual process links developmental robustness with generative modeling, relates collective AI within the same framework as multicellular coordination, and suggests (near-)criticality as a dynamical setpoint naturally supporting the required balance between stability and adaptability in cognitive systems. To the general framework of the Free-Energy Principle [28,29], which in its most general form assumes only a dynamical system that factors into conditionally-independent components, it adds architectural principles with which the generative models of systems with sufficient degrees of freedom and internal structure – including living systems at all scales – can be expected to comply. These include, in particular, remappings that radically decrease state-space dimensionality, such as the localization of objects to regions of 4D spacetime, and navigation procedures that rely on coarse-graining and cheap heuristics, as seen in bacterial stress response [258] or the inference of object persistence from trajectory shape in primates [259]. By situating remapping and navigation as invariants of cognition, we suggest a substrate-independent principle that links diverse intelligences across scales. This perspective emphasizes not just descriptive parallels, but also operational mechanisms for how adaptive and resilient minds – biological, artificial, or hybrid – emerge from shared (self-)organizational dynamics.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Michael Levin reports financial support was provided by Astonishing Labs, Inc. Michael Levin reports financial support

was provided by Templeton World Charity Foundation Inc. Michael Levin reports financial support was provided by John Templeton Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank members of the Levin Lab for helpful discussions, and Tomika Gotch for help with the manuscript. We gratefully acknowledge support for this work provided through a sponsored research agreement with Astonishing Labs, and the [Templeton World Charity Foundation, Inc. \(TWCF0606\)](#). This publication was made possible through the support of Grant 62212 from the [John Templeton Foundation](#). The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation.

Supplementary material

Supplementary material associated with this article can be found in the online version at [10.1016/j.plrev.2026.06.009](https://doi.org/10.1016/j.plrev.2026.06.009)

References

- [1] Levin M. Technological approach to mind everywhere: an experimentally-grounded framework for understanding diverse bodies and minds. *Front Syst Neurosci* 2022;16. <https://doi.org/10.3389/fnsys.2022.768201>
- [2] Fields C, Levin M. Competency in navigating arbitrary spaces as an invariant for analyzing cognition in diverse embodiments. *Entropy* 2022;24(6):819.
- [3] Fields C, Levin M. Scale-free biology: integrating evolutionary and developmental thinking. *BioEssays* 2020;42(8):1900228.
- [4] Lyon P, Keijzer F, Arendt D, Levin M. Reframing cognition: getting down to biological basics. *Philos Trans R Soc B Biol Sci* 2021;376(1820):20190750. <https://doi.org/10.1098/rstb.2019.0750>
- [5] Abramson CI, Levin M. Behaviorist approaches to investigating memory and learning: a primer for synthetic biology and bioengineering. *Commun Integr Biol* 2021;14(1):230–47. PMID: 34925687; <https://doi.org/10.1080/19420889.2021.2005863>
- [6] Pezzulo G, Levin M. Top-down models in biology: explanation and control of complex living systems above the molecular level. *J R Soc Interface* 2016;13(124):20160555.
- [7] Lyon P. The cognitive cell: bacterial behavior reconsidered. *Front Microbiol* 2015;6–2015. <https://doi.org/10.3389/fmicb.2015.00264>
- [8] Keijzer F, van Duijn M, Lyon P. What nervous systems do: early evolution, input-output, and the skin brain thesis. *Adapt Behav* 2013;21(2):67–85. <https://doi.org/10.1177/1059712312465330>
- [9] Maturana HR, Varela FJ. *Autopoiesis and cognition: the realization of the living*; vol. 42. Springer Science & Business Media; 2012.
- [10] Pattee HH, Raczaszek-Leonardi J, Pattee HH. Cell psychology: an evolutionary approach to the symbol-matter problem. In: LAWS, LANGUAGE and LIFE: Howard Pattee's classic papers on the physics of symbols with contemporary commentary. Dordrecht: Springer Netherlands. ISBN 978-94-007-5161-3; 2012, p. 165–79. https://doi.org/10.1007/978-94-007-5161-3_11
- [11] Ashby WR. *Design for a brain: the origin of adaptive behaviour*. New York: John Wiley & Sons; 1952.
- [12] James W. *The principles of psychology*. New York: Henry Holt and Company; 1890. <https://doi.org/10.1037/10538-000>
- [13] Levin M. Bioelectric signaling: reprogrammable circuits underlying embryogenesis, regeneration, and cancer. *Cell* 2021;184(8):1971–89.
- [14] Baluška F, Miller WB, Reber AS. Biomolecular basis of cellular consciousness via subcellular nanobrain. *Int J Mol Sci* 2021;22(5). <https://doi.org/10.3390/ijms22052545>
- [15] Levin M, Keijzer F, Lyon P, Arendt D. Uncovering cognitive similarities and differences, conservation and innovation. *Philos Trans R Soc B Biol Sci* 2021;376(1821):20200458. <https://doi.org/10.1098/rstb.2020.0458>
- [16] Lyon P. Of what is “minimal cognition” the half-baked version? *Adapt Behav* 2020;28(6):407–24. <https://doi.org/10.1177/1059712319871360>
- [17] Bayne T, Brainard D, Byrne RW, Chittka L, Clayton N, Heyes C, et al. What is cognition? *Curr Biol* 2019;29(13):R608–R615. <https://doi.org/10.1016/j.cub.2019.05.044>
- [18] Fields C, Glazebrook JF, Levin M. Minimal physicalism as a scale-free substrate for cognition and consciousness. *Neurosci Conscious* 2021;2021(2):niab013. <https://doi.org/10.1093/nc/niab013>
- [19] Levin M. Bioelectric networks: the cognitive glue enabling evolutionary scaling from physiology to mind. *Anim Cogn* 2023;26(6):1865–91. <https://doi.org/10.1007/s10071-023-01780-3>
- [20] Clawson WP, Levin M. Endless forms most beautiful 2.0: teleonomy and the bioengineering of chimaeric and synthetic organisms. *Biol J Linn Soc* 2022;139(4):457–86. <https://doi.org/10.1093/biolinnean/blac073>
- [21] Davies J, Levin M. Synthetic morphology with agential materials. *Nat Rev Bioeng* 2023;1(1):46–59. <https://doi.org/10.1038/s44222-022-00001-9>
- [22] Lagasse E, Levin M. Future medicine: from molecular pathways to the collective intelligence of the body. *Trends Mol Med* 2023;29(9):687–710. <https://doi.org/10.1016/j.molmed.2023.06.007>
- [23] Levin M. The multiscale wisdom of the body: collective intelligence as a tractable interface for next-generation biomedicine. *BioEssays* 2025;47(3):e202400196. <https://doi.org/10.1002/bies.202400196>
- [24] Ha D, Schmidhuber J. Recurrent world models facilitate policy evolution. In: Bengio S, Wallach HM, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3–8, 2018, Montréal, Canada*. 2018, p. 2455–67. <https://proceedings.neurips.cc/paper/2018/hash/2de5d16682c3c35007e4e92982f1a2ba-Abstract.html>
- [25] Craik K JW. *The nature of explanation*. Cambridge: Cambridge University Press; 1967.
- [26] Maes L, Lidec QL, Scieur D, LeCun Y, Balestrieri R. LeWorldModel: stable end-to-end joint-embedding predictive architecture from pixels. 2026. <https://arxiv.org/abs/2603.19312>
- [27] Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 2010;11(2):127–38. <https://doi.org/10.1038/nrn2787>
- [28] Ramstead MJ, Sakthivadivel DAR, Heins C, Koudahl M, Millidge B, Da Costa L, et al. On Bayesian mechanics: a physics of and by beliefs. *Interface Focus* 2022;13:20220029.
- [29] Friston KJ, Da Costa L, Sakthivadivel DAR, Heins C, Pavliotis GA, Ramstead MJ, et al. Path integrals, particular kinds, and strange things. *Phys Life Rev* 2023;47:35–62.
- [30] Pearl J. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann: San Mateo, CA; 1988.
- [31] Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, O'Doherty J, Pezzulo G. Active inference and learning. *Neurosci Biobehav Rev* 2016;68:862–79. <https://doi.org/10.1016/j.neubiorev.2016.06.022>
- [32] Parr T, Pezzulo G, Friston KJ. *Active inference: the free energy principle in mind, brain, and behavior*. The MIT Press; 2022. <https://doi.org/10.7551/mitpress/12441.001.0001>
- [33] Fields C, Friston K, Glazebrook JF, Levin M. A free energy principle for generic quantum systems. *Prog Biophys Mol Biol* 2022;173:36–59. <https://doi.org/10.1016/j.pbiomolbio.2022.05.006>

- [34] Biswas S, Clawson W, Levin M. Learning in transcriptional network models: computational discovery of pathway-level memory and effective interventions. *Int J Mol Sci* 2022;24(1):285.
- [35] Fields C, Levin M. Regulative development as a model for origin of life and artificial life studies. *Biosystems* 2023;229:104927. <https://doi.org/10.1016/j.biosystems.2023.104927>
- [36] Levin M. Darwin's agential materials: evolutionary implications of multiscale competency in developmental biology. *Cell Mol Life Sci* 2023;80(6):142.
- [37] MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, et al. The evolution of self-control. *Proc Natl Acad Sci* 2014;111(20):E2140–E2148. <https://doi.org/10.1073/pnas.1323533111>
- [38] Seed A, Byrne R. Animal tool-use. *Curr Biol* 2010;20(23):R1032–R1039. <https://doi.org/10.1016/j.cub.2010.09.042>
- [39] Shettleworth SJ. Cognition, evolution, and behavior. Oxford University Press; 2009.
- [40] Emery NJ, Clayton NS. The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 2004;306(5703):1903–7. <https://doi.org/10.1126/science.1098410>
- [41] O'Keefe J, Nadel L. The hippocampus as a cognitive map. Oxford University Press; 1978.
- [42] Tolman EC. Cognitive maps in rats and men. *Psychol Rev* 1948;55(4):189–208. <https://doi.org/10.1037/h0061626>
- [43] Levin M, Pietak AM, Bischof J. Planarian regeneration as a model of anatomical homeostasis: recent progress in biophysical and computational approaches. In: *Seminars in cell & developmental biology*; vol. 87. Elsevier; 2019, p. 125–44.
- [44] Durant F, Morokuma J, Fields C, Williams K, Adams DS, Levin M. Long-term, stochastic editing of regenerative anatomy via targeting endogenous bioelectric gradients. *Biophys J* 2017;112(10):2231–43.
- [45] Emmons-Bell M, Durant F, Tung A, Pietak A, Miller K, Kane A, et al. Regenerative adaptation to electrochemical perturbation in planaria: a molecular analysis of physiological plasticity. *iScience* 2019;22:147–65. <https://doi.org/10.1016/j.isci.2019.11.014>
- [46] Biswas S, Manicka S, Hoel E, Levin M. Gene regulatory networks exhibit several kinds of memory: quantification of memory in biological and random transcriptional networks. *iScience* 2021;24(3):102131.
- [47] Tanaka EM, Reddien PW. The molecular and cellular basis of regeneration: from planaria to vertebrates. *Dev Cell* 2016;38(5):459–65.
- [48] Levin M. Morphogenetic fields in embryogenesis, regeneration, and cancer: non-local control of complex patterning. *Biosystems* 2012;109(3):243–61.
- [49] Wolpert L, Tickle C, Arias AM. Principles of development. Oxford University Press, USA; 2015.
- [50] Rogers A, Kovaleva O, Rumshisky A. A primer in BERTology: what we know about how BERT works. *Trans Assoc Comput Linguist* 2020;8:842–66.
- [51] Yang J, Jin H, Tang R, Han X, Feng Q, Jiang H, et al. Harnessing the power of LLMs in practice: a survey on ChatGPT and beyond. *ACM Trans Knowl Discov Data* 2024;18(6):1–32.
- [52] Schaal S. Dynamic movement primitives-a framework for motor control in humans and humanoid robotics. In: *Adaptive motion of animals and machines*. Springer; 2006, p. 261–80.
- [53] Baltrušaitis T, Ahuja C, Morency L-P. Multimodal machine learning: a survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell* 2018;41(2):423–43.
- [54] Kandel ER. The molecular biology of memory storage: a dialogue between genes and synapses. *Science* 2001;294(5544):1030–8.
- [55] Martin SJ, Grimwood PD, Morris R GM. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annu Rev Neurosci* 2000;23(1):649–711.
- [56] Fenton AA. Remapping revisited: how the hippocampus represents different spaces. *Nat Rev Neurosci* 2024;25(6):428–48.
- [57] Tonegawa S. Somatic generation of antibody diversity. *Nature* 1983;302(5909):575–81.
- [58] Schatz DG, Swanson PC. V (D) J recombination: mechanisms of initiation. *Annu Rev Genet* 2011;45(1):167–202.
- [59] Laurenti E, Göttgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* 2018;553(7689):418–26.
- [60] Szathmáry E, Smith JM. The major evolutionary transitions. *Nature* 1995;374(6519):227–32. <https://doi.org/10.1038/374227a0>
- [61] Szathmáry E. Toward major evolutionary transitions theory 2.0. *Proc Natl Acad Sci* 2015;112(33):10104–11. <https://doi.org/10.1073/pnas.1421398112>
- [62] Watson RA, Wagner GP, Pavlicev M, Weinreich DM, Mills R. The evolution of phenotypic correlations and “developmental memory”. *Evolution* 2014;68(4):1124–38. <https://doi.org/10.1111/evo.12337>
- [63] Watson RA, Mills R, Buckley CL, Kouvaris K, Jackson A, Powers ST, et al. Evolutionary connectionism: algorithmic principles underlying the evolution of biological organisation in evo-devo, evo-eco and evolutionary transitions. *Evol Biol* 2015;43(4):553–81. <https://doi.org/10.1007/s11692-015-9358-z>
- [64] Power DA, Watson RA, Szathmáry E, Mills R, Powers ST, Doncaster CP, et al. What can ecosystems learn? Expanding evolutionary ecology with learning theory. *Biol Direct* 2015;10(1). <https://doi.org/10.1186/s13062-015-0094-1>
- [65] Watson RA, Szathmáry E. How can evolution learn? *Trends Ecol Evol* 2016;31(2):147–57. <https://doi.org/10.1016/j.tree.2015.11.009>
- [66] Kouvaris K, Clune J, Kounios L, Brede M, Watson RA. How evolution learns to generalise: using the principles of learning theory to understand the evolution of developmental organisation. *PLOS Comput Biol* 2017;13(4):e1005358. <https://doi.org/10.1371/journal.pcbi.1005358>
- [67] Buckley CL, Lewens T, Levin M, Millidge B, Tschantz A, Watson RA. Natural induction: spontaneous adaptive organisation without natural selection. *Entropy* 2024;26(9). <https://doi.org/10.3390/e26090765>
- [68] Watson RA, Levin M, Lewens T. Evolution by natural induction. *Interface Focus* 2025a;15(6):20250025. <https://doi.org/10.1098/rsfs.2025.0025>
- [69] Watson RA, Levin M, Lewens T. Evolution by natural induction II: further interactions with natural selection. *Interface Focus* 2025b;15(6):20250036. <https://doi.org/10.1098/rsfs.2025.0036>
- [70] Jacob F. Evolution and tinkering. *Science* 1977;196(4295):1161–6. <https://doi.org/10.1126/science.860134>
- [71] Wagner GP, Altenberg L. Perspective: complex adaptations and the evolution of evolvability. *Evolution* 1996;50(3):967–76. <https://doi.org/10.1111/j.1558-5646.1996.tb02339.x>
- [72] Wagner A. Robustness and evolvability: a paradox resolved. *Proc R Soc B Biol Sci* 2007;275(1630):91–100. <https://doi.org/10.1098/rspb.2007.1137>
- [73] Mitchell KJ, Cheney N. The genomic code: the genome instantiates a generative model of the organism. *Trends Genet* 2025;41(6):462–79. <https://doi.org/10.1016/j.tig.2025.01.008>
- [74] Hartl B, Levin M. What does evolution make? learning in living lineages and machines. *Trends Genet* 2025;41(6):480–96. <https://doi.org/10.1016/j.tig.2025.04.002>
- [75] Levin M, Watson R. Machines all the way up and cognition all the way down: updating the machine metaphor in biology. *Semin Cell Dev Biol* 2026;177:178:103668. <https://doi.org/10.1016/j.semdb.2026.103668>
- [76] Fields C, Levin M. Thoughts and thinkers: on the complementarity between objects and processes. *Phys Life Rev* 2025;52:256–73. <https://doi.org/10.1016/j.phlev.2025.01.008>
- [77] Benson NC, Butt OH, Datta R, Radoeva PD, Brainard DH, Aguirre GK. The retinotopic organization of striate cortex is well predicted by surface topology. *Curr Biol* 2012;22(21):2081–5.
- [78] Moerel M, De Martino F, Formisano E. An anatomical and functional topography of human auditory cortical areas. *Front Neurosci* 2014;8:225.
- [79] Nakashima A, Takeuchi H. Roles of odorant receptors during olfactory glomerular map formation. *Genesis* 2024;62(3):e23610.
- [80] Luo L, Flanagan JG. Development of continuous and discrete neural maps. *Neuron* 2007;56(2):284–300.
- [81] Hensch TK. Critical period plasticity in local cortical circuits. *Nat Rev Neurosci* 2005;6(11):877–88.
- [82] Griffin DR. Listening in the dark: the acoustic orientation of bats and men. Yale University Press; 1958.
- [83] Gilbert CD, Wiesel TN. Receptive field dynamics in adult primary visual cortex. *Nature* 1992;356(6365):150–2.
- [84] Strelnikov K, Rouger J, Demonet J-F, Lagleyre S, Fraysse B, Deguine O, et al. Visual activity predicts auditory recovery from deafness after adult cochlear implantation. *Brain* 2013;136(12):3682–95.
- [85] Kalaska JF. From intention to action: motor cortex and the control of reaching movements. In: *Sternad D (eds) Progress in motor control: a multidisciplinary perspective*. Springer US, Boston, MA; 2009:139–78. https://doi.org/10.1007/978-0-387-77064-2_8
- [86] Harrison TC, Murphy TH. Motor maps and the cortical control of movement. *Curr Opin Neurobiol* 2014;24:88–94.
- [87] Lebedev MA, Nicolelis M AL. Brain-machine interfaces: past, present and future. *Trends Neurosci* 2006;29(9):536–46.

- [88] Bruno D, Schurmann Vignaga S. Addenbrooke's cognitive examination III in the diagnosis of dementia: a critical review. *Neuropsychiatr Dis Treat* 2019;15:441–7.
- [89] Nour MM, McNamee DC, Liu Y, Dolan RJ. Trajectories through semantic spaces in schizophrenia and the relationship to ripple bursts. *Proc Natl Acad Sci* 2023;120(42):e2305290120.
- [90] Schmidhuber J. A possibility for implementing curiosity and boredom in model-building neural controllers. In: *Proc. of the international conference on simulation of adaptive behavior: from animals to animats*. 1991, p. 222–7.
- [91] Friston K, Adams RA, Perrinet L, Breakspear M. Perceptions as hypotheses: saccades as experiments. *Front Psychol* 2012;3:151.
- [92] Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis Res* 2000;40(10–12):1489–506.
- [93] Gottlieb J, Oudeyer P-Y, Lopes M, Baranes A. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn Sci* 2013;17(11):585–93.
- [94] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst* 2020;33:1877–901.
- [95] Dong Q, Li L, Dai D, Zheng C, Ma J, Li R, et al. A survey on in-context learning. 2022; arXiv:2301.00234.
- [96] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [97] Sutton RS, Barto AG, et al. Reinforcement learning: an introduction. Cambridge, MA: MIT Press; 1998.
- [98] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature* 2015;518(7540):529–33.
- [99] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in Neural Information Processing Systems (NeurIPS)*; vol. 30. 2017, p. 5998–6008.
- [100] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. *Adv Neural Inf Process Syst (NeurIPS)* 2014;27:2672–80.
- [101] Filliat D, Meyer J-A. Map-based navigation in mobile robots: I. A review of localization strategies. *Cogn Syst Res* 2003;4(4):243–82.
- [102] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: *International Conference on Machine Learning. PMLR*; 2017, p. 1126–35.
- [103] Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. *Behav Brain Sci* 2017;40:e253.
- [104] Epstein RA, Patai EZ, Julian JB, Spiers HJ. The cognitive map in humans: spatial navigation and beyond. *Nat Neurosci* 2017;20(11):1504–13.
- [105] Alvernhe A, Save E, Poucet B. Local remapping of place cell firing in the Tolman detour task. *Eur J Neurosci* 2011;33(9):1696–705.
- [106] He Q, McNamara TP, Brown TI. Manipulating the visibility of barriers to improve spatial navigation efficiency and cognitive mapping. *Sci Rep* 2019;9(1):11567.
- [107] Schuck NW, Niv Y. Sequential replay of nonspatial task states in the human hippocampus. *Science* 2019;364(6447):eaaw5181.
- [108] Garvert MM, Dolan RJ, Behrens TEJ. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife* 2017;6:e17086.
- [109] Iaria G, Burles F. Developmental topographical disorientation. *Trends Cogn Sci* 2016;20(10):720–2.
- [110] Pio-Lopez L, Kuchling F, Tung A, Pezzulo G, Levin M. Active inference, morphogenesis, and computational psychiatry. *Front Comput Neurosci* 2022;16:988977.
- [111] Pisella L, Mattingley JB. The contribution of spatial remapping impairments to unilateral visual neglect. *Neurosci Biobehav Rev* 2004;28(2):181–200.
- [112] Horsman D, Kendon V, Stepney S. The natural science of computing. *Comms ACM* 2017;60(8):31–4.
- [113] Hartshorne R. Algebraic geometry (Graduate texts in mathematics, No. 52). Springer, Berlin; 1977.
- [114] Font B, Spivak D. Seven sketches in compositionality: an invitation to applied category theory. 2018. arXiv:1803.05316v1 [math.CT].
- [115] Abramsky S, Brandenburger A. The sheaf-theoretic structure of non-locality and contextuality. *New J Phys* 2011;13:113036.
- [116] Fields C, Glazebrook JF. Distributed information and computation in generic quantum systems. Springer, Cham; 2025.
- [117] Zwier MC, Chong LT. Reaching biological timescales with all-atom molecular dynamics simulations. *Curr Opin Pharmacol* 2010;10:745–52.
- [118] Fuster D. Generative deep learning. Sebastopol, CA: O'Reilly Media; 2 ed.; 2023.
- [119] Sutton RS. The bitter lesson. Incomplete ideas blog; 2019. <https://www.incompleteideas.net/IncIdeas/BitterLesson.html>.
- [120] Yousefi M, Collins J. Learning the bitter lesson: empirical evidence from 20 years of CVPR proceedings. In: Peled-Cohen L, Calderon N, Lissak S, Reichart R, editors. *Proceedings of the 1st workshop on NLP for science (NLP4Science)*. Miami, FL, USA: Association for Computational Linguistics; 2024, p. 175–87. <https://doi.org/10.18653/v1/2024.nlp4science-1.15>
- [121] Li X, Peng L, Wang Y-P, Zhang W. Open challenges and opportunities in federated foundation models towards biomedical healthcare. *BioData Min* 2025;18(1). <https://doi.org/10.1186/s13040-024-00414-9>
- [122] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Bengio Y, LeCun Y, (Eds.), 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2–4, 2013, workshop track proceedings. 2013, <http://arxiv.org/abs/1301.3781>.
- [123] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. In: Bengio Y, LeCun Y, (Eds.), 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, conference track proceedings. 2015, <http://arxiv.org/abs/1409.0473>.
- [124] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. In: 9th International Conference on Learning Representations, ICLR 2021, virtual event, Austria, May 3–7, 2021. OpenReview.net; 2021, <https://openreview.net/forum?id=YicbFdNTTy>.
- [125] Jha RD, Zhang C, Shmatikov V, Morris JX. Harnessing the universal geometry of embeddings. 2025. <https://doi.org/10.48550/ARXIV.2505.12540>
- [126] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, et al. Learning transferable visual models from natural language supervision. In: Meila M, Zhang T, editors. *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18–24 July 2021, virtual event; vol. 139 of Proceedings of Machine Learning Research. PMLR*; 2021, p. 8748–63. <http://proceedings.mlr.press/v139/radford21a.html>.
- [127] Krotov D, Hoover B, Ram P, Pham B. Modern methods in associative memory. 2025; arXiv:2507.06211
- [128] Ramsauer H, Schäfl B, Lehner J, Seidl P, Widrich M, Gruber L, et al. Hopfield networks is all you need. In: *International Conference on Learning Representations (ICLR)*. OpenReview; 2021.
- [129] Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci* 1982;79(8):2554–8. <https://doi.org/10.1073/pnas.79.8.2554>
- [130] Kanerva P. Hyperdimensional computing: an introduction to computing in distributed representation with high-dimensional random vectors. *Cogn Comput* 2009;1(2):139–59.
- [131] Burrello A, Schindler KA, Benini L, Rahimi A. Hyperdimensional computing with local binary patterns: one-shot learning of seizure onset and identification of ictogenic brain regions using short-time iEEG recordings. *IEEE Trans Biomed Eng* 2020;67(2):601–13.
- [132] Lindsey J, Gurnee W, Ameisen E, Chen B, Pearce A, Turner NL, et al. On the biology of a large language model. *Transform Circuits Thread* 2025; <https://transformer-circuits.pub/2025/attribution-graphs/biology.html>.
- [133] Zhang D, Li Z-Z, Zhang M-L, Zhang J, Liu Z, Yao Y, et al. From system 1 to system 2: a survey of reasoning large language models. *IEEE Trans Pattern Anal Mach Intell* 2026;48(3):3335–54. <https://doi.org/10.1109/tpami.2025.3637037>
- [134] Lawsen A. Comment on the illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity. 2025. <https://doi.org/10.48550/ARXIV.2506.09250>
- [135] Shojaei P, Mirzadeh I, Alizadeh K, Horton M, Bengio S, Farajtabar M. The illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity. 2025. <https://doi.org/10.48550/ARXIV.2506.06941>
- [136] Kojima T, Gu SS, Reid M, Matsuo Y, Iwasawa Y. Large language models are zero-shot reasoners. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, (Eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28, - December 9, 2022*. 2022, http://papers.nips.cc/paper_files/paper/2022/hash/8bb0d291acd4ac0f6ef112099c16f326-Abstract-Conference.html.
- [137] Peng B, Alcaide E, Anthony Q, Albalak A, Arcadinho S, Biderman S, et al. RWKV: reinventing RNNs for the transformer era. 2023; arXiv:2305.13048

- [138] Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. In: First Conference on Language Modeling (COLM). 2024, <https://openreview.net/forum?id=tEYskw1VY2>.
- [139] Beck M, Pöppel K, Lippe P, Kurlle R, Blies PM, Klambauer G, et al. xLSTM 7b: a recurrent LLM for fast and efficient inference. In: Forty-second International Conference on Machine Learning (ICML). 2025, <https://openreview.net/forum?id=LV3DpKD08B>.
- [140] Geiping J, McLeish SM, Jain N, Kirchenbauer J, Singh S, Bartoldson BR, et al. Scaling up test-time compute with latent reasoning: a recurrent depth approach. In: ES-FoMo III: 3rd workshop on Efficient Systems for Foundation Models. 2025, <https://openreview.net/forum?id=D6o6Bwtq7h>.
- [141] Dehghani M, Gouws S, Vinyals O, Uszkoreit J, Kaiser L. Universal transformers. In: International Conference on Learning Representations. 2019, <https://openreview.net/forum?id=HyzdRiR9Y7>.
- [142] Fan Y, Du Y, Ramchandran K, Lee K. Looped transformers for length generalization. In: The thirteenth International Conference on Learning Representations. 2025, <https://openreview.net/forum?id=2edigk8yoU>.
- [143] Dawid A, LeCun Y. Introduction to latent variable energy-based models: a path towards autonomous machine intelligence. 2023. <https://doi.org/10.48550/ARXIV.2306.02572>
- [144] Assran M, Duval Q, Misra I, Bojanowski P, Vincent P, Rabbat MG, et al. Self-supervised learning from images with a joint-embedding predictive architecture. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17–24, 2023. IEEE; 2023, p. 15619–29. <https://doi.org/10.1109/CVPR52729.2023.01499>
- [145] Assran M, Bardes A, Fan D, Garrido Q, Howes R, Mojtaba, et al. V-JEPA 2: self-supervised video models enable understanding, prediction and planning. 2025; <https://arxiv.org/abs/2506.09985>.
- [146] Friston KJ, Moran RJ, Nagai Y, Taniguchi T, Gomi H, Tenenbaum J. World model learning and inference. Neural Netw 2021;144:573–90. <https://doi.org/10.1016/J.NEUNET.2021.09.011>
- [147] Leadholm N, Clay V, Knudstrup S, Lee H, Hawkins J. Thousand-brains systems: sensorimotor intelligence for rapid, robust learning and inference. 2025; <https://arxiv.org/abs/2507.04494>.
- [148] Huh M, Cheung B, Wang T, Isola P. Position: the Platonic representation hypothesis. In: Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21–27, 2024. OpenReview.net; 2024, <https://openreview.net/forum?id=BH8TYy0r6u>.
- [149] Levin M. Ingressing minds: causal patterns beyond genetics and environment in natural, synthetic, and hybrid embodiments. OSF preprint https://doi.org/10.31234/osf.io/5g2xj_v1
- [150] Kumar A, Clune J, Lehman J, Stanley KO. Questioning representational optimism in deep learning: the fractured entangled representation hypothesis. 2025. <https://doi.org/10.48550/ARXIV.2505.11581>
- [151] Whitelam S, Selin V, Park S-W, Tamblyn I. Correspondence between neuroevolution and gradient descent. Nat Commun 2021;12(1). <https://doi.org/10.1038/s41467-021-26568-2>
- [152] Bardes A, Ponce J, LeCun Y. VicReg: variance-invariance-covariance regularization for self-supervised learning. In: The tenth International Conference on Learning Representations, ICLR 2022, virtual event, April 25–29, 2022. OpenReview.net; 2022, <https://openreview.net/forum?id=xm6YD62D1Ub>.
- [153] Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 2013;36(3):181–204. <https://doi.org/10.1017/s0140525x12000477>
- [154] Levin M. Artificial intelligences: a bridge toward diverse intelligence and humanity's future. Adv Intell Syst <https://doi.org/10.1002/aisy.202401034>
- [155] McMillen P, Levin M. Collective intelligence: a unifying concept for integrating biology across scales and substrates. Commun Biol 2024;7(1):378.
- [156] Zhao X, Kang Z, Feng A, Levine S, Song D. Learning to reason without external rewards. 2025; <https://arxiv.org/abs/2505.19590>.
- [157] Wang L, Ma C, Feng X, Zhang Z, Yang H, Zhang J, et al. A survey on large language model based autonomous agents. Front Comput Sci 2024;18(6):186345. <https://doi.org/10.1007/S11704-024-40231-1>
- [158] Park JS, O'Brien JC, Cai CJ, Morris MR, Liang P, Bernstein MS. Generative agents: interactive simulacra of human behavior. In: Follmer S, Han J, Steimle J, Riche NH, (Eds.), Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, UIST 2023, San Francisco, CA, USA, 29 October 2023- 1 November 2023. ACM; 2023, p. 2:1–2:22. <https://doi.org/10.1145/3586183.3606763>
- [159] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, (Eds.), Advances in Neural Information Processing Systems; vol. 33. Curran Associates, Inc.; 2020, p. 9459–74. https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf.
- [160] Borgeaud S, Mensch A, Hoffmann J, Cai T, Rutherford E, Millican K, et al. Improving language models by retrieving from trillions of tokens. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S, (Eds.), Proceedings of the 39th International Conference on Machine Learning; vol. 162 of Proceedings of Machine Learning Research. PMLR; 2022, p. 2206–40. <https://proceedings.mlr.press/v162/borgeaud22a.html>.
- [161] Yao S, Zhao J, Yu D, Du N, Shafraan I, Narasimhan KR, et al. ReAct: synergizing reasoning and acting in language models. In: The eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1–5, 2023. OpenReview.net; 2023, https://openreview.net/forum?id=WE_vluYUL-X.
- [162] Mialon G, Dessì R, Lomeli M, Nalmpantis C, Pasunuru R, Railleanu R, et al. Augmented language models: a survey. Trans Mach Learn Res 2023;2023. <https://openreview.net/forum?id=jh7wH2AzKK>.
- [163] Shinn N, Cassano F, Gopinath A, Narasimhan K, Yao S. Reflexion: language agents with verbal reinforcement learning. In: Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S, (Eds.), Advances in Neural Information Processing Systems 36: annual conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10, - 16, 2023. 2023, http://papers.nips.cc/paper_files/paper/2023/hash/1b44b878bb782e6954cd888628510e90-Abstract-Conference.html.
- [164] Huang W, Xia F, Xiao T, Chan H, Liang J, Florence P, et al. Inner monologue: embodied reasoning through planning with language models. In: Liu K, Kulic D, Ichnowski J, (Eds.), Conference on Robot Learning, CoRL 2022, 14–18 December 2022, Auckland, New Zealand; vol. 205 of Proceedings of Machine Learning Research. PMLR; 2022, p. 1769–82. <https://proceedings.mlr.press/v205/huang23c.html>.
- [165] Li K, Hopkins AK, Bau D, Viégas FB, Pfister H, Wattenberg M. Emergent world representations: exploring a sequence model trained on a synthetic task. In: The eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1–5, 2023. OpenReview.net; 2023, https://openreview.net/forum?id=DeG07_TcZvT.
- [166] Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW, (Eds.), Advances in Neural Information Processing Systems; vol. 34. Curran Associates, Inc.; 2021, p. 8780–94. https://proceedings.neurips.cc/paper_files/paper/2021/file/49ad23d1ec9fa4bd8d77d02681df5cfa-Paper.pdf.
- [167] Nichol AQ, Dhariwal P. Improved denoising diffusion probabilistic models. In: Meila M, Zhang T, (Eds.), Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18–24 July 2021, virtual event; vol. 139 of Proceedings of Machine Learning Research. PMLR; 2021, p. 8162–71. <http://proceedings.mlr.press/v139/nichol21a.html>.
- [168] Song J, Meng C, Ermon S. Denoising diffusion implicit models. In: 9th International Conference on Learning Representations, ICLR 2021, virtual event, Austria, May 3–7, 2021. OpenReview.net; 2021, <https://openreview.net/forum?id=StlgiaRCHLP>.
- [169] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, (Eds.), Advances in Neural Information Processing Systems; vol. 33. Curran Associates, Inc.; 2020, p. 6840–51. https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bfc8e8584af0d967f1ab10179ca4b-Paper.pdf.
- [170] Sohl-Dickstein J, Weiss EA, Maheswaranathan N, Ganguli S. Deep unsupervised learning using nonequilibrium thermodynamics. In: Bach FR, Blei DM, (Eds.), Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015; vol. 37 of JMLR Workshop and Conference Proceedings. JMLR.org; 2015, p. 2256–65. <http://proceedings.mlr.press/v37/sohl-dickstein15.html>.
- [171] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, p. 10684–95.
- [172] Brooks T, Peebles B, Holmes C, DePue W, Guo Y, Jing L, et al. Video generation models as world simulators. OpenAI technical report; 2024. <https://openai.com/research/video-generation-models-as-world-simulators>.

- [173] Alonso E, Jelley A, Micheli V, Kanervisto A, Storkey AJ, Pearce T, et al. Diffusion for world modeling: visual details matter in Atari. In: Globersons A, Mackey L, Belgrave D, Fan A, Paquet U, Tomczak JM, et al., (Eds.), *Advances in Neural Information Processing Systems 38: annual conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10, - 15, 2024*. 2024. http://papers.nips.cc/paper_files/paper/2024/hash/6b8de0373d53d4a501249547084bed43-Abstract-Conference.html.
- [174] Yan X, Jin Y. EMO-DM: a diffusion model for evolutionary multi-objective optimization. 2024. <https://arxiv.org/abs/2401.15931>.
- [175] Krishnamoorthy S, Mashkaria SM, Grover A, et al. Diffusion models for black-box optimization. In: Krause A, Brunskill E, Cho K, Engelhardt B, Sabato S, Scarlett J, (Eds.), *Proceedings of the 40th International Conference on Machine Learning*; vol. 202 of *Proceedings of Machine Learning Research*. PMLR; 2023, p. 17842–57. <https://proceedings.mlr.press/v202/krishnamoorthy23a.html>.
- [176] Zhang Y, Hartl B, Hazan H, Levin M. Diffusion models are evolutionary algorithms. In: *The thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24–28, 2025*. OpenReview.net; 2025. <https://openreview.net/forum?id=xVefsBbG2O>.
- [177] Hartl B, Zhang Y, Hazan H, Levin M. Heuristically adaptive diffusion-model evolutionary strategy. *Adv Sci* 2026;13(20). <https://doi.org/10.1002/adv.202511537>
- [178] Jumper J, Evans R, Pritzel A, Green T, Figurnov N, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583–9. <https://doi.org/10.1038/s41586-021-03819-2>
- [179] Watson JL, Juergens D, Bennett NR, Trippie BL, Yim J, Eisenach HE, et al. De novo design of protein structure and function with RFdiffusion. *Nature* 2023;620(7976):1089–100. <https://doi.org/10.1038/s41586-023-06415-8>
- [180] Kenneweg P, Dandinasivara R, Luo X, Hammer B, Schönhuth A. Generating synthetic genotypes using diffusion models. *Bioinformatics* 2025;41:i484–i492. <https://doi.org/10.1093/bioinformatics/btaf209>
- [181] Li Z, Ni Y, Huygelen T, Das A, Xia G, Stan G-B, et al. Latent diffusion model for DNA sequence generation. In: *NeurIPS 2023 AI for science workshop*. 2023, <https://openreview.net/forum?id=lxymZngWen>.
- [182] Avdeyev P, Shi C, Tan Y, Dudnyk K, Zhou J. Dirichlet diffusion score model for biological sequence generation. In: Krause A, Brunskill E, Cho K, Engelhardt B, Sabato S, Scarlett J, (Eds.), *Proceedings of the 40th International Conference on Machine Learning*; vol. 202 of *Proceedings of Machine Learning Research*. PMLR; 2023, p. 1276–301. <https://proceedings.mlr.press/v202/avdeyev23a.html>.
- [183] Zhang T, Zhao Z, Ren J, Zhang Z, Zhang H, Wang G. cDiffusion: diffusion-based efficient generation of high quality scRNA-Seq data with classifier-free guidance. *Brief Bioinf* 2025;26(1):bbaf071. <https://doi.org/10.1093/bib/bbaf071>
- [184] DaSilva LF, Senan S, Patel ZM, Reddy AJ, Gabbita S, Nussbaum Z, et al. DNA-diffusion: leveraging generative models for controlling chromatin accessibility and gene expression via synthetic regulatory elements. *bioRxiv* <https://doi.org/10.1101/2024.02.01.578352>
- [185] Li K, Li J, Tao Y, Wang F. stDiff: a diffusion model for imputing spatial transcriptomics through single-cell transcriptomics. *Brief Bioinf* 2024;25(3):bbae171. <https://doi.org/10.1093/bib/bbae171>
- [186] Luo E, Hao M, Wei L, Zhang X. scDiffusion: conditional generation of high-quality single-cell data using diffusion model. *Bioinformatics* 2024;40(9):btae518. <https://doi.org/10.1093/bioinformatics/btae518>
- [187] Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods* 2018;15(12):1053–8. <https://doi.org/10.1038/s41592-018-0229-2>
- [188] Guo Z, Liu J, Wang Y, Chen M, Wang D, Xu D, et al. Diffusion models in bioinformatics and computational biology. *Nat Rev Bioeng* 2023;2(2):136–54. <https://doi.org/10.1038/s44222-023-00114-9>
- [189] Ambrogioni L. In search of dispersed memories: generative diffusion models are associative memory networks. *Entropy* 2024;26(5). <https://doi.org/10.3390/e26050381>
- [190] Raya G, Ambrogioni L. Spontaneous symmetry breaking in generative diffusion models. In: Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S, (Eds.), *Advances in Neural Information Processing Systems*; vol. 36. Curran Associates, Inc.; 2023, p. 66377–89. https://proceedings.neurips.cc/paper_files/paper/2023/file/d0da30e312b75a3ffid9e91f8bc1b0-Paper-Conference.pdf.
- [191] Ambrogioni L. The statistical thermodynamics of generative diffusion models: phase transitions, symmetry breaking, and critical instability. *Entropy* 2025;27(3). <https://doi.org/10.3390/e27030291>
- [192] Scolocchi A, Favero A, Wyart M. A phase transition in diffusion models reveals the hierarchical nature of data. *Proc Natl Acad Sci* 2025;122(1):e2408799121. <https://doi.org/10.1073/pnas.2408799121>
- [193] Hartl B, Risi S, Levin M. Evolutionary implications of self-assembling cybernetic materials with collective problem-solving intelligence at multiple scales. *Entropy* 2024;26(7). <https://doi.org/10.3390/e26070532>
- [194] Ho J, Salimans T. Classifier-free diffusion guidance. In: *NeurIPS 2021 workshop on deep generative models and downstream applications*. 2021, <https://openreview.net/forum?id=qw8AKxfYbl>.
- [195] Friston K, Levin M, Sengupta B, Pezzullo G. Knowing one's place: a free-energy approach to pattern regulation. *J R Soc Interface* 2015;12(105):20141383. <https://doi.org/10.1098/rsif.2014.1383>
- [196] Whitelam S. Generative thermodynamic computing. *Phys Rev Lett* 2026a;136(3). <https://doi.org/10.1103/kwyy-1xln>
- [197] Whitelam S. Training thermodynamic computers by gradient descent. *Proc Natl Acad Sci* 2026b;123(14). <https://doi.org/10.1073/pnas.2528413123>
- [198] Shing M, Koyama M, Akiba T. DiffusionBlocks: block-wise neural network training via diffusion interpretation. In: *The fourteenth International Conference on Learning Representations*. 2026, <https://openreview.net/forum?id=pwVSmK71cS>.
- [199] Hoover B, Liang Y, Pham B, Panda R, Strobel H, Chau DH, et al. Energy transformer. In: *Thirty-seventh conference on Neural Information Processing Systems*. 2023, <https://openreview.net/forum?id=MbwVNE9KS>.
- [200] Li X, Yeh A G-O. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *Int J Geogr Inf Sci* 2002;16(4):323–43. <https://doi.org/10.1080/13658810210137004>
- [201] Mordvintsev A, Randazzo E, Fouts C. Growing isotropic neural cellular automata. *The 2022 Conference on Artificial Life*, MIT Press; 2022. https://doi.org/10.1162/isal_a_00552
- [202] Langton CG. *Artificial life: an overview*. MIT Press; 1997.
- [203] Langton CG. Studying artificial life with cellular automata. *Phys D Nonlinear Phenom* 1986;22(1–3):120–49.
- [204] Von Neumann J, Burks AW, et al. Theory of self-reproducing automata. *IEEE Trans Neural Netw* 1966;5(1):3–14.
- [205] Gardner M. The fantastic combinations of John Conway's new solitaire game "Life". *Sci Am* 1970;223(4):120–3. <http://www.jstor.org/stable/24927642>.
- [206] Wolfram S. *A new kind of science*. Wolfram Media; 2002. <https://www.wolframscience.com>.
- [207] Randazzo E, Mordvintsev A, Fouts C. Growing steerable neural cellular automata. *The 2023 Conference on Artificial Life*, MIT Press; 2023. https://doi.org/10.1162/isal_a_00564
- [208] Cavuoti L, Sacco F, Randazzo E, Levin M. Adversarial takeover of neural cellular automata. *The 2022 Conference on Artificial Life*, MIT Press; 2022. https://doi.org/10.1162/isal_a_00521
- [209] Randazzo E, Mordvintsev A, Niklasson E, Levin M. Adversarial reprogramming of neural cellular automata. *Distill* 2021;6(5). <https://doi.org/10.23915/distill.00027.004>
- [210] Niklasson E, Mordvintsev A, Randazzo E, Levin M. Self-organising textures. *Distill* 2021;6(2). <https://doi.org/10.23915/distill.00027.003>
- [211] Randazzo E, Mordvintsev A, Niklasson E, Levin M, Greydanus S. Self-classifying MNIST digits. *Distill* 2020;5(8). <https://doi.org/10.23915/distill.00027.002>
- [212] Mordvintsev A, Randazzo E, Niklasson E, Levin M. Growing neural cellular automata. *Distill* 2020;5(2). <https://doi.org/10.23915/distill.00023>
- [213] Montero ML, Plantec E, Nisioti E, Pedersen JW, Risi S, et al. Meta-learning an evolvable developmental encoding. 2024. <https://doi.org/10.48550/arXiv.2406.09020>
- [214] Pio-Lopez L, Bischof J, LaPalme JV, Levin M. The scaling of goals from cellular to anatomical homeostasis: an evolutionary simulation, experiment and analysis. *Interface Focus* 2023;13(3):20220072. <https://doi.org/10.1098/rsfs.2022.0072>
- [215] Earle S, Yildiz O, Togelius J, Hegde C. Pathfinding neural cellular automata. 2023 <https://arxiv.org/abs/2301.06820>.

- [216] Katsuhiko Endo KY. Neural cellular maze solver. 2023. <https://umu1729.github.io/pages-neural-cellular-maze-solver/>
- [217] Hartl B, Levin M, Zöttl A. Neuroevolution of decentralized decision-making in N-bead swimmers leads to scalable and robust collective locomotion. *Commun. Phys.* 8 194 2025. <https://doi.org/10.1038/s42005-025-02101-5>.
- [218] Sudhakaran S, Grbic D, Li S, Katona A, Najarro E, Glanois C, et al. Growing 3D artefacts and functional machines with neural cellular automata. The 2021 Conference on Artificial Life, MIT Press; 2021. https://doi.org/10.1162/isal_a_00451
- [219] Pontes-Filho S, Walker K, Najarro E, Nichele S, Risi S. A unified substrate for body-brain co-evolution. In: From cells to societies: collective learning across scales. 2022. <https://openreview.net/forum?id=BcgXSzk6b5>.
- [220] Najarro E, Sudhakaran S, Glanois C, Risi S. HyperNCA: growing developmental networks with neural cellular automata. In: ICLR 2022 Workshop on from cells to societies: collective learning across scales; 2022. <https://doi.org/10.48550/ARXIV.2204.11674>
- [221] Pathak D, Lu C, Darrell T, Isola P, Efros AA. Learning to control self-assembling morphologies: a study of generalization via modularity. Red Hook, NY, USA: Curran Associates Inc.; 2019.
- [222] Pio-Lopez L, Hartl B, Levin M. Aging as a loss of goal-directedness: an evolutionary simulation and analysis unifying regeneration with anatomical rejuvenation. *Adv Sci* 2025;12(46):e09872.
- [223] Hartl B, Levin M, Pio-Lopez L. Neural cellular automata: applications to biology and beyond classical AI. *Phys Life Rev* 2026;56:94–108. <https://doi.org/10.1016/j.plrev.2025.11.010>
- [224] Bielawski K, Gaylinn N, Lunn C, Motia K, Bongard J, et al. Evolving hierarchical neural cellular automata. In: Proceedings of the Genetic and Evolutionary Computation Conference. Melbourne VIC Australia: ACM. ISBN 979-8-4007-0494-9; 2024, p. 78–86. <https://doi.org/10.1145/3638529.3654150>
- [225] Pande R, Grattarola D. Hierarchical neural cellular automata. The 2023 Conference on Artificial Life, MIT Press; 2023. https://doi.org/10.1162/isal_a_00601
- [226] Palm RB, Duque MG, Sudhakaran S, Risi S. Variational neural cellular automata. In: The tenth International Conference on Learning Representations, ICLR 2022, virtual event, April 25–29, 2022. OpenReview.net; 2022. <https://openreview.net/forum?id=7fFO4cMBx9>.
- [227] Kaplan R, Friston KJ. Planning and navigation as active inference. *Biol Cybern* 2018;112(4):323–43. <https://doi.org/10.1007/S00422-018-0753-2>
- [228] Da Costa L, Parr T, Sajid N, Veselic S, Neacsu V, Friston K. Active inference on discrete state-spaces: a synthesis. *J Math Psychol* 2020;99:102447. <https://doi.org/10.1016/j.jmp.2020.102447>
- [229] Gisin N. Indeterminism in physics, classical chaos and Bohmian mechanics: are real numbers really real? *Erkenntnis* 2019;86(6):1469–81. <https://doi.org/10.1007/s10670-019-00165-8>
- [230] Ma S, Wang H, Ma L, Wang L, Wang W, Huang S, et al. The era of 1-bit LLMs: all large language models are in 1.58 bits. 2024. <https://arxiv.org/abs/2402.17764>.
- [231] Nanda N, Lee A, Wattenberg M. Emergent linear representations in world models of self-supervised sequence models. In: Belinkov Y, Hao S, Jumelet J, Kim N, McCarthy A, Mohebbi H, (Eds.), Proceedings of the 6th BlackboxNLP workshop: analyzing and interpreting neural networks for NLP, BlackboxNLP@EMNLP 2023, Singapore, December 7, 2023. Association for Computational Linguistics; 2023, p. 16–30. <https://doi.org/10.18653/v1/2023.BLACKBOXNLP-1.2>
- [232] Gurnee W, Tegmark M. Language models represent space and time. In: The twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7–11, 2024. OpenReview.net; 2024. <https://openreview.net/forum?id=jE8xbmvFin>.
- [233] y Arcas BA. What is intelligence?: Lessons from AI about evolution, computing, and minds. MIT Press; 2025.
- [234] Lindsey J. Emergent introspective awareness in large language models. 2026. <https://doi.org/10.48550/ARXIV.2601.01828>
- [235] von Oswald J, Niklasson E, Randazzo E, Sacramento J, Mordvintsev A, Zhmoginov A, et al. Transformers learn in-context by gradient descent. In: Krause A, Brunskill E, Cho K, Engelhardt B, Sabato S, Scarlett J, (Eds.), International Conference on Machine Learning, ICML 2023, 23–29 July 2023, Honolulu, Hawaii, USA. Proceedings of Machine Learning Research; PMLR; 2023, p. 35151–74. <https://proceedings.mlr.press/v202/von-oswald23a.html>.
- [236] Dai D, Sun Y, Dong L, Hao Y, Ma S, Sui Z, et al. Why can GPT learn in-context? Language models secretly perform gradient descent as meta-optimizers. In: Rogers A, Boyd-Graber JL, Okazaki N, (Eds.), Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9–14, 2023. Findings of ACL; Association for Computational Linguistics; 2023, p. 4005–19. <https://doi.org/10.18653/v1/2023.FINDINGS-ACL.247>
- [237] Shi H, Xu Z, Wang H, Qin W, Wang W, Wang Y, et al. Continual learning of large language models: a comprehensive survey. *ACM Comput Surv* 2026;58(5):120:1–120:42. <https://doi.org/10.1145/3735633>
- [238] Pfeifer R, Bongard J. How the body shapes the way we think: a new view of intelligence. MIT Press; 2006.
- [239] Cangelosi A, Bongard J, Fischer MH, Nolfi S. Embodied intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-662-43505-2; 2015, p. 697–714. https://doi.org/10.1007/978-3-662-43505-2_37
- [240] Bongard J, Levin M. There's plenty of room right here: biological systems as evolved, overloaded, multi-scale machines. *Biomimetics* 2023;8(1). <https://doi.org/10.3390/biomimetics8010110>
- [241] Landauer R. Irreversibility and heat generation in the computing process. *IBM J Res Dev* 1961;5(3):183–91. <https://doi.org/10.1147/rd.53.0183>
- [242] Jiang Y, Zhang Y, Shen X, Backes M, Zhang Y. “Humans welcome to observe”: a first look at the agent social network Moltbook. 2026. <https://arxiv.org/abs/2602.10127>.
- [243] Holtz D. The anatomy of the Moltbook social graph. 2026. <https://arxiv.org/abs/2602.10131>.
- [244] De Marzo G, Garcia D. Collective behavior of AI agents: the case of Moltbook. 2026. <https://arxiv.org/abs/2602.09270>.
- [245] Durand EA, Joffily M, Khamassi M. A diffusion model-based approach to active inference. In: 2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS). 2024, p. 75–80. <https://doi.org/10.1109/FMLDS63805.2024.00024>
- [246] Yokozawa R, Fujii K, Nomura Y, Murata S. Deep active inference with diffusion policy and multiple timescale world model for real-world exploration and navigation. 2025. <https://doi.org/10.48550/ARXIV.2510.23258>
- [247] Levin M. The computational boundary of a “Self”: developmental bioelectricity drives multicellularity and scale-free cognition. *Front Psychol* 2019;10. <https://doi.org/10.3389/fpsyg.2019.02688>
- [248] Hengen KB, Shew WL. Is criticality a unified setpoint of brain function? *Neuron* <https://doi.org/10.1016/j.neuron.2025.05.020>
- [249] Dunkelmann S, Radons G. Neural networks and Abelian sandpile models of self-organized criticality. In: Proceedings of International Conference Artificial Neural Networks. Springer-Verlag; 1994, p. 867–70.
- [250] Hesse J, Gross T. Self-organized criticality as a fundamental property of neural systems. *Front Syst Neurosci* 2014;8. <https://doi.org/10.3389/fnsys.2014.00166>
- [251] Beggs JM, Timme N. Being critical of criticality in the brain. *Front Physiol* 2012;3. <https://doi.org/10.3389/fphys.2012.00163>
- [252] Pezzulo G, Parr T, Friston KJ. Shared worlds, shared minds. *EMBO Rep* 2025;26(17):4197–202. <https://doi.org/10.1038/s44319-025-00549-8>
- [253] Bak P, Tang C, Wiesenfeld K. Self-organized criticality. *Phys Rev A* 1988;38(1):364–74. <https://doi.org/10.1103/physrev.38.364>
- [254] Bak P, Chen K. Self-organized criticality. *Sci Am* 1991;264(1):46–53. <http://www.jstor.org/stable/24936753>.
- [255] Shin D, Cho K-H. Critical transition and reversion of tumorigenesis. *Exp Mol Med* 2023;55(4):692–705. <https://doi.org/10.1038/s12276-023-00969-3>
- [256] Xiong R, Yang Y, He D, Zheng K, Zheng S, Xing C, et al. On layer normalization in the transformer architecture. In: Daumé IH, Singh A, (Eds.), Proceedings of the 37th International Conference on Machine Learning; vol. 119 of Proceedings of Machine Learning Research. PMLR; 2020, p. 10524–33. <https://proceedings.mlr.press/v119/xiong20b.html>.
- [257] Alekseev S. Subcritical signal propagation at initialization in normalization-free transformers. 2026. <https://arxiv.org/abs/2604.11890>.
- [258] Gottesman S. Trouble is coming: signaling pathways that regulate general stress responses in bacteria. *J Biol Chem* 2019;249:11685–700.
- [259] Flombaum JL, Scholl BJ, Santos LR. Spatiotemporal priority as a fundamental principle of object persistence. Oxford University Press; 2008, p. 135–64.
- [260] Farinella-Ferruzza N. The transformation of a tail into limb after xenoplastic transplantation. *Experientia* 1956;12(8):304–5. <https://doi.org/10.1007/bf02159624>
- [261] Vandenberg LN, Adams DS, Levin M. Normalized shape and location of perturbed craniofacial structures in the *Xenopus* tadpole reveal an innate ability to achieve correct morphology. *Dev Dyn* 2012;241(5):863–78. <https://doi.org/10.1002/dvdy.23770>
- [262] Lobo D, Solano M, Bubenik GA, Levin M. A linear-encoding model explains the variability of the target morphology in regeneration. *J R Soc Interface* 2014;11(92):20130918.

- [263] Emmons-Bell M, Durant F, Hammelman J, Bessonov N, Volpert V, Morokuma J, et al. Gap junctional blockade stochastically induces different species-specific head anatomies in genetically wild-type *girardia dorotocephala* flatworms. *Int J Mol Sci* 2015;16(11):27865–96. <https://doi.org/10.3390/ijms161126065>
- [264] Durant F, Bischof J, Fields C, Morokuma J, LaPalme J, Hoi A, et al. The role of early bioelectric signals in the regeneration of planarian anterior/posterior polarity. *Biophys J* 2019;116(5):948–61. <https://doi.org/10.1016/j.bpj.2019.01.029>
- [265] Levin M. Self-improvising memory: a perspective on memories as agential, dynamically reinterpreting cognitive glue. *Entropy* 2024;26(6). <https://doi.org/10.3390/e26060481>
- [266] Levin M. Life, death, and self: fundamental questions of primitive cognition viewed through the lens of body plasticity and synthetic organisms. *Biochem Biophys Res Commun* 2021;564:114–33. Rethinking Cognition: From Animal to Minimal; <https://doi.org/10.1016/j.bbrc.2020.10.077>
- [267] Ananthaswamy A. A new approach to computation reimagines artificial intelligence. *Quanta Mag* 2023; Accessed: 2025-12-08; <https://www.quantamagazine.org/a-new-approach-to-computation-reimagines-artificial-intelligence-20230413/>.
- [268] Alammr J. The illustrated transformer. 2018. Accessed: 2025-12-08. <https://jalammar.github.io/illustrated-transformer/>.
- [269] Pennington J, Socher R, Manning CD. GloVe: global vectors for word representation. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 2014, p. 1532–43.
- [270] Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. *Trans Assoc Comput Linguist* 2017;5:135–46.
- [271] Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep contextualized word representations. In: Walker MA, Ji H, Stent A, (Eds.), *Proceedings of the 2018 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1–6, 2018, volume 1 (Long papers)*. Association for Computational Linguistics; 2018, p. 2227–37. <https://doi.org/10.18653/V1/N18-1202>
- [272] McCann B, Bradbury J, Xiong C, Socher R. Learned in translation: contextualized word vectors. In: Guyon I, von Luxburg U, Bengio S, Wallach HM, Fergus R, Vishwanathan SVN, et al., (Eds.), *Advances in Neural Information Processing Systems 30: annual conference on Neural Information Processing Systems 2017, December 4–9, 2017, Long Beach, CA, USA*. 2017, p. 6294–305. <https://proceedings.neurips.cc/paper/2017/hash/20c86a628232a67e7bd46f76fba7ce12-Abstract.html>.
- [273] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1 (long and short papers)*. 2019, p. 4171–86.
- [274] Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, et al. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Jurafsky D, Chai J, Schluter N, Tetreault JR, (Eds.), *Proceedings of the 58th annual meeting of the Association for Computational Linguistics, ACL 2020, online, July 5–10, 2020*. Association for Computational Linguistics; 2020, p. 7871–80. <https://doi.org/10.18653/V1/2020.ACL-MAIN.703>
- [275] Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. Technical report, OpenAI; 2018. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
- [276] Spot Intelligence. Embeddings from language models - ELMo and beyond. 2023. Accessed: 2025-12-08. <https://spotintelligence.com/2023/12/26/embeddings-from-language-models-elmo/>.
- [277] McCloud S. *Understanding comics: the invisible art*. Tundra Publishing; 1993. https://en.wikipedia.org/wiki/Understanding_Comics.
- [278] Kingma DP, Welling M. Auto-encoding variational Bayes. In: *2nd International Conference on Learning Representations (ICLR)*. 2014.
- [279] de Lamo CV, Gidey HK, Lenz A, Knoll A. Fundamentals of building autonomous LLM agents. 2025. <https://doi.org/10.48550/ARXIV.2510.09244>
- [280] Pezzulo G, LaPalme J, Durant F, Levin M. Bistability of somatic pattern memories: stochastic outcomes in bioelectric circuits underlying regeneration. *Philos Trans R Soc Lond B Biol Sci* 2021;376(1821):20190765. <https://doi.org/10.1098/rstb.2019.0765>