Why AI systems don't learn and what to do about it

Lessons on autonomous learning from cognitive science

Emmanuel Dupoux[1,2], **Yann LeCun**[3], **Jitendra Malik**[1,4]

[1]FAIR at META, [2]École des Hautes Études en Sciences Sociales, [3]NYU, [4]UC Berkeley

We critically examine the limitations of current AI models in achieving autonomous learning and propose a learning architecture inspired by human and animal cognition. The proposed framework integrates learning from observation (System A) and learning from active behavior (System B) while flexibly switching between these learning modes as a function of internally generated meta-control signals (System M). We discuss how this could be built by taking inspiration on how organisms adapt to real-world, dynamic environments across evolutionary and developmental timescales.

Date: March 17, 2026
Correspondence: dpx at @meta.com

Meta

Both AI and Cognitive Science emerged in the 1950's in the post-war intellectual ferment which brought together neural modeling, computation, information, and control. While the objectives differ –creating intelligent machines vs. scientific understanding of brains and behavior, the intellectual trajectories of AI and cognitive science have overlapped with varying degree of cross-fertilization. Today, the successes of Deep Learning ushered an era of deeper cross-disciplinary interactions. As AI models tackle high level human abilities like language, visual understanding and reasoning, they incorporate concepts and evaluation methods from the cognitive and neural sciences. Conversely, AI systems provide sorely needed quantitative theories of cognitive processes that can be tested against empirical data. *Paradoxically, given the importance of deep learning, one key component of human intelligence remains out of reach for current AI models: the ability to learn as humans do.*

Forewords

The dominant AI research paradigm today relies on hyperscaling of text-based LLMs with ever larger models, data and compute. But even prominent architects of this approach such as Ilya Sutskever[a] and Andrei Karpathy [b] suggest we may be hitting diminishing returns. Areas of concern include (1) confronting the "data wall" on quality text data (2) inability to learn new things beyond current human knowledge because of the absence of interaction with the environment (Silver and Sutton, 2025) (3) excessively language-centrism as opposed to spatial, embodied and grounded reasoning in the physical world (4) lack of continual life-long learning (self-improvement after deployment). While these critiques echo long standing controversies within cognitive science on the non-verbal cognition (Johnson-Laird, 1983), and situated interactions (Piaget, 1952; Vygotsky and Cole, 1978) in intelligence, it behooves us as scientists to take stock of progress from both fields and look beyond the current paradigm. What could come next?

[a]Business insider
[b]Medium

1 What is autonomous learning?

Consider the distinction between children and current AI models. Children learn and act from birth. They flexibly choose what to attend, what to learn, when to act or observe, and more generally how to switch between different learning modes (Botvinick et al., 2019; Shenhav et al., 2017). For example, a toddler trying a new toy may explore it randomly (*learning through action*; Gopnik et al. 2017), or by watching a peer, attempt to imitate the goal or gesture, depending on context (*learning through observation*; Gergely et al. 2002; Tomasello 1999). They may follow a caretaker's verbal instruction on how to use the toy (*learning through communication*; Csibra and Gergely 2009), or take a pause and daydream about the various ways to use the toy (*learning through imagination*; Redshaw and Suddendorf 2016).

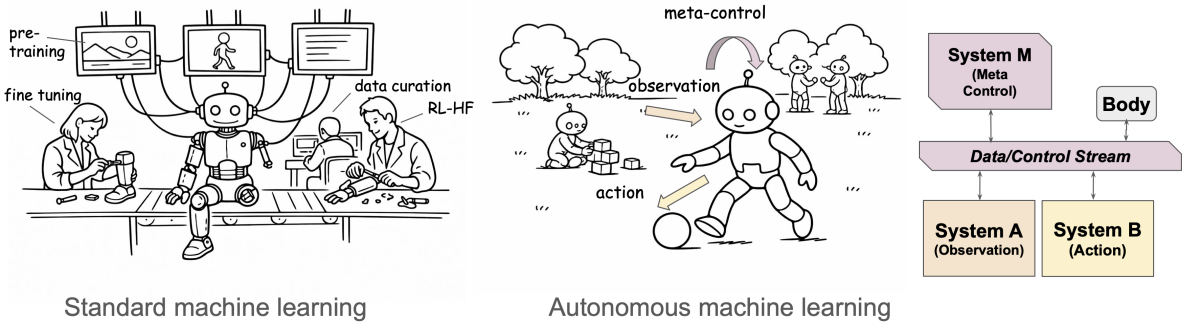

Figure 1 Standard machine learning (left). The machine does not learn by itself; it requires an assembly line of research engineers and data scientists collecting, formatting, and curating different kinds of data, each used to train successively different components of the model, each with specificly engineered loss and reward functions. The machine is then left with no ability to learn from its experience. **Autonomous machine learning (right).** The agent is learning directly in interaction with the world; the sources of data are generarated by the agent itself through different learning modes (learning by observation, by action, which can be extended to higher modes like learning by verbal interaction or self-play). Our proposed architecture include a meta controler enabling learning while operating in the real world. *(Drawings from ChatGPT).*

In contrast, AI models, once deployed, learn essentially *nothing*; their mode of operation is fixed, and if not adapted to their environment, a new model has to be rebuilt using new data by human experts-in-the-loop (Hadsell et al., 2020). Furthermore, the different learning modes exemplified in children are typically siloed into distinct machine learning paradigms (e.g., self-supervised learning, supervised learning, reinforcement learning), each requiring specific data curation pipelines and training recipes; when the different modes are mixed, it is mainly through rigid sequences of training recipes established through trial and error by human experts and tuned to particular applications (chatbots, coding assistants, etc.). In other words, in current AI systems, learning is *outsourced* to human experts instead of being an *intrinsic* capability.

Arguably, the inability to learn may explain some difficulties of AI systems to be deployed in real life. AI systems are built by optimizing an objective over a fixed set of *training data*, typically lifted from the internet. However, once deployed in real life, this system may be confronted with new data that diverge significantly from this distribution, with unpredictable consequences. This phenomenon known as *domain mismatch* cannot be fixed by merely increasing the training set size, as real-life data always contains new, unseen cases (*heavy tailed*) and keeps changing over time (*non-stationarity*) (Geirhos et al., 2020; Koh et al., 2021). Modern AI addresses domain mismatch by breaking down model training into two phases: *pretraining* using large generic datasets, and *fine tuning* using data more appropriate to the target application (Bommasani et al., 2022). This is a first step in the right direction, but it still requires considerable human involvement, and there is no guarantee that this will work, as the system has not been primarily built to be fine-tunable or adaptable, especially on unfiltered raw data. In contrast, in biological organism, domain mismatch is mitigated by enabling the agent to learn and adapt from the data directly available in its environment, allowing for species-specific cognitive adaptability. Humans are particularly adaptable in this respect, having shown the ability to spread very quickly to a variety of different ecological niches (Boyd et al., 2011).

In this paper, we start from the idea that autonomous learning should be considered a *core capability*, essential for building reliable AI systems that can operate in the real world. Conversely, we take it that the development of adaptable AI systems can benefit cognitive science by providing quantitative models capable of addressing long-standing debates about the nature and origins of human intelligence. The main contributions of the paper are the following: we identify three conceptual and technical roadblocks that have so far limited the development of autonomous learning, and we propose possible diractions to address them. This should be as a high level roadmap which we hope will be useful in inspiring future cross-disciplinary work and collaborations.

The first roadblock is conceptual: existing approaches to learning remain *fragmented* across subfields, making it difficult to integrate them within a unified framework. In Section 2, we argue that a path for integration is first to recognize two fundamental learning modes: learning through observation (*System A*) and learning through action (*System B*), and then classify the different ways in which these modes can interact with

one another. The second roadblock is the *externalisation* of learning which is currently practiced in AI. To address this, we propose in Section 3 a *meta-control architecture (System M)* that coordinates information flow between the learning components enabling to reproduce automatically the learning and data filtering recipes typically done by hand. We show that such system open up the possibility of higher order learning modes only found in some large brain species, like learning by through communication and imagination. The third roadblock is the lack of effective methods to build such architectures at scale. In Section 4, we propose an *evolutionary* inspired *bilevel optimization* approach to jointly learn the meta-control model and the initial states of the system A and B components to achieve robust real-world behavior. In Section 5, we conclude by reviewing recent progress and outlining promising directions for research at the interface of AI and cognitive science in this emerging area of autonomous learning.

2 Integrating Observation and Action

The ability to learn has been studied within a rich set of traditions and methods in the natural, social, and formal sciences. For lack of space, we will not attempt even a cursory review of this vast topic here, but rather point out one major fault line across these fields that separates what can be called *observation-based* versus *action-based* learning. In the first case, the organism is conceived passively accumulating sensory input, and learning through building a statistical or a predictive model of its input data (Saffran et al., 1996; Rao and Ballard, 1999). We call the set of learning mechanisms that fall in this bucket: *System A*. In the second case, the organism is conceived as an agent interacting in the world and trying to achieve a particular goal through the adjustment of its actions against the observed feedback from the environment (Sutton and Barto, 2018; Schultz et al., 1997). We call the relevant set of mechanisms *System B*[1].

For methodological or historical reasons, these two conceptions have yielded distinct subfields within each of the relevant sciences, with little interaction between them (even using different terminology, as in Table 1 for System A). In the following sections, we describe their strengths and limitations and then outline why they need to be combined to account for how learning really takes place in living organisms.

2.1 System A: Learning from Observation

From a cognitive point of view, System A learning situations abound in early childhood where infants' abilities to act on the world are limited. For instance, infants initially can discriminate faces from multiple species at 6 months (e.g., human and monkey faces), but by 9 months they become specialized for human faces and lose sensitivity to non-human faces (Pascalis et al., 2002). Similarly, while newborns can distinguish phonetic contrasts from many languages, between 6 and 12 months their perception improves for the sounds of their native language and degrades for non-native ones (Werker and Tees, 1984). Infants are thought to learn a model of the world through observation, making them able to predict the future based on past observations: they are surprised by 'impossible' events or magic tricks (Spelke et al., 1992; Baillargeon, 2004; Spelke and Kinzler, 2007). All of these phenomena have been attributed to learning mechanisms capable of extracting the statistical distribution of sensory stimuli and/or predicting future stimuli conditioned on past events, both being instances of, both being instance of System A algorithms (Saffran and Kirkham, 2018).

From an AI perspective, such algorithms include Self-Supervised Learning (SSL) models applied to static datasets or passively collected sensory streams (Chen et al., 2020; He et al., 2020; Devlin et al., 2019). These paradigms can be classified based on their modality, data type, and structure. Some systems operate on a single modality such as text, images, or audio, while others combine modalities, for instance, vision and language (Radford et al., 2021). Some work on symbolic, discrete data like tokens, while others learn directly from continuous sensory input. Finally, some models explicitly exploit the spatial or sequential structure of their inputs, such as grids or time series.

Abstractly, one can describe this class of algorithms in the following way. Let data come from a distribution $\mathcal{D}$ ($x \sim \mathcal{D}$). We define a *task generator* $\mathcal{G}$ that, given a raw sample x, produces an input–target pair:

[1]This terminology is reminescent of Kahneman's System 1 and 2, but is actually orthogonal to it, see Appendix A for a discussion.

Table 1 Sample observational learning problems (System A) from different sensory inputs, proposed mechanisms with corresponding terminology in Cognitive Science and AI fields and sample AI algorithms (non-exhaustive).

Input	Task	Proposed Learning Mechanisms		(sample) AI Algorithms
		CogSci terminology	AI terminology	
Speech	Learning phonetic categories	Distributional Learning / Perceptual Learning	Self-Supervised Learning / Acoustic Unit Discovery	CPC (van den Oord et al., 2018), HuBERT (Hsu et al., 2021)
Language	Syntax acquisition	Statistical Learning	Language Modeling	GPT (Radford et al., 2018), BERT (Devlin et al., 2019), GSLM (Lakhotia et al., 2021)
Images	Face recognition, Object categorization	Perceptual Learning	Self-Supervised Learning	SimCLR (Chen et al., 2020), MoCo (He et al., 2020), DINO (Caron et al., 2021), I-JEPA (Assran et al., 2023)
Video	Intuitive physics	Perceptual Learning	Predictive World Modeling	PredNet (Lotter et al., 2017), V-JEPA (Bardes et al., 2024)
Language + Vision	Learning words and sentence meaning	Cross-situational Learning	Multimodal Language Modeling	CLIP (Radford et al., 2021), Flamingo (Alayrac et al., 2022), Transfusion (Zhou et al., 2024)

$$(x_{\text{in}}, x_{\text{tar}}) = \mathcal{G}(x) \tag{1}$$

We want to learn a representation $z = f_\theta(x)$ with parameters θ that minimizes a loss. The training objective is:

$$\theta^\star = \arg\min_\theta \; \mathbb{E}_{x\sim\mathcal{D}} \; \mathcal{L}\big(f_\theta(x_{\text{in}}), x_{\text{tar}}\big) \tag{2}$$

System A has several strengths. It scales well with large datasets and is capable of discovering abstract latent representations that can be organized hierarchically across different levels of abstraction, from low-level sensory features to high-level conceptual categories (Bengio et al., 2013), and can support robust transfer to downstream tasks (Devlin et al., 2019; Caron et al., 2021).

However, it also faces limitations, as is apparent in the formulations in Equations 1 and 2. System A models need access to $\mathcal{D}$ and a task generator $\mathcal{G}$, both of which typically require considerable human expertise, and have to be tailored to each domain with care (Tian et al., 2020). They lack any built-in mechanism to decide what data might be useful or should be acquired next (Settles, 2009). Furthermore, their representations are disconnected from the agent's ability to act, making it hard to ground what they learn in real-world behavior (Bisk et al., 2020). Last but not least, because they are based purely on observation, they struggle to distinguish between correlation and causation (Schölkopf, 2019).

2.2 System B: Learning from Action

From a cognitive point of view, typical System B learning situations can be found in basic motor learning in children, like learning to walk. Here, observation of the world does not necessarily help, as children do not initially attempt to imitate other agents' mode of locomotion (Adolph et al., 2012). Rather, through trial and error they go through various stages using non-bipedal modes (rolling, crawling), before they develop the

ability to stand and take a few wobbling steps and finally develop a mature gait. To a certain extent, vocal learning follows a similar path, whereby initial vocal explorations by infants are not similar to their linguistic input, and even arise in children with hearing loss (Oller and Eilers, 1988).

From a machine learning point of view, the class of System B algorithms comprises learning mechanisms that operate through interaction. Acting is to intervene on the world through a sequence of *actions* a_t, to reach a given *goal* (optimizing some *reward* r over some *time horizon* T). The world is characterized in terms of its *states* s_t, and *transition dynamics* $M(s_{t+1}|s_t, a_t)$. The transition dynamics of the agent is called a *policy* (π). If the world dynamics were totally known in advance and relatively simple, the optimal sequence of actions could be derived mathematically without any learning all (as in *control theory*, Bertsekas 2019). If the world dynamics are unknown, then the system has to learn about it to optimize its actions (as for *reinforcement learning* and *planning*; Sutton and Barto 2018; Russell and Norvig 2020; Moerland et al. 2023). The general problem can be formulated:

$$\text{Maximize } J(\pi) = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r(s_t, a_t)\right] \tag{3}$$

Where:

$s_{t+1} \sim M(.\|s_t, a_t)$	Dynamic model of the world
$a_t \sim \pi(.\|s_t)$	Agent's policy
$r(s, a)$	Reward function
$\gamma \in [0, 1)$	Discount factor

Table 2 Popular paradigms addressing System B optimization problems, differing in how the world model M and the policy π are defined, whether they are learned from data in a training phase, and how they are used at inference time, with selected examples of applications in the modeling of animal behavior.

Paradigm	**W. Model M (train., infer.)**	**Policy π (train., infer.)**	**When is optimization solved?**	**Example**
Control Theory	fixed (no,no)	derived analytically (no, direct)	offline (design time)	Spinal reflexes, saccadic eye movement (Robinson, 1981)
Adaptive Control	adaptive param. (yes, no)	derived analytically (no, direct)	online (parameter estimation)	Motor adaptation (Shadmehr and Mussa-Ivaldi, 1994)
Model-Free RL	none	NN (yes, direct)	training time (compilation of experience into reactive policy)	{Habitual actions (Schultz et al., 1997)
Model-Based RL	NN (yes, no / unrolled)	NN (yes, direct / search)	both training and inference	Goal-directed behaviors, foraging strategies (Daw et al., 2011)
Planning	simulator / NN (no/yes, unrolled)	computed online (no,search)	inference time (thinking before acting)	Detour planning, mental simulation (Pfeiffer and Foster, 2013)

Methods to optimize the objective function depend on whether the world transition dynamics and reward function are known or must be learned, and whether one searches through a space of actions, or learns a policy that directly predicts the next action to be taken given the world state (see Table 2). These systems also vary in their reward sources, which may be provided externally through task-specific signals or generated internally through curiosity, novelty, or empowerment (Schmidhuber, 1991; Pathak et al., 2017; Mohamed and Rezende, 2015). The size and structure of the action space also vary widely. Simple environments use a small set of discrete actions, but real-world tasks often require complex, continuous, and high-dimensional action

spaces (Lillicrap et al., 2016). Exploration strategies can be random, curiosity-driven, or guided by a goal or policy (Ecoffet et al., 2021).

System B has important strengths: it is grounded in control and interaction, enabling it to learn directly from sparse or delayed outcomes, making it naturally suited for real-time and adaptive behavior. It can also discover truly novel solutions via search (Silver et al., 2017). However, it also faces major limitations. Primarily, it is notoriously sample-inefficient, often requiring large numbers of interactions to learn even simple tasks (Dulac-Arnold et al., 2021). It struggles in high-dimensional or open-ended action spaces. Furthermore, it depends on having well-specified reward functions and interpretable actions, which are rarely available in naturalistic settings (Amodei et al., 2016).

2.3 System A Helping System B

At an intuitive level, learning through action is easy when the number of possible actions is limited, and the world states are easy to track. This is typically the case in games like chess or video games. This is less the case in real life where the action space grows exponentially with the number of degrees of freedom (roughly 200 to 300 for robotics or animation), and world states are virtually unlimited. In humans and animals, one dominant idea is that the sensory/motor system provides a kind of curriculum by limiting the resolution of sensors at birth (children are myopic) and the effective degrees of freedom (with very synergistic muscles) (Turkewitz and Kenny, 1982; Bernstein, 1967). Even there, the search space is vastly larger than in a game of chess, and this is where System A can help by providing compressed representations for states and actions, predictive world models, and intrinsic reward signals that would make learning and planning more tractable (Ha and Schmidhuber, 2018; Yarats et al., 2020).

Abstract representations of states and actions. By observing the world as experienced by the agent, System A can learn abstract representations of observations through SSL methods. This can be used as a proxy for the representation of *world states* that are more abstract and more compact than raw sensory data (pixels or sound waves). For instance, CURL (Laskin et al., 2020) uses contrastive vision pretraining to derive a compact representation from pixels that can be fed to an RL agent learning policies for Atari games, with performance equivalent to an agent trained on hand-coded world states. (see also ACT; Zhao et al. 2023). Similarly, instead of working from raw pixels, many robotics papers leverage a pretrained vision encoder to provide useful features (Nair et al., 2022; Radosavovic et al., 2022).

Similarly, by observing sequences of actions taken by the agent, SSL techniques can yield abstract representations of the *action space* or group successive actions into skills (see DIAYN, Eysenbach et al. 2019; action chunking, Li et al. 2025). These principles are applied to robotics to reduce the dimensionality of action space (eg. CLAM, Liang et al. 2025). In Radosavovic et al. (2024b,a), generatively modeling sensorimotor trajectories in humanoid locomotion (in simulation), facilitates subsequent RL and application to real ,challenging terrain.

Forward looking, System A can learn latent action spaces from unlabelled video (eg. LAWM; Tharwat et al. 2025) or infer *goals* or *rewards* from raw videos, enabling inverse RL and direct imitation of tasks or goals (Sermanet et al., 2017; Ma et al., 2023).

Predictive World Models. Among System A models, predictive models learn to predict future states based on past states, capturing the dynamics of the environment. This addresses a critical problem in (model-free) RL, which is the combinatorial explosion of the search space. Predictive models, when conditionned on self-actions can turn System B into model-based RL, enabling planning instead of blind trial-and-error.

Notable models include PlaNet (Hafner et al., 2019), Dreamer (Hafner et al., 2020) and SPR (Schwarzer et al., 2021). The ability to learn an internal simulation that can fully replace an external environment has been demonstrated in video games and Go with Mu-zero Schrittwieser et al. (2020). The scaling of these ideas to more complex environment is still underway. Predictive World models are split between pixel-based generative models (e.g., Genie (Bruce et al., 2024), Gato (Reed et al., 2022)), and Video Joint Embedding Predictive Architecture (V-JEPA) that predict in latent space (Bardes et al., 2024), enabling better modeling of physics (Garrido et al., 2025), and showing fast transfer to robotics applications (Assran et al., 2025).

Intrinsic reward signals. Reinforcement Learning hasclassically been confronted with the exploration/exploitation dilemma, where exploitation tries to optimize immediate reward, but may fail because of incomplete

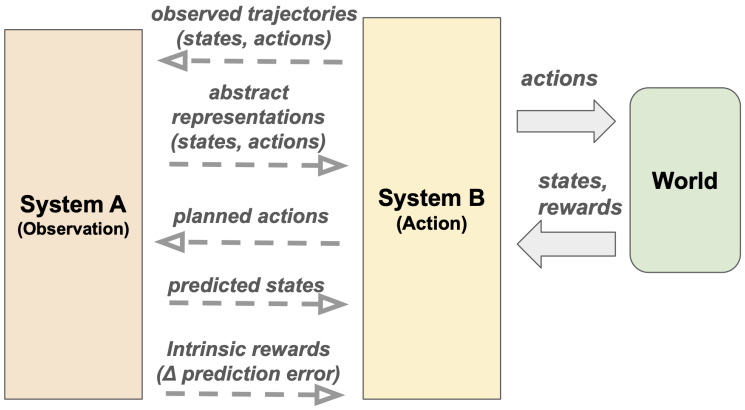

Figure 2 Summary of modes of interactions between Systems A and B. System A provides System B with predictions of future states conditioned on past states and actions, with hierarchical abstractions over possible actions, and a SSL loss that can be used for curiosity/exploration. System B through its action provides rich and task relevant input for System A to learn from.

knowledge about the effect of actions on the world, and exploration improves world knowledge but may delay immediate reward. System A can help by providing intrinsic reward signals like prediction errors, uncertainty, or novelty, enabling the agent to explore efficiently, and shift to exploitation once it is confident (Oudeyer et al., 2007; Schmidhuber, 2010; Aubret et al., 2023). These ideas have been applied to the video game domains (e.g., Pathak et al. 2017; Badia et al. 2020; Kayal et al. 2025), but applications to robotics remain limited (Tang et al., 2025; Taylor et al., 2021).

Each of these mechanisms reduces the burden on System B by simplifying the search space of model-free RL, by reducing the dimensionality of representations, providing forward models to guide search, and exploration rewards to reduce uncertainty. Much remains to be done in scaling these different approaches to real-life situations.

2.4 System B Helping System A

System A's primary limitation is its reliance on passive or static data. Without guidance or data curation, it may fail to learn useful representations from uninformative, noisy, or irrelevant data streams (Lavechin et al., 2023; Oquab et al., 2024; Gadre et al., 2023). But even human newborns are not passive observers: they can affect their data stream by orienting their visual and auditory attention to specific parts of the sensory field (e.g., orientation towards faces or speech sounds; Morton and Johnson 1991; Vouloumanos and Werker 2007). As they grow in motor autonomy, they are more and more able to seek out specific sources of information useful for their goals. System B, through active behavior, can help collect better data and provide grounding for learned representations. Gibson (1966)'s notion of active perception: "We see in order to move and we move in order to see" is a classic statement of the active gathering of information driven by a goal.

System B can support System A learning in two fundamental ways: directly, by helping to optimize System A's own predictive objectives (*active SSL*), or indirectly, by exploring the environment in ways that yield task-relevant or informative trajectories (*goal-directed SSL*).

Active self-supervised learning. System B is explicitly optimizing System A's ability to represent or generalize. For instance, through eye or head movement, System B can select a particularly 'interesting' portion of the sensory data to learn from, 'interesting' being defined by System A itself such as uncertainty, prediction error, or learning progress (Gottlieb et al., 2013; Oudeyer and Kaplan, 2007). This resembles curriculum learning or active data selection, but achieved through actions in the world (Smith et al., 2018). This logic can be extended to interventions that help disambiguate perception by revealing causal relationships that would be missed through passive observation (Agrawal et al., 2016).

Goal-directed self-supervised learning. System B optimizes its own task-related reward, and provides data to System A as a by-product. This results in data that may not be optimized for System A's loss but nonetheless

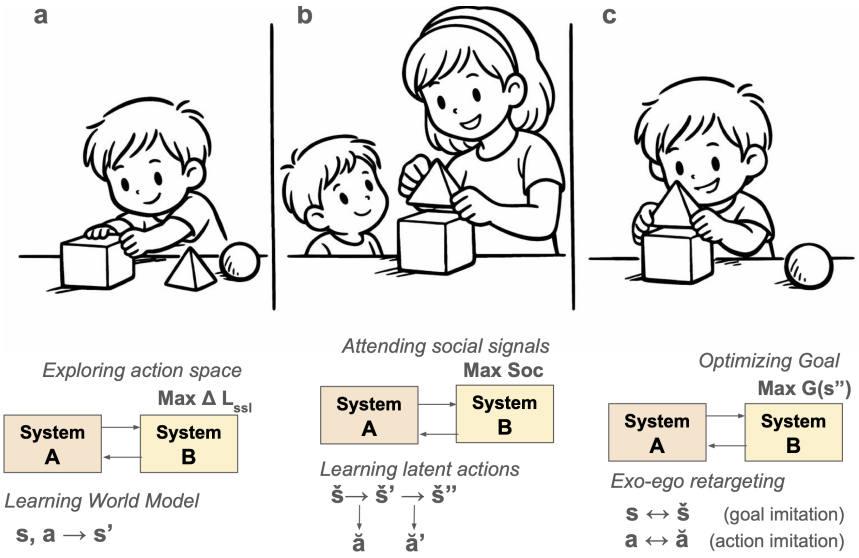

Figure 3 Interactions between learning modes for imitation learning. (a) **Self Play.** System B provides action, state trajectories to System A that learns a World Model, and provides a prediction-based intrinsic reward signal to system B. (b) **Social Observation**. System B directs the attention to peers that provide System A with complex trajectories from which it infers latent actions. (c) **Retargeted imitation**. System A learns to map exocentric actions and states to egocentric ones, helping system B to achieve goal-directed behavior. *(image from ChatGPT)*

offers rich and grounded input that supports the development of task-relevant representations (Pathak et al., 2017; Pong et al., 2020).

In both of these cases, System B may enrich the data available to System A by generating parallel action/perception datasets that support cross-modal learning or even a form of supervised learning, to the extent that actions are lower-variance and more reliable than noisy, ambiguous sensory inputs.

2.5 Towards deeper integration of learning modes

Figure 2 summarizes the interactions between the two systems discussed so far. In AI, deep integration between System A and System B has already been succesful in constrained domains. In games, agents like MuZero (Schrittwieser et al., 2020) and Dreamer (Hafner et al., 2023) couple learned latent dynamics with action planning to achieve superhuman performance. Such integration is gaining traction in robotics where vision-language-action models leverage massive, passively trained representations to directly guide motor execution (Driess et al., 2023; Zitkovich et al., 2023). However, in current systems, the learning recipe and runtime execution remains rigidly fixed by human engineers, while in living organisms, System A and B interact far more autonomously and fluidly. Classical AI has proposed *cognitive architectures* (ACT, SOAR; Anderson et al. 2004; Laird 2012) to formalize such flexibility, which have recently been adapted to deep learning. LeCun (2022) proposes an architecture integrating self supervived learning of world models and planning, within an energy based approach, enabling flexible specification of tasks through a *configurator*. Flexbility is also the target of *Global Workspace Theory* architectures (Goyal and Bengio, 2022), which are still in early stage of developmennt (Goyal et al., 2022).

In this section, we exemplify the flexibility problem with *imitation learning* (also called *social learning*)—a complex behavior requiring continuous toggling between passive observation and active motor correction—before turning to our proposed architecture, System M.

Imitation learning consists of reproducing an action performed by a conspecifics. It has been documented in several species (Whiten and Ham, 1992) and in human children (Meltzoff, 2007; Rizzolatti and Craighero, 2004). Unpacking this capability reveals that it relies on tightly integrated Systems A and B learning modes (see Figure 3). The organism needs to learn a sequence of observed action in conspecifics (System A world modeling), then map this sequence to a corresponding sequence of its own actions. This is not obvious, because the space of obervations (images) is not the same as the space of actions (motor commands). As the

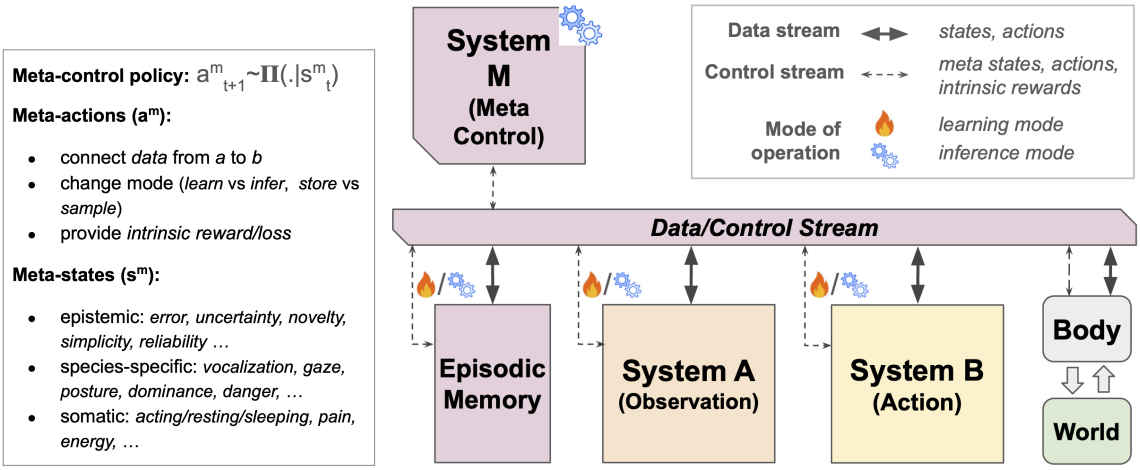

Figure 4 Blueprint of a cognitive architecture featuring System M as an autonomous orchestrator. System M acts as a central control plane that automates data routing and training recipes. High-bandwidth *data streams* (e.g., plain arrows) carry raw sensory inputs, motor commands, and latent representations between System A (perception/world modeling), System B (action/policy), and an episodic memory buffer. Low-bandwidth *control streams* (e.g., thin dashed arrows) carry telemetry: System M monitors internal meta-states (such as prediction errors or uncertainty) and outputs routing commands (meta-actions) to dynamically open or close data pathways, effectively assembling and disassembling learning and inference pipelines on the fly.

bodies between learner and demonstrators are not the same, this creates a 'retargeting problem', which can be addressed by trying to reproduce the observed state changes (goal imitation) or by learning a correspondence between exocentric and egocentric actions (action imitation). Solving this problem involves learning a common world model based on self-play and social observation, and learning a policy to reproduce these actions. These phases of observation and action likely alternating in cases of complex skills requiring a hierarchy of subskills.

Current approaches in robotics also learn from human examples, but they sidestep the retargeting problem by using data from *teleoperation* (where humans directly provide motor actions for the agent Fu et al. 2024). This limits the scale at which new skills can be added to robots, as well as the speed and agility of the learned skills. Recent approaches pretrain models using video data of human actions and learn latent actions (Baker et al., 2022; Tharwat et al., 2025). This reduces the amount of teleoperation data without eliminating it. A more flexibly integrated System A and System B model would be able to learn from videos of humans without any teleoperation data. In the next section, we explore what such a flexible architecture could look like.

3 Meta-control for autonomous learning (System M)

In humans and animals, System A and System B are active from birth, exhibiting a continuous, fluid, and autonomous interaction across cognitive domains (e.g., visuo-motor skills, social skills, and language). In current AI systems, the switching between learning modes is managed offline through rigid training recipes designed and controlled by human experts in a pipeline generally referred to as *MLOps* (Zhengxin et al., 2023). A quintessential example is the training pipeline of Large Language Models (LLMs), which strictly forces a massive phase of unsupervised next-token prediction (System A) followed by a subsequent, disconnected phase of reinforcement learning from human feedback (System B; Ouyang et al. 2022).

A truly autonomous AI would integrate components that automate the traditional MLOps human functions of *data sourcing and curation*, of building and adjusting *training recipes*, and of *benchmarking* performance and *monitoring* learning signals (Sculley et al., 2015; Paleyes et al., 2022). In turn, this would require integrating in the learning architecture, an *episodic memory* to store and replay raw or processed data, and a *central orchestrator* to dynamically apply training recipes and route data streams.

At this stage, we can only offer a blueprint, a conceptual sketch requiring substantial future work to fully instantiate. We call the orchestrator **System M** (for Meta-control) (see Figure 4). It would operate much like

the Control Plane in *Software-Defined Networking* (Kreutz et al., 2015). It does not process the high-bandwidth *data streams* of raw sensory inputs or motor commands directly. Instead, System M can be formalized as executing a (meta-)policy: it monitors low-dimensional telemetry or *meta-states* (e.g., prediction errors, uncertainty, or somatic signals) and outputs *meta-actions*. Paralleling the executive routing functions of the biological prefrontal cortex (Miller and Cohen, 2001), these meta-actions consist of dynamically interconnecting auxiliary processors (System A, System B) and episodic memory. By opening and closing data pathways, System M assembles and disassembles learning and inference pipelines on the fly.

Unlike the policies of System B, which are learned over the organism's lifetime via gradient descent or reinforcement, we propose that System M's core routing policy is hardwired—an evolutionarily fixed transition table that dictates when to explore, when to plan, and when to act[2]. In the following sections, we first review how such meta-control functions in biological agents before mapping these principles to autonomous AI.

3.1 Inspiration for Meta-Control in Humans and Animals

Biological agents exhibit meta-control through multiple mechanisms that regulate learning and behavior in context-sensitive ways. These mechanisms can be grouped into three broad categories that parallel the engineering components of artificial systems: *Input Selection*, *Loss/Reward Modulation*, and *Mode of Operation Control*.

Input Selection. The bandwidth of sensory space is overwhelmingly large, requiring organisms to typically attend to only a subset of their data stream. For instance, infants preferentially attend to faces or vocalizations, a form of hardwired data curation that boosts System A's social and language learning (Morton and Johnson, 1991; Vouloumanos and Werker, 2007). Furthermore, they allocate attention to visual sequences of intermediate complexity, effectively generating a developmental curriculum that optimizes learning gains (Kidd et al., 2012). At a lower level, sensory streams are accompanied by internal uncertainty estimates, allowing the organism to selectively discount noisy modalities during multimodal integration (Ernst and Banks, 2002). More generally, both children and adults actively explore and select their inputs during learning tasks, acting as intuitive scientists who sample data to resolve uncertainty (Gureckis and Markant, 2012; Gopnik, 2012; Gopnik et al., 1999).

Loss/Reward Modulation. The objective functions that organisms attempt to optimize are not fixed. At the developmental scale, critical periods illustrate that specific learning components are highly plastic only at certain developmental stages, effectively implementing a biological learning rate schedule (Smith and Gasser, 2005; Hensch, 2005). Special modes of learning and memory consolidation are triggered during sleep or rest states (Diekelmann and Born, 2010). During activity, agents transition between exploratory behavior (sampling new options to reduce uncertainty) and exploitative behavior (maximizing known rewards) based on the volatility and predictability of their environment (Daw et al., 2006; Wilson et al., 2014). Social signals are also powerful modulators of learning strategies. Animals will learn preferentially from dominant conspecifics (Kendal et al., 2015). Children dynamically modulate their learning updates by prioritizing pedagogically cued demonstrations (Csibra and Gergely, 2009), but also learn according to estimated reliability and trustworthiness (Harris and Corriveau, 2011).

Mode of Operation Control. System A and System B can be run independently in inference or learning modes, with their inputs and outputs dynamically routed to each other or to episodic memory. Most spectacularly, during sleep, motor outputs and primary sensory inputs are actively gated off by the orchestrator, while episodic memory, System A, and System B remain highly active for both inference and offline learning, as evidenced by coordinated neural replay during REM and slow-wave sleep (Rasch and Born, 2013). During wake states, organisms flexibly arbitrate between *habitual* control (System B reactive policies, or model-free behavior) and *goal-directed* planning (System A world-model simulation, or model-based behavior) depending on reward stability, task volatility, and the degree of task mastery (Daw et al., 2005; Dolan and Dayan, 2013). Other examples include flexibly toggling between observational social learning and direct trial-and-error problem solving depending on task novelty and success rates, as observed in corvids and macaques (Subiaul

[2]This may seem a bit extreme considering that humans display the ability to acquire specific learning strategies through culture (eg, going to school). This is not incompatible with a fixed system M, however, as illustrated in Appendix B where we show that advanced learning modes like learning by communication and by imagination can be implemented through a fixed system M.

et al., 2004; Taylor et al., 2012). Similarly, human children leverage internal uncertainty estimates to switch into specialized learning modes, initiating exploratory play or engaging in metacognitive help-seeking when they recognize their own models are insufficient (Lyons and Ghetti, 2010).

To summarize, meta-control can be modeled as a policy, $\pi(a^m|s^m)$, which maps an internal *meta-state* to a corresponding *meta-action* (Figure 4). The input meta-states can be summarized into three distinct types: *epistemic signals*, which are derived by monitoring the internal operation of cognitive components (e.g., confidence, prediction error, learning gain, or novelty); *species-specific signals*, which are high-priority environmental configurations recognized by pre-wired evolutionary detectors (e.g., direct gaze, dominance displays, looming stimuli, or heights); and finally, for embodied agents, *somatic signals* derived directly from the physical body (e.g., energy levels, pain detectors, or arousal states). Meta-actions consist of dynamically connecting or disconnecting the input and output data streams of the subcomponents, turning them on and off in various operating modes (e.g., learning, inference, or optimization), providing them targets or internal rewards, and accessing episodic memory for memory replay or randomized batch learning. Together, these meta-actions enable the system to autonomously assemble and disassemble entire training and inference pipelines on the fly.

3.2 Meta-Control in AI systems

While a full specification of the possible routing circuits and meta signals is outside the scope of this paper, we refer the reader to Appendix B for advanced operating modes, and Appendix C for further ethological examples of meta-signals. In addition, we provide below a non-exhaustive list of relevant work in the AI literature where isolated fragments of System M are actively being developed.

Input Selection. In machine learning, the autonomous curation of data streams is explored through the lens of *Active Learning*, where epistemic meta-signals (like model uncertainty or ensemble variance) help query the most informative data points, drastically reducing sample complexity (Ren et al., 2021). In Reinforcement Learning, *Prioritized Experience Replay* (PER), past experiences are sampled with a probability proportional to their temporal difference error (a direct equivalent of biological prediction error) (Schaul et al., 2016). *Mixtures of Experts* were an early implementation of dynamic routing of subsystems (Jacobs et al., 1991), and have been applied to deal with noisy or incomplete modalities (e.g., Han et al. 2024; Mai et al. 2026).

Loss/Reward Modulation. The autonomous, dynamic adjustment of objective functions is actively researched in the fields of *Intrinsic Motivation* and *Unsupervised Environment Design* (UED), where meta-states like novelty or prediction error, enable curiosity-driven exploration that mirrors biological exploratory play (Oudeyer and Kaplan, 2007; Bellemare et al., 2016; Pathak et al., 2017). *Auto-Curriculum* algorithms dynamically modulate the difficulty of the training environment generating tasks at the frontier of the agent's capabilities (Portelas et al., 2020), while *Continual Learning* methods like Elastic Weigh Consolidation mimick the biological critical periods discussed earlier Kirkpatrick et al. (2017).

Mode of Operation Control. The dynamic arbitration between fast inference (System B) and deliberate planning (System A) has initially been explored in RL through a meta-controller optimizing an 'imagination budget' by switching from world-model simulation (Monte Carlo Tree Search), to a reactive model-free policy based on task difficulty (Hamrick et al., 2017). Similarly, in Hierarchical Reinforcement Learning (HRL), a high-level manager policy does not output motor actions directly, but rather acts as a router that toggles lower-level, specialized sub-policies on and off depending on the task context (Bacon et al., 2017). These ideas are also explored through *inference-time compute scaling*. LLM-based reasoning and agentic models using *'Thinking' Tokens* to solve complex tasks (Yao et al., 2023; Snell et al., 2024) can also be taught to switch between reactive responses and more expensive search Lin et al. (2023); Zelikman et al. (2024).

Despite these advances, modern AI still lacks a unified Control Plane that integrates all these meta-functions—input selection, reward modulation, and operational routing—into a single, cohesive architecture. Next, we explore how such a system could be built.

4 Bootstrapping Autonomous Learning: An Evolutionary-Developmental Framework

Our proposed architecture with Systems A, B, and M is an overall blueprint for autonomous learning. Building a functional model may prove challenging due to the interdependencies between the three components: if System A relies on action-generated data to acquire grounded representations, and System B in turn depends on perceptual structure to guide efficient action, how can either system be initialized so that learning can begin? Likewise, if System M is crucial for orchestrating learning in the other systems, but itself depends on well-calibrated uncertainty or error signals produced by them, how can it be set up to ensure robust learning across diverse environments? In the following sections, we draw from biology the distinction between adaptations at the evolutionary versus developmental scale (Evo/Devo), to outline a strategy for resolving this chicken-and-egg (and rooster) problem.

4.1 Evo/Devo Scales for Organisms

Even a cursory examination of natural organisms reveals that none start from randomly initialized neural networks (Zador, 2019). Animals inherit a highly specified species-typical nervous system that unfolds over developmental time. This inherited structure constrains and guides learning by providing *inductive biases* that shape what can be learned, how rapidly, and through which modalities (Karmiloff-Smith, 1992; Johnson, 2001; Gallistel, 1990, 2013). In computational terms, such biases can be understood as an initial *architectural* and *parametric configuration*, coupled with a developmental program that provides an *internal curriculum* (Bengio et al., 2009) that gradually increases the complexity of perception, action, and learning (Thelen and Smith, 1994). Empirically documented mechanisms include synaptic growth and pruning, temporally regulated plasticity, critical periods, spontaneous neural activity (Huttenlocher, 1979; Hensch, 2005; Hadders-Algra, 2018; Molnár et al., 2020), and progressive increase in visual acuity and motor degrees of freedom (Turkewitz and Kenny, 1982; Turvey, 1990). Through these processes, organisms acquire increasingly sophisticated representations and action policies starting from a deliberately simplified regime (Elman, 1993), but by no means a tabula rasa state.

One could argue, however, that this merely shifts the bootstrapping problem from ontogeny to phylogeny, without explaining how inductive biases and developmental curricula arise in the first place. Here we can offer only speculative remarks. Large-brained, highly behaviorally flexible organisms are evolutionarily recent, and early life forms likely operated with comparatively simple sensorimotor loops in environments where modest behavioral plasticity was sufficient (Striedter, 2005; Paulin and Cahill-Lane, 2021; Keijzer, 2015). Evidently, large brains are not a systematic outcome of evolution, as brainless organisms still occupy the majority of the biomass (Bar-On et al., 2018). Nevertheless, evidence suggests that during the Cambrian, a fraction of this biomass transitioned into more cognitively sophisticated organisms with improved sensory and effector organs (see a review in Hsieh et al. 2022) correlatively to increasingly complex ecosystems including predation and competition (Plotnick et al., 2010). The putative relationship between niche complexity, brain complexity and learning abilities remains a fascinating and controversial topic (Krause et al., 2022; Dukas and Ratcliffe, 2019; Arendt et al., 2016), but suggests a pathway from simple to increasingly sophisticated autonomous learning capabilities.

4.2 Evo/Devo for autonomous AI

In machine learning, the evolutionary–developmental distinction is also present, but instantiated in strikingly different terms. The developmental scale is the realm of current machine learning algorithms, as discussed above as System A and B. The evolutionary scale is, for the most part, manifested through the standard practices of scientific research (distributed research teams, publications, open sourcing, etc.). In this sense, System M is implemented through humans—research scientists, engineers, and students.

Here we propose to formalize the problem in a unified fashion (Figure 5): each agent is defined by a set of parameters ϕ that correspond to the information in its genetic code. At "birth", ϕ is used to specify an architecture comprising Systems A, B, and M with their initial parameters (A_0,B_0,M_0). During the developmental scale (inner loop), the agent learns in interaction with its environment by updating Systems A

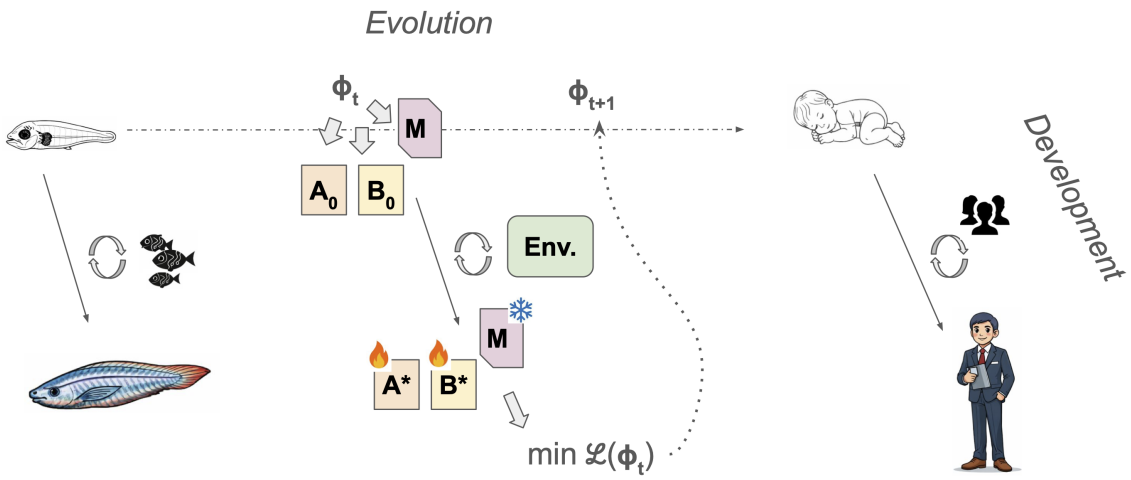

Figure 5 Evo/Devo framework for building autonomous learning agents. Learning takes place at two scales. In the developmental scale, the learner's architecture (A, B and M) is initialized from meta parameter ϕ. A and B update their parameters through interaction with the environment controlled by a fixed controler M. In the the evolutionary scale, *phi* is updated to optimize a fitness function $\mathcal{L}$ measured over the life cycle of the system. *(images from ChatGPT).*

and B as controlled by a fixed System M. On the evolutionary timescale (outer loop), ϕ is optimized through an externally handcrafted fitness function $\mathcal{L}$ computed at the end of the life cycle of each agent. The fitness function (and the environments) are the only handcrafted parts of the system. The update rules, initialization, data filtering, curriculum, etc., are all provided through System M.

$$\phi_{t+1} = \arg\min_{\phi_t} \mathcal{L}(A_0 : A_K, B_0 : B_K)$$

subject to

$$A_0, B_0, M = \text{Init}(\phi_t)$$
$$A_{i+1}, B_{i+1} = \text{Update}(M, A_i, B_i, Env)$$

where

Init is the initialization procedure
Update is the inner loop update rule
Env is the interactive environment

This problem belongs to the class of *bilevel optimization* problems. Methods to solve it include research on meta-learning (Schmidhuber, 1987; Thrun and Pratt, 1998; Andrychowicz et al., 2016; Finn et al., 2017), neural architecture search (Stanley and Miikkulainen, 2002; Zoph and Le, 2017), curriculum learning (Bengio et al., 2009; Graves et al., 2017), intrinsic motivation (Oudeyer and Kaplan, 2007), exploration strategies (Bellemare et al., 2016; Pathak et al., 2017), and continual learning (Ring, 1997; Thrun, 1998; De Lange et al., 2021), although these elements have not yet been integrated to deliver an autonomous learning system.

Note that bilevel optimization problems are challenging and have typically been studied in relatively simple situations. One challenge is that at the outer level, a whole life cycle is just one data point. In order to optimize ϕ, one needs to run millions of simulated life cycles which themselves imply learning over millions of datapoints. This requires considerable feats of improving memory and compute efficiency of the basic architecture. In addition, while bilevel optimization can be solved using a variety of techniques, both gradient-based and gradient-free methods (see Sinha et al. 2018 for a review), extending such techniques to much larger architectures yields severe scalability issues (see Lorraine et al. 2020; Real et al. 2019; Metz et al. 2021). A second challenge is that what has to be optimized is itself a dynamic system comprising a learner and an environment in interaction. Here, we would suggest that an *Evolutionary Curriculum*, gradually increasing the diversity and unpredictability of the environments, would help allowing the three components to co-evolve and solve the chicken–egg–rooster problem (Oudeyer and Kaplan, 2007; Leike et al., 2017).

Before closing, let us note that we are not proposing that the full complexity of living organisms need to be reproduced to construct autonomous systems. These principles can be applied to systems behaving adaptively

in simpler environments as has been done with adaptive control, for instance. One can view our proposal as an extension of adaptive control to more complex problems with the tools of modern AI.

5 Conclusions

Contrary to animals, current AI systems don't learn autonomously. Their learning is restricted to an off-line MLOps pipeline involving a large team of experts who prepare the data, build the training recipes and adjust them according to performance metrics. Once deployed, the models do not learn anymore, and adaptation to specific use cases is on the users through prompting or fine tuning.

Our analysis shows that current AI systems are lacking are three key abilities that are found across the animal kingdom: the ability to select their own training data (*active learning*), the ability to flexibly switch between learning modes (*meta control*), the ability sense their own performance (*meta-cognition*). In this paper, we identified three roadblocks that stand in the way to unlocking these missing abilities: the necessity to merge AI techniques from well siloed paradigms (self-supervised learning, reinforcement learning), the necessity to build an integrated cognitive architecture that automates the MLops pipeline through additional data routing, orchestration, and internal sensing components (the A-B-M architecture), the necessity to build these components jointly through an evolutionary/development scheme using simulated environments.

5.1 Why is Autonomous Learning Useful?

Systems that learn autonomously and adaptively like animals or humans could offer a powerful leap toward more general, robust, and flexible intelligence. Such systems would learn from raw experience, side stepping the entire MLOps pipeline, enabling them to operate in complex, changing, or poorly understood environments. Like children or animals, they could generalize from limited examples, explore new tasks through curiosity, and develop an embodied, grounded understanding of the physical and social world (Lake et al., 2017). These features would make autonomous AI well-suited for real-world deployment in uncertain or dynamic settings—ranging from home robots to scientific discovery.

Beyond practical utility, they also offer a unique opportunity to reverse-engineer natural intelligence, accelerating the scientific understanding of the neural mechanisms and developmental trajectories of cognition (Turing, 1950; Hassabis et al., 2017) across different environments and cultures.

5.2 Why is it Difficult?

Building a machine that learns like children do has been envisioned since the inception of AI (Turing, 1950), but several technical and ethical challenges remain.

Simulators and Environments. Training tightly coupled Systems A, B, and M requires environments that should both be realistic and optimized for speed. On the realism side, it should provide diverse sensory modalities to enable observational learning (System A), rich embodied affordances and goal-conditioned tasks to support interaction and control (System B), and sufficiently diverse and non-stationary dynamics to train a robust meta-controller (System M). On the speed side, faster than real time is necessary for the evolutionary scale to be applicable. Procedural generation can help scale diversity, but requires careful design to avoid trivial patterns or degenerate exploration (Cobbe et al., 2020). Incorporating social agents or enabling teacher–learner interactions is especially challenging at scale (Crosby et al., 2019; Savva et al., 2019).

Evaluation and Benchmarking. As agents become more general, task-specific benchmarks lose their diagnostic value, creating the need for new evaluation paradigms. We propose distinguishing between *unit tests*, which assess individual learning components in isolation, and *integration tests*, which evaluate their combined behavior.

Unit tests should target the core competencies of each system: sample-efficient perceptual generalization for System A; few-shot task adaptation under sparse rewards for System B; and, for System M, efficient switching between learning modes in non-stationary environments, as well as the emergence of more advanced capabilities such as learning through communication or imagination.

Integration tests should assess end-to-end learning performance in realistic settings. One promising approach is to compare humans and AI agents in terms of learning speed on novel tasks—for example, the number of trials required to learn a new videogame (Lake et al., 2017), or the number of hours of exposure needed to acquire a language at a level comparable to that of human children (Dupoux, 2018). Such benchmarks should emphasize few-shot, data-efficient generalization and closely mirror developmental learning processes (Hill et al., 2020; Chevalier-Boisvert et al., 2019).

Scaling Bi-Level Optimization. Optimizing lifelong learning processes in complex environments is both computationally demanding and highly sensitive to curriculum design. Addressing these challenges requires progress along several complementary directions. First, we need more efficient inner-loop learners with low data and computational requirements. Second, meta-objectives must be designed to effectively shape priors, intrinsic reward structures, and communicative tendencies. Third, appropriate strategies for curriculum scheduling and environment sampling are necessary to accelerate the emergence of robust, autonomous learning architectures.

Ethical Issues. Developing AI systems that learn in ways analogous to humans or animals raises novel ethical concerns that go beyond those associated with current AI technologies. In particular, autonomous learning introduces new trade-offs between flexibility, safety, and societal oversight.

A first challenge concerns the tension between *adaptability* and *controllability*. As systems are granted greater autonomy in exploratory learning modes, it becomes harder to guarantee that they remain aligned with intended objectives. Mitigating this risk may require explicit auditing mechanisms and the ability to intervene in or constrain the meta-control system (System M).

A second risk is *alignment hacking*. Although animals evolved to optimize reproductive fitness, their everyday behavior is often driven by proxy objectives such as exploration or play, and can occasionally give rise to maladaptive outcomes, including addiction or self-harm. These behaviors arise because biological agents optimize internally generated signals that may become mismatched to their environment (Tooby and Cosmides, 1992; Buss, 2019). Autonomous artificial agents that rely on similar proxy signals may face analogous vulnerabilities.

A third concern relates to *over-trusting*. As artificial agents become more human-like in their behavior and learning trajectories, users may increasingly anthropomorphize them, leading to emotional attachment, misplaced trust, or opportunities for manipulation (Turkle, 2011). Addressing this risk requires transparency about system capabilities and limitations, as well as mechanisms that ensure meaningful societal and user control.

Finally, autonomous learning systems often depend on bodily or somatic signals to guide adaptation. To the extent that these signals are processed in ways functionally analogous to pain or fear in biological organisms, this raises unresolved questions about the *moral status* of such agents (Gunkel, 2012; Birhane et al., 2021).

6 The Path Forward

To build agents capable of autonomous, open-ended learning, we must move beyond disjointed, hand-designed training paradigms and rigid execution. Already, the AI field is moving beyond fixed systems with frontier topics such as *runtime adaptation* and *inference-time compute*, including Large Language Models (e.g., test-time training with verifier-driven selection (Moradi et al., 2025), synergistic adaptation (Xu et al., 2025), and adaptive retrieval (Sun et al., 2026)), Vision-Language Models (Lei et al., 2025; Kojima et al., 2025), speech recognition (Fang et al., 2026), and core Reinforcement Learning (Chehade et al., 2025; Bagatella et al., 2025). Closely related ideas are also gaining traction in robotics under adjacent terminology such as online adaptive learning (Yuan et al., 2026), deployment-time correction (Welte et al., 2026), test-time reinforcement learning (Liu et al., 2026), and test-time mixture of world models (Jang et al., 2026). These adaptations, however exciting, are still minor variations over an overall rigid system, compared to children who learn a whole language and new skills at 'test time'.

The challenges are considerable and we are probably decades away from fully autonomous, broad scope learning systems. Our proposed architecture (System A-B-M) is a tentative blueprint that we hope can inspire research by providing a unified conceptualization of this problem space, and a path towards building actual

autonomous learning systems—inspired by natural intelligence, but not strictly bound to replicate it. The alignment of such adaptive systems with humans goals, and the autonomy–controllability trade-offs, are of paramount importance, and should be considered within an evolutionary–developmental framework through the careful design of fitness rewards and interactive environments. But even before fully autonomous learning systems are achievable, the successes and failures in building such systems will be scientifically invaluable, providing quantitative models of how biological organisms successfully learn and adapt in the wild, and offering insights on the very nature of learning and intelligence.

Appendix

A What's the link with System 1 and 2?

Kahneman's (Kahneman, 2011) distinction is about modes of inference (fast parallel inference, as in a forward pass in a DNN versus slow multistep inference, as in tree search or chain of thought). Our distinction is about modes of learning (from static data versus from interaction). Even though these distinctions seem similar, one could argue that they are actually orthogonal: one could learn from static data through simple backpropagation, or after having run simulation or imagination steps; conversely, even though learning from interaction does require at least one step of action, this action step could be generated reflexively through a learned policy rather than search.

		Modes of learning	
		System A	System B
Modes of inference	System 1	predictive coding; statistical learning	policy learning
	System 2	counterfactual reasoning; causal learning	learning through planning

Table A.1 Link between System 1/2 and System A/B

B Two Advanced Meta-Controlled Learning Modes: Communication and Imagination

In species living in complex and unpredictable environments, a capable System M enables new functional configurations, which we illustrate briefly here.

B.1 Learning from Communication

One of the functions of System M is to select "important" inputs to learn from, and across many species, social context is highly salient. We call this learning through communication, and it serves as a mechanism for accelerated, distributed learning (Tomasello, 1999; Heyes, 2018).

This ranges from basic attention-based mechanisms to complex cultural transmission. Below is a structured summary of key forms of social learning, grouped by increasing complexity.

Basic Observational and Associative Learning. Individuals adjust their behavior by noticing where others direct their attention or how they react to stimuli. Example: A child becomes interested in a toy only after seeing another child play with it (Bandura, 1977; Moore, 2013).

Behavioral Copying. Learners replicate others' actions or outcomes, either by imitating specific movements or by finding new ways to reach the same goal. Example: A chimpanzee uses a different method to retrieve food after seeing another (typically dominant) chimpanzee succeed (Whiten et al., 2005; Subiaul et al., 2004).

Guided Learning. Learning is supported by intentional or structured social interaction with a teacher, using methods such as shared attention, social referencing, behavioral scaffolding, or demonstrations. Example: A parent shows how to use an object while making eye contact with a child (Bruner, 1983; Wood et al., 1976). Such a mode of learning can extend to mimicking gestures that have no obvious goals (like learning rituals or complex recipes).

Higher-Level Social Learning. Learners internalize norms, generalize across contexts, and pass on knowledge culturally through abstract symbolic formats like language (Csibra and Gergely, 2009; Henrich, 2016). Example: Children are explicitly taught about social rules, follow verbal instructions.

While the first two forms of social learning have been documented across many species (mammals and birds), guided learning is more rarely observed, and only humans exhibit the higher-level form of social learning (Tomasello, 2009).

System M supports learning through communication by attending to communicative triggers (e.g., pointing, direct gaze, imperative intonation), and routing the highlighted inputs for System A or B learning. The strength of this learning episode can be one-shot and modulated by System M based on perceived social importance or trust in the teacher (epistemic vigilance; Sperber et al., 2010).

In standard AI systems, learning through communication is done externally, through a team of data scientists curating the data from reliable sources, or through post-training methods. But because this process is entirely externalized, current systems are unable to learn socially or exert epistemic vigilance regarding the source of their data.

B.2 Learning from Imagination

Another mode of operation enabled by System M is the ability to learn from internally generated inputs (learning from imagination). This mode of operation has been documented in varying degrees of complexity across species.

Memory replay at rest. During pauses in activity (e.g., at decision points or after receiving a reward), rodents experience a reactivation of the place cells they have recently visited, in the same order or reverse order, typically at a much faster rate. This has been linked to decision making, near-term planning, and updating value functions in reinforcement learning (Foster and Wilson, 2006; Diba and Buzsáki, 2007; Mattar and Daw, 2018).

Memory replay during sleep. During non-REM sleep, recent experiences are replayed at a compressed time scale, often in longer episodes including novel combinations. This has been linked to memory consolidation and the formation of long-term schemas or generalization (Wilson and McNaughton, 1994; Diekelmann and Born, 2010; Kumaran et al., 2016).

Long-horizon planning. Humans and some animals show evidence of problem-solving without prior trial and error in tasks involving multiple steps of tool use, delayed gratification, or counterfactual reasoning (Suddendorf and Corballis, 2007; Redshaw and Suddendorf, 2016; Tulving, 2005). Such phenomena highlight imagination-like simulation as a substrate for flexible foresight

System M can support these modes of operation by switching Systems A and B into inference mode, routing input information from memory (instead of the sensors), and routing output information (e.g., actions) to internal simulation. It can then trigger learning on the successful imagined trajectories. This highlights the flexibility of System M—not just as a router, but as an enabler of qualitatively new learning regimes (Ha and Schmidhuber, 2018; Hafner et al., 2020).

C Examples of Meta-Control signals for autonomous learning in Humans and Animals

In Table C.1, we list examples of meta-states that have been found in humans and animals to have direct effect on input selection (attention, gaze), on learning efficiency (loss/reward modulation) or on mode of Control. We sort them according to our three way category of species-specific, epistemic and somatic signals. Interestingly, most of these signals are relatively simple to compute or rest on rudimentary computations (for instance, the preference for faces boils down to a preference to for a dark T pattern on a white background), making it relatively simple to emerge in a meta learning setting. They could inspire the design of meta-controlers in artificial agents depending on the application.

Acknowledgements

This paper is the result of several years of discussion between the authors and an interdisciplinary workshop on autonomous learning at META in July 2, 2025 in New York. We thank the participants of this workshop, in particular Karen Adolf, Catherine Tamis Lemonda, Pulkit Agrawal, Carl Vondrick, Linda Smith, Allison Gopnick, Brenden Lake, Mido Asran, Michael Henaf, Alessandro Lazaric, Mahi Luthra, Juan Pino, Jiayi Shen and Pascale Fung who shared useful insights. We are especially grateful to Shiry Ginosar for her detailed, incisive, and highly useful comments on an earlier version of this paper. A first version of Section 2 has been published as a section in Fung et al. (2025). The work was conducted while JM and YLC were at META. ED in his EHESS role was supported by the Agence Nationale pour la Recherche (ANR-17-EURE0017 Frontcog, ANR10-IDEX-0001-02 PSL*) and an ERC grant (InfantSimulator), these granting agencies declining responsibilities for the views and opinions expressed. AI tools were used in the conception stage, to discuss the logic and structure of the paper, to produce the graphical elements of figures 1 and 3, and in the final stage for a stylistic / grammatical check and rechecking of the bibliography.

Meta-states	Species	Effects (sample references)
Species-Specific signals		
Faces and vocalizations	Human infants	IS (Johnson et al., 1991; Vouloumanos and Werker, 2007); LE (Ferry et al., 2010)
Gaze direction	Human infants, dogs, corvids	IS (gaze following: Farroni et al. 2002; Tomasello et al. 1998)
Pedagogical signals *(high pitch, direct gaze; pointing)*	Human infants	IS (Fernald, 1985), LE (Thiessen et al., 2005), MC (shifts from memorization to generalization; Csibra and Gergely 2009; Gergely et al. 2002)
Self-propelled, biomechanical motion	Human infants, chicks	IS (Simion et al., 2008; Scholl and Tremoulet, 2000), LE (Nairne et al., 2013); MC (switches to teleological reasoning; Csibra et al. 1999)
Potential threats *(looming, snake or spider-like objects)*	Human infants, mammals	IS (Ball and Tronick, 1971; Yilmaz and Meister, 2013; DeLoache and LoBue, 2009); LE (Õhman and Mineka, 2001; Cook and Mineka, 1989); MC (freezing/high vigilance; Fanselow 1994)
Dominant, prestigious or in-group conspecifics	Human infants, primates	IS (Shepherd et al., 2006; Chudek et al., 2012; Kinzler et al., 2007); LE (Buttelmann et al., 2013); MC (shifts from high to low level imitation; Haun and Tomasello 2011)
Epistemic signals		
Reliable conspecifics / Selective trust	Human infants, primates, dogs	IS (Poulin-Dubois et al., 2011; Schmid et al., 2017); LE (Koenig et al., 2004; Zmyj et al., 2010); MC (shifts to exploration when unreliable; Gweon et al. 2014; Takaoka et al. 2015)
Logical/semantic contradictions or conflicts	Human adults, infants, primates	IS (Baillargeon et al., 1985); LE?? (Stahl and Feigenson, 2015); MC (switches to world-model simulation; Botvinick et al. 2001)
Unexpected outcomes	Mammals, birds, insects	IS (Sokolov, 1963); LE (Pearce and Hall, 1980; Schultz et al., 1997); MC (shifts to exploration: Dayan and Balleine 2002)
Uncertainty	Human infants, primates	LE (bayesian multimodal fusion; Ernst and Banks 2002); MC (halts exploitation (opt-out); Goupil et al. 2016; Hampton 2001)
Stimuli of intermediate complexity	Human infants, primates	IS (Kidd et al. 2012), LE (Kang et al. 2009)
Somatic signals		
Sleep	All animals	MC, LE, IS (disconnection of sensory inputs, boost in learning, memory replay) (Buzsáki, 2015; Tononi and Cirelli, 2014)
Rest	All animals	MC, LE, IS (sensory input attenuation, learning by imagination, memory replay) (Kam et al., 2011; Buckner et al., 2008; Raichle, 2015)
Pain	All animals	MC (interrupts all goal-directed plans; switch to reactive actions, followed by replay and world modeling), LE (single shot learning) (Eccleston and Crombez, 1999)
Hunger	All animals	MC (high exploration, boost in goal directed), IS(enhanced detection of food) (Balleine and Dickinson, 1998; Kolling et al., 2012).
Stress	All animals	MC (low stress: world modeling simulation, exploration; high stress: reactive policies, exploitation) (Schwabe and Wolf, 2009)

Table C.1 Non exhaustive list of meta-states in humans and animals, and their effects interpreted in terms of Input Selection (IS: attracting gaze or attention), Learning Efficacy (LE: boosting learning rate) and Mode Control (MC) in humans and animals.

References

Karen E. Adolph, Whitney G. Cole, Meghana Komati, Jessie S. Garciaguirre, Daryaneh Badaly, Jesse M. Kurz, Gina L. Korleski, and Rebecca L. Freedland. How do you learn to walk? thousands of steps and dozens of falls per day. *Psychological Science*, 23(11):1387–1394, 2012. doi: 10.1177/0956797612446346.

Pulkit Agrawal, Ashvin Nair, Pieter Abbeel, Jitendra Malik, and Sergey Levine. Learning to poke by poking: Experiential learning of intuitive physics. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 29, pages 5074–5082, 2016.

Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob L Menick, Sebastian Borgeaud, Andy Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikoł aj Bińkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karén Simonyan. Flamingo: a visual language model for few-shot learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 23716–23736. Curran Associates, Inc., 2022.

Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*, 2016. URL https://arxiv.org/abs/1606.06565.

John R. Anderson, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An integrated theory of the mind. *Psychological Review*, 111(4):1036–1060, 2004. doi: 10.1037/0033-295X.111.4.1036.

Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando de Freitas. Learning to learn by gradient descent by gradient descent. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 3981–3989, 2016.

Detlev Arendt, Maria Antonietta Tosches, and Heather Marlow. From nerve net to nerve ring, nerve cord and brain—evolution of the nervous system. *Nature Reviews Neuroscience*, 17(1):61–72, 2016.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15619–15629, 2023. doi: 10.1109/CVPR52729.2023.01500.

Mido Assran, Adrien Bardes, David Fan, Quentin Garrido, Russell Howes, Mojtaba, Komeili, Matthew Muckley, Ammar Rizvi, Claire Roberts, Koustuv Sinha, Artem Zholus, Sergio Arnaud, Abha Gejji, Ada Martin, Francois Robert Hogan, Daniel Dugas, Piotr Bojanowski, Vasil Khalidov, Patrick Labatut, Francisco Massa, Marc Szafraniec, Kapil Krishnakumar, Yong Li, Xiaodong Ma, Sarath Chandar, Franziska Meier, Yann LeCun, Michael Rabbat, and Nicolas Ballas. V-jepa 2: Self-supervised video models enable understanding, prediction and planning, 2025. URL https://arxiv.org/abs/2506.09985.

Arthur Aubret, Laetitia Matignon, and Salima Hassas. An information-theoretic perspective on intrinsic motivation in reinforcement learning: A survey. *Entropy*, 25(2):327, 2023. doi: 10.3390/e25020327.

Pierre-Luc Bacon, Jean Harb, and Doina Precup. The option-critic architecture. In *AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Adrià Puigdomènech Badia, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, Bilal Piot, Steven Kapturowski, Olivier Tieleman, Martín Arjovsky, Alexander Pritzel, Andrew Bolt, and Charles Blundell. Never give up: Learning directed exploration strategies. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.

Marco Bagatella, Mert Albaba, Jonas Hübotter, Georg Martius, and Andreas Krause. Test-time offline reinforcement learning on goal-related experience. In *Proceedings of the Pre-Training in the Wild Workshop at ICML 2025*, 2025. OpenReview.

Renée Baillargeon. Infants' physical world. *Current directions in psychological science*, 13(3):89–94, 2004. doi: 10.1111/j.0963-7214.2004.00281.x.

Renée Baillargeon, Elizabeth S. Spelke, and Stanley Wasserman. Object permanence in five-month-old infants. *Cognition*, 20(3):191–208, 1985. doi: 10.1016/0010-0277(85)90008-3.

Bowen Baker, Ilge Akkaya, Peter Zhokhov, Joost Huizinga, Jie Tang, Adrien Ecoffet, Bradley Weiner, Joel Lehman, and Jeff Clune. Video pretraining (vpt): Learning to act by watching unlabeled online videos. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 24639–24654, 2022.

William Ball and Edward Tronick. Infant responses to impending collision: Optical and real. *Science*, 171(3973): 818–820, 1971.

Bernard W. Balleine and Anthony Dickinson. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*, 37(4-5):407–419, 1998. doi: 10.1016/s0028-3908(98)00033-1.

Albert Bandura. *Social Learning Theory*. Prentice Hall, Englewood Cliffs, NJ, 1977.

Yinon M Bar-On, Rob Phillips, and Ron Milo. The biomass distribution on earth. *Proceedings of the National Academy of Sciences*, 115(25):6506–6511, 2018.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mahmoud Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video. *arXiv preprint arXiv:2404.08471*, 2024. doi: 10.48550/arXiv.2404.08471.

Marc G. Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Rémi Munos. Unifying count-based exploration and intrinsic motivation. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1471–1479, 2016.

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*, pages 41–48, 2009. doi: 10.1145/1553374.1553380.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. doi: 10.1109/TPAMI.2013.50.

Nikolai A. Bernstein. *The Co-ordination and Regulation of Movements*. Pergamon Press, Oxford, UK, 1967.

Dimitri P. Bertsekas. *Reinforcement Learning and Optimal Control*. Athena Scientific, Belmont, MA, 2019. ISBN 978-1886529397.

Abeba Birhane, Jelle van Dijk, and Rens Huygen. The ethics of embodied ai. *AI & Society*, 36(3):705–715, 2021.

Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, Nicolas Pinto, and Joseph Turian. Experience grounds language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8718–8735, 2020. doi: 10.18653/v1/2020.emnlp-main.703.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. On the opportunities and risks of foundation models. 2022. URL https://arxiv.org/abs/2108.07258.

Matthew Botvinick, Jane X. Wang, William Dabney, Kevin J. Miller, and Zeb Kurth-Nelson. Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, 23(5):408–422, 2019. doi: 10.1016/j.tics.2019.02.006.

Matthew M. Botvinick, Todd S. Braver, Deanna M. Barch, Cameron S. Carter, and Jonathan D. Cohen. Conflict monitoring and cognitive control. *Psychological Review*, 108(3):624–652, 2001. doi: 10.1037/0033-295X.108.3.624.

Robert Boyd, Peter J. Richerson, and Joseph Henrich. The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, 108(Supplement 2):10918–10925, 2011. doi: 10.1073/pnas.1100290108.

Jake Bruce, Michael Dennis, Ashley Edwards, Jack Parker-Holder, Yuge (Jimmy) Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, Yusuf Aytar, Sarah Bechtle, Feryal Behbahani, Stephanie Chan, Nicolas Heess, Lucy Gonzalez, Simon Osindero, Sherjil Ozair, Scott Reed, Jingwei Zhang, Konrad Zolna, Jeff Clune, Nando De Freitas, Satinder Singh, and Tim Rocktäschel. Genie: generative interactive environments. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, ICML'24, 2024.

Jerome Bruner. *Child's Talk: Learning to Use Language*. W. W. Norton & Company, New York, NY, 1983.

Randy L. Buckner, Jessica R. Andrews-Hanna, and Daniel L. Schacter. The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, 1124(1):1–38, 2008. doi: 10.1196/annals.1440.011.

David M. Buss. *Evolutionary Psychology: The New Science of the Mind*. Routledge, 6th edition, 2019.

David Buttelmann, Norbert Zmyj, Moritz Daum, and Malinda Carpenter. Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, 84(2):422–428, 2013. doi: 10.1111/j.1467-8624.2012.01860.x.

György Buzsáki. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus*, 25(10):1073–1188, 2015. doi: 10.1002/hipo.22488.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9650–9660, 2021. doi: 10.1109/ICCV48922.2021.00951.

Mohamad Chehade, Amrit Singh Bedi, Souradip Chakraborty, Amy Zhang, and Hao Zhu. Tram: Test-time risk adaptation with mixture of agents. In *International Conference on Learning Representations (ICLR) 2026 Submission*, 2025. OpenReview.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119, pages 1597–1607, 2020.

Maxime Chevalier-Boisvert, Dzmitry Bahdanau, Salem Lahlou, Lucas Willems, Chitwan Saharia, Thien Huu Nguyen, and Yoshua Bengio. Babyai: A platform to study the sample efficiency of grounded language learning. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, 2019.

Maciej Chudek, Sarah Heller, Susan Birch, and Joseph Henrich. Prestige-biased cultural learning: bystander's differential attention to potential models influences children's learning. *Evolution and Human Behavior*, 33(1):46–56, 2012. doi: 10.1016/j.evolhumbehav.2011.05.005.

Karl Cobbe, Christopher Hesse, Jacob Hilton, and John Schulman. Leveraging procedural generation to benchmark reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119, pages 2048–2056, 2020.

Michael Cook and Susan Mineka. Observational conditioning of fear to fear-relevant versus fear-irrelevant stimuli in rhesus monkeys. *Journal of Abnormal Psychology*, 98(4):448–459, 1989. doi: 10.1037/0021-843X.98.4.448.

Matthew Crosby, Benjamin Beyret, and Marta Halina. The animal-ai testbed and competition. In *NeurIPS Workshop on Critiquing and Correcting Trends in Machine Learning*, 2019.

Gergely Csibra and György Gergely. Natural pedagogy. *Trends in Cognitive Sciences*, 13(4):148–153, 2009. doi: 10.1016/j.tics.2009.01.005.

Gergely Csibra, György Gergely, Szilvia Bíró, Orsolya Koós, and Matthew Brockbank. Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, 3(4):136–143, 1999. doi: 10.1016/S1364-6613(99)01298-2.

Nathaniel D. Daw, Yael Niv, and Peter Dayan. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12):1704–1711, 2005. doi: 10.1038/nn1560.

Nathaniel D. Daw, John P. O'Doherty, Peter Dayan, Ben Seymour, and Raymond J. Dolan. Cortical substrates for exploratory decisions in humans. *Nature*, 441:876–879, 2006.

Nathaniel D. Daw, Samuel J. Gershman, Ben Seymour, Peter Dayan, and Raymond J. Dolan. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215, 2011. doi: 10.1016/j.neuron.2011.02.027.

Peter Dayan and Bernard W. Balleine. Reward, motivation, and reinforcement learning. *Neuron*, 36(2):285–298, 2002. doi: 10.1016/S0896-6273(02)00963-7.

Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3366–3385, 2021. doi: 10.1109/TPAMI.2021.3057446.

Judy S. DeLoache and Vanessa LoBue. The narrow fellow in the grass: Human infants associate snakes and fear. *Developmental Science*, 12(1):201–207, 2009.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 4171–4186, 2019. doi: 10.18653/v1/N19-1423.

Kamran Diba and György Buzsáki. Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, 10(10):1241–1242, 2007. doi: 10.1038/nn1961.

Susanne Diekelmann and Jan Born. The memory function of sleep. *Nature Reviews Neuroscience*, 11(2):114–126, 2010. doi: 10.1038/nrn2762.

Raymond J. Dolan and Peter Dayan. Goals and habits in the brain. *Neuron*, 80(2):312–325, 2013.

Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. Palm-e: an embodied multimodal language model. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 8469–8488. JMLR.org, 2023.

Reuven Dukas and John M Ratcliffe. *Cognitive ecology II*. University of Chicago Press, 2019.

Gabriel Dulac-Arnold, Nir Levine, Daniel J. Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Machine Learning*, 110(9): 2419–2468, 2021. doi: 10.1007/s10994-021-05961-4.

Emmanuel Dupoux. Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, 173:43–59, 2018. doi: 10.1016/j.cognition.2017.11.008.

Christopher Eccleston and Geert Crombez. Pain demands attention: A cognitive–affective model of the interruptive function of pain. *Psychological Bulletin*, 125(3):356–366, 1999. doi: 10.1037/0033-2909.125.3.356.

Adrien Ecoffet, Joost Huizinga, Joel Lehman, Kenneth O. Stanley, and Jeff Clune. First return, then explore. *Nature*, 590(7847):580–586, 2021. doi: 10.1038/s41586-020-03157-9.

Jeffrey L. Elman. Learning and development in neural networks: the importance of starting small. *Cognition*, 48(1): 71–99, 1993. doi: 10.1016/0010-0277(93)90058-4.

Marc O. Ernst and Martin S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002. doi: 10.1038/415429a.

Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. Diversity is all you need: Learning skills without a reward function. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, 2019.

Linghan Fang, Tianxin Xie, and Li Liu. Boosting asr robustness via test-time reinforcement learning with audio-text semantic rewards. *arXiv preprint arXiv:2603.05231*, 2026.

Michael S. Fanselow. Neural organization of the defensive behavior system responsible for fear. *Psychonomic Bulletin & Review*, 1(4):429–438, 1994. doi: 10.3758/BF03210947.

Teresa Farroni, Gergely Csibra, Francesca Simion, and Mark H. Johnson. Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, 99(14):9602–9605, 2002.

Anne Fernald. Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8(2):181–195, 1985. doi: 10.1016/S0163-6383(85)80005-9.

Alissa L. Ferry, Susan J. Hespos, and Sandra R. Waxman. Words, but not melodies, facilitate object categorization in 9-month-old infants. *Developmental Science*, 13(2):29–34, 2010. doi: 10.1111/j.1467-7687.2009.00899.x.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning (ICML)*, pages 1126–1135, 2017.

David J. Foster and Matthew A. Wilson. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440:680–683, 2006. doi: 10.1038/nature04587.

Zipeng Fu, Tony Z. Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. *arXiv preprint arXiv:2401.02117*, 2024.

Pascale Fung, Yoram Bachrach, Asli Celikyilmaz, Kamalika Chaudhuri, Delong Chen, Willy Chung, Emmanuel Dupoux, Hongyu Gong, Hervé Jégou, Alessandro Lazaric, Arjun Majumdar, Andrea Madotto, Franziska Meier, Florian Metze, Louis-Philippe Morency, Théo Moutakanni, Juan Pino, Basile Terver, Joseph Tighe, Paden Tomasello, and Jitendra Malik. Embodied ai agents: Modeling the world, 2025. URL https://arxiv.org/abs/2506.22355.

Samir Yitzhak Gadre, Gabriel Ilharco, Alex Fang, Jonathan Hayase, Georgios Smyrnis, Thao Nguyen, Ryan Marten, Mitchell Wortsman, Dhruba Ghosh, Jieyu Zhang, Eyal Orgad, Rahim Entezari, Giannis Daras, Sarah Pratt, Vivek Ramanujan, Yonatan Bitton, Kalyani Marathe, Stephen Mussmann, Richard Vencu, Mehdi Cherti, Ranjay Krishna, Ali Farhadi, Shiry Ginosar, Alexander Toshev, Christopher Ré, Yair Carmon, Simon Kornblith, Romain Beaumont, Moritz Hardt, Robert-Jan Ness, Sarah Hooker, and Ludwig Schmidt. DataComp: In search of the next generation of multimodal datasets. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 27092–27112, 2023.

C. R. Gallistel. *The Organization of Learning*. MIT Press, Cambridge, MA, 1990. ISBN 9780262071260.

C. R. Gallistel. The neuroscience of learning: Beyond the Hebbian synapse. *Annual Review of Psychology*, 64:169–200, 2013. doi: 10.1146/annurev-psych-113011-143807.

Quentin Garrido, Nicolas Ballas, Mahmoud Assran, Adrien Bardes, Laurent Najman, Michael Rabbat, Emmanuel Dupoux, and Yann LeCun. Intuitive physics understanding emerges from self-supervised pretraining on natural videos. *arXiv preprint arXiv:2502.11831*, 2025.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A. Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, 2020. doi: 10.1038/s42256-020-00257-z.

György Gergely, Harold Bekkering, and Ildikó Király. Rational imitation in preverbal infants. *Nature*, 415:755, 2002. doi: 10.1038/415755a.

James J. Gibson. *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston, MA, 1966.

Alison Gopnik. Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, 337(6102):1623–1627, 2012. doi: 10.1126/science.1223416.

Alison Gopnik, Andrew N. Meltzoff, and Patricia K. Kuhl. *The Scientist in the Crib: Minds, Brains, and How Children Learn*. William Morrow & Co, New York, NY, 1999.

Alison Gopnik, Shaun O'Grady, Christopher G. Lucas, Thomas L. Griffiths, Adrienne Wente, Sophie Bridgers, Rosie Aboody, Hoki Fung, and Ronald E. Dahl. Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences*, 114(30): 7892–7899, 2017. doi: 10.1073/pnas.1705834114.

Jacqueline Gottlieb, Pierre-Yves Oudeyer, Manuel Lopes, and Adrien Baranes. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, 17(11):585–593, 2013. doi: 10.1016/j.tics.2013.09.001.

Louise Goupil, Margaux Romand-Monnier, and Sid Kouider. Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences*, 113(13):3492–3496, 2016. doi: 10.1073/pnas.1515129113.

Anirudh Goyal and Yoshua Bengio. Inductive biases for deep learning of higher-level cognition. *Proceedings of the Royal Society A*, 478(2266):20210068, 2022. doi: 10.1098/rspa.2021.0068.

Anirudh Goyal, Aniket Didolkar, Alex Lamb, Kartikeya Badola, Nan Rosemary Ke, Nasim Rahaman, Jonathan Binas, Charles Blundell, Michael Mozer, and Yoshua Bengio. Coordination among neural modules through a shared global workspace. In *Proceedings of the 10th International Conference on Learning Representations (ICLR)*, 2022.

Alex Graves, Marc G. Bellemare, Jacob Menick, Rémi Munos, and Koray Kavukcuoglu. Automated curriculum learning for neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70, pages 1311–1320, 2017.

David J. Gunkel. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press, 2012.

Todd M. Gureckis and Douglas B. Markant. Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, 7(5):464–481, 2012. doi: 10.1177/1745691612454304.

Hyowon Gweon, Hannah Pelton, Jaclyn A. Konopka, and Laura E. Schulz. Sins of omission: Children incompletely explore when teachers are under-informative. *Cognition*, 132(3):335–342, 2014. doi: 10.1016/j.cognition.2014.04.013.

David Ha and Jürgen Schmidhuber. World models. In *International Conference on Learning Representations (ICLR)*, 2018.

Mijna Hadders-Algra. Early human motor development: From variation to selection. *Neuroscience & Biobehavioral Reviews*, 90:411–427, 2018. doi: 10.1016/j.neubiorev.2018.05.010.

Raia Hadsell, Dushyant Rao, Andrei A. Rusu, and Razvan Pascanu. Embracing change: Continual learning in deep neural networks. *Trends in Cognitive Sciences*, 24(12):1028–1040, 2020. doi: 10.1016/j.tics.2020.09.004.

Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning latent dynamics for planning from pixels. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97, pages 2555–2565, 2019.

Danijar Hafner, Timothy Lillicrap, Jimmy Ba, Mohammad Norouzi, and James Davidson. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations (ICLR)*, 2020.

Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*, 2023. URL https://arxiv.org/abs/2301.04104.

Robert R. Hampton. Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, 98(9):5359–5362, 2001. doi: 10.1073/pnas.101074698.

Jessica B. Hamrick, Andrew J. Ballard, Razvan Pascanu, Oriol Vinyals, Nicolas Heess, and John Battenberg. Metacontrol for adaptive imagination-based optimization. In *International Conference on Learning Representations (ICLR)*, 2017.

Xing Han, Huy Nguyen, Carl William Harris, Nhat Ho, and Suchi Saria. Fusemoe: Mixture-of-experts transformers for fleximodal fusion. In *Advances in Neural Information Processing Systems 37*, 2024.

Paul L. Harris and Kathleen H. Corriveau. Young children's selective trust in informants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1179–1187, 2011. doi: 10.1098/rstb.2010.0321.

Demis Hassabis, Dharshan Kumaran, Christopher Summerfield, and Matthew Botvinick. Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258, 2017. doi: 10.1016/j.neuron.2017.06.011.

Daniel B. M. Haun and Michael Tomasello. Children conform to the behavior of peers; other great apes stick with what they know. *Psychological Science*, 22(12):1421–1428, 2011. doi: 10.1177/0956797611418712. Actually published in 2011, correcting citation year.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9712–9721, 2020. doi: 10.1109/CVPR42600.2020.00973.

Joseph Henrich. *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton University Press, Princeton, NJ, 2016.

Takao K. Hensch. Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience*, 6(11):877–888, 2005. doi: 10.1038/nrn1787.

Cecilia Heyes. *Cognitive Gadgets: The Cultural Evolution of Thinking*. Harvard University Press, Cambridge, MA, 2018.

Felix Hill, Andrew Lampinen, Rafael Schneider, Stephen Clark, Matthew Botvinick, Brenden Lake, and Adam Santoro. Grounded language learning fast and slow. In *International Conference on Learning Representations (ICLR)*, 2020.

Shannon Hsieh, Roy E Plotnick, and Andrew M Bush. The phanerozoic aftermath of the cambrian information revolution: sensory and cognitive complexity in marine faunas. *Paleobiology*, 48(3):397–419, 2022.

Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3451–3460, 2021. doi: 10.1109/TASLP.2021.3122291.

Peter R. Huttenlocher. Synaptic density in human frontal cortex—developmental changes and effects of aging. *Brain Research*, 163(2):195–205, 1979. doi: 10.1016/0006-8993(79)90349-4.

Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991.

Jinwoo Jang, Minjong Yoo, Sihyung Yoon, and Honguk Woo. Test-time mixture of world models for embodied agents in dynamic environments. *arXiv preprint arXiv:2601.22647*, 2026.

Mark H. Johnson. Functional brain development in humans. *Nature Reviews Neuroscience*, 2(7):475–483, 2001. doi: 10.1038/35081509.

Mark H. Johnson, Suzanne Dziurawiec, Hadyn Ellis, and John Morton. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40(1-2):1–19, 1991. doi: 10.1016/0010-0277(91)90045-6.

Philip Nicholas Johnson-Laird. *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press, 1983.

Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York, NY, 2011. ISBN 9780374275631.

Julia W. Y. Kam, Elizabeth Dao, Joelle Farley, Kevin Fitzpatrick, Jonathan Smallwood, Jonathan W. Schooler, and Todd C. Handy. Slow fluctuations in attentional control of sensory tracking. *Journal of Cognitive Neuroscience*, 23 (2):460–470, 2011. doi: 10.1162/jocn.2010.21443.

Min Jeong Kang, Ming Hsu, Ian M. Krajbich, George Loewenstein, Samuel M. McClure, Joseph T.-Y. Wang, and Colin F. Camerer. The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, 20(8):963–973, 2009. doi: 10.1111/j.1467-9280.2009.02402.x.

Annette Karmiloff-Smith. *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press, Cambridge, MA, 1992. ISBN 9780262111690.

Aya Kayal, Eduardo Pignatelli, and Laura Toni. The impact of intrinsic rewards on exploration in reinforcement learning. *Neural Computing and Applications*, pages 1–35, 2025. doi: 10.1007/s00521-025-10874-y.

Fred Keijzer. Moving and sensing without input and output: early nervous systems and the origins of the animal sensorimotor organization. *Biology & Philosophy*, 30(3):311–331, 2015.

Rachel L. Kendal, Lydia M. Hopper, Andrew Whiten, Sarah F. Brosnan, Susan P. Lambeth, Steven J. Schapiro, and William Hoppitt. Chimpanzees copy dominant and knowledgeable individuals: An empirical test of the prestige bias hypothesis. *Evolution and Human Behavior*, 36(1):65–72, 2015. doi: 10.1016/j.evolhumbehav.2014.09.002.

Celeste Kidd, Steven T. Piantadosi, and Richard N. Aslin. The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, 7(5):e36399, 2012. doi: 10.1371/journal.pone.0036399.

Katherine D. Kinzler, Emmanuel Dupoux, and Elizabeth S. Spelke. The native language of social cognition. *Proceedings of the National Academy of Sciences*, 104(30):12577–12580, 2007. doi: 10.1073/pnas.0705305104.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. doi: 10.1073/pnas.1611835114.

Melissa A. Koenig, Fabrice Clément, and Paul L. Harris. Trust in testimony: Children's use of true and false statements. *Psychological Science*, 15(10):694–698, 2004. doi: 10.1111/j.0956-7976.2004.00742.x.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Swan, William Sandina, Chhavi Shiragur, Sara Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. WILDS: A benchmark of in-the-wild distribution shifts. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139, pages 5637–5664, 2021.

Yuto Kojima, Jiarui Xu, Xueyan Zou, and Xiaolong Wang. Lora-ttt: Low-rank test-time training for vision-language models. *arXiv preprint arXiv:2502.02069*, 2025.

Nils Kolling, Timothy E. J. Behrens, Rogier B. Mars, and Matthew F. S. Rushworth. Neural mechanisms of foraging. *Science*, 336(6077):95–98, 2012. doi: 10.1126/science.1216930.

Mark A Krause, Karen L Hollis, and Mauricio R Papini. *Evolution of learning and memory mechanisms*. Cambridge University Press, 2022.

Diego Kreutz, Fernando M. V. Ramos, Paulo Esteves Verissimo, Christian Esteve Rothenberg, Siamak Azodolmolky, and Steve Uhlig. Software-defined networking: A comprehensive survey. *Proceedings of the IEEE*, 103(1):14–76, 2015. doi: 10.1109/JPROC.2014.2371999.

Dharshan Kumaran, Demis Hassabis, and James L. McClelland. What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends in Cognitive Sciences*, 20(7):512–534, 2016. doi: 10.1016/j.tics.2016.05.004.

John E. Laird. *The Soar Cognitive Architecture*. MIT Press, Cambridge, MA, 2012.

Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40:e253, 2017.

Kushal Lakhotia, Eugene Kharitonov, Wei-Ning Hsu, Yossi Adi, Adam Polyak, Benjamin Bolte, Tu-Anh Nguyen, Jade Copet, Alexei Baevski, Abdelrahman Mohamed, and Emmanuel Dupoux. On generative spoken language modeling from raw audio. *Transactions of the Association for Computational Linguistics*, 9:1336–1354, 2021. doi: 10.1162/tacl_a_00430.

Michael Laskin, Aravind Srinivas, and Pieter Abbeel. CURL: Contrastive unsupervised representations for reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119, pages 5639–5650, 2020.

Marvin Lavechin, Yaya Sy, Hadrien Titeux, María Andrea Cruz Blandón, Okko Räsänen, Hervé Bredin, Emmanuel Dupoux, and Alejandrina Cristia. BabySLM: language-acquisition-friendly benchmark of self-supervised spoken language models. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 4588–4592, 2023. doi: 10.21437/Interspeech.2023-978.

Yann LeCun. A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27. *OpenReview*, 2022. URL https://openreview.net/forum?id=BZ5a1r-kVsf.

Yuqing Lei, Yingjun Du, Yawen Huang, Xiantong Zhen, and Ling Shao. Metatpt: Meta test-time prompt tuning for vision-language models. *arXiv preprint arXiv:2512.12268*, 2025.

Jan Leike, Miljan Martic, Victoria Krakovna, Pedro A. Ortega, Tom Everitt, Andrew Lefrancq, Laurent Orseau, and Shane Legg. AI safety gridworlds. *arXiv preprint arXiv:1711.09883*, 2017. URL https://arxiv.org/abs/1711.09883.

Qiyang Li, Zhiyuan Zhou, and Sergey Levine. Reinforcement learning with action chunking. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, 2025.

Anthony Liang, Pavel Czempin, Matthew Hong, Yutai Zhou, Erdem Biyik, and Stephen Tu. CLAM: Continuous latent action models for robot learning from unlabeled demonstrations. *arXiv preprint arXiv:2505.04999*, 2025.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. In *Proceedings of the 4th International Conference on Learning Representations (ICLR)*, 2016.

Bill Yuchen Lin, Yicheng Fu, Karina Yang, Faeze Brahman, Shiyu Huang, Chandra Bhagavatula, Prithviraj Ammanabrolu, Yejin Choi, and Xiang Ren. Swiftsage: A generative agent with fast and slow thinking for complex interactive tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

Changyu Liu, Yiyang Liu, Taowen Wang, Qiao Zhuang, James Chenhao Liang, Wenhao Yang, Renjing Xu, Qifan Wang, Dongfang Liu, and Cheng Han. On-the-fly vla adaptation via test-time reinforcement learning. *arXiv preprint arXiv:2601.06748*, 2026.

Jonathan Lorraine, Paul Vicol, and David Duvenaud. Optimizing millions of hyperparameters by implicit differentiation. In *Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 108, pages 1540–1552, 2020.

William Lotter, Gabriel Kreiman, and David Cox. Deep predictive coding networks for video prediction and unsupervised learning. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*, 2017.

Kristen E. Lyons and Simona Ghetti. Metacognitive development in early childhood: New questions about old assumptions. In A. Efklides and P. Misailidi, editors, *Trends and prospects in metacognition research*, page 259–278. Springer Science + Business Media, 2010. doi: https://doi.org/10.1007/978-1-4419-6546-2_12.

Yecheng Jason Ma, Shagun Sodhani, Dinesh Jayaraman, Osbert Bastani, Vikash Kumar, and Amy Zhang. VIP: Towards universal visual reward and representation via value-implicit pre-training. In *Proceedings of the 10th International Conference on Learning Representations (ICLR)*, 2023.

Sijie Mai, Shiqin Han, and Haifeng Hu. Addressing missing and noisy modalities in one solution: Unified modality-quality framework for low-quality multimodal data. *arXiv preprint arXiv:2603.02695*, 2026.

Marcelo G. Mattar and Nathaniel D. Daw. Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, 21:1609–1617, 2018. doi: 10.1038/s41593-018-0232-z.

Andrew N. Meltzoff. 'like me': a foundation for social cognition. *Developmental Science*, 10(1):126–134, 2007. doi: 10.1111/j.1467-7687.2007.00574.x.

Luke Metz, C. Daniel Freeman, Samuel S. Schoenholz, and Tal Kachman. Gradients are not all you need. *arXiv preprint arXiv:2111.05803*, 2021. URL https://arxiv.org/abs/2111.05803.

Earl K. Miller and Jonathan D. Cohen. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1):167–202, 2001. doi: 10.1146/annurev.neuro.24.1.167.

Thomas M. Moerland, Joost Broekens, Aske Plaat, and Catholijn M. Jonker. Model-based reinforcement learning: A survey. *Foundations and Trends® in Machine Learning*, 16(1):1–118, 2023. doi: 10.1561/2200000086.

Shakir Mohamed and Danilo Jimenez Rezende. Variational information maximisation for intrinsically motivated reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 28, pages 2125–2133, 2015.

Zoltán Molnár, Heiko J Luhmann, and Patrick O Kanold. Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science*, 370(6514):eabb2153, 2020.

Chris Moore. Social learning in children. *Cognitive Science*, 37(1):142–167, 2013.

Mohammad Mahdi Moradi, Hossam Amer, Sudhir Mudur, Weiwei Zhang, Yang Liu, and Walid Ahmed. Continuous self-improvement of large language models by test-time training with verifier-driven sample selection. *arXiv preprint arXiv:2505.19475*, 2025.

John Morton and Mark H. Johnson. CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review*, 98(2):164–181, 1991. doi: 10.1037/0033-295X.98.2.164.

Suraj Nair, Aravind Rajeswaran, Vikash Kumar, Chelsea Finn, and Abhinav Gupta. R3m: A universal visual representation for robot manipulation. In *Proceedings of the 6th Conference on Robot Learning (CoRL)*, 2022.

James S. Nairne, Joshua E. VanArsdall, Josefa N. S. Pandeirada, Megan Cogdill, and Tyler LeFront. Survival processing: The animacy effect in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(1): 276–287, 2013. doi: 10.1037/a0029022.

Arne Õhman and Susan Mineka. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, 108(3):483–522, 2001. doi: 10.1037/0033-295X.108.3.483.

D. Kimbrough Oller and Rebecca E. Eilers. The role of audition in infant babbling. *Child Development*, 59(2):441–449, 1988. doi: 10.2307/1130323.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael G. Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research (TMLR)*, 2024. URL https://openreview.net/forum?id=a68SUt6zFt.

Pierre-Yves Oudeyer and Frédéric Kaplan. What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, 1:6, 2007. doi: 10.3389/neuro.12.006.2007.

Pierre-Yves Oudeyer, Frdric Kaplan, and Verena V Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation*, 11(2):265–286, 2007. doi: 10.1109/TEVC.2006.890271.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pages 27730–27744, 2022.

Andrei Paleyes, Raoul-Gabriel Urma, and Neil D. Lawrence. Challenges in deploying machine learning: A survey of case studies. *ACM Computing Surveys (CSUR)*, 55(6):1–29, 2022. doi: 10.1145/3533378.

Olivier Pascalis, Michelle de Haan, and Charles A Nelson. Is face processing species-specific during the first year of life? *Science*, 296(5571):1321–1323, 2002. doi: 10.1126/science.1070223.

Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning (ICML)*, pages 2778–2787. PMLR, 2017.

Michael G Paulin and Joseph Cahill-Lane. Events in early nervous system evolution. *Topics in Cognitive Science*, 13 (1):25–44, 2021.

John M. Pearce and Geoffrey Hall. A model for pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87(6):532–552, 1980. doi: 10.1037/0033-295X.87.6.532.

Brad E. Pfeiffer and David J. Foster. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447):74–79, 2013. doi: 10.1038/nature12112.

Jean Piaget. *The origins of intelligence in children*. International University Press, New York, 1952.

Roy E Plotnick, Stephen Q Dornbos, and Junyuan Chen. Information landscapes and sensory ecology of the cambrian radiation. *Paleobiology*, 36(2):303–317, 2010.

Vitchyr H. Pong, Murtaza Dalal, Steven Lin, Ashvin Nair, Shikhar Bahl, and Sergey Levine. Skew-fit: State-covering self-supervised reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119, pages 7783–7792, 2020.

Rémy Portelas, Cédric Colas, Lilian Weng, Katja Hoffman, and Pierre-Yves Oudeyer. Automatic curriculum learning for deep rl: A short survey. *arXiv preprint arXiv:2003.04664*, 2020.

Diane Poulin-Dubois, Ivy Brooker, and Alexandra Polonia. Infants prefer to attend to a reliable looker. *British Journal of Developmental Psychology*, 29(2):341–353, 2011. doi: 10.1111/j.2044-835X.2010.02019.x.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. Technical report, OpenAI, 2018.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139, pages 8748–8763, 2021.

Ilija Radosavovic, Tete Xiao, Stephen James, Pieter Abbeel, Jitendra Malik, and Trevor Darrell. Real-world robot learning with masked visual pre-training. In *Proceedings of the 6th Conference on Robot Learning (CoRL)*, 2022.

Ilija Radosavovic, Sarthak Kamat, Trevor Darrell, and Jitendra Malik. Learning humanoid locomotion over challenging terrain. *arXiv preprint arXiv:2410.03654*, 2024a.

Ilija Radosavovic, Bike Zhang, Baifeng Shi, Jathushan Rajasegaran, Sarthak Kamat, Trevor Darrell, Koushil Sreenath, and Jitendra Malik. Humanoid locomotion as next token prediction. *Advances in neural information processing systems*, 37:79307–79324, 2024b.

Marcus E. Raichle. The brain's default mode network. *Annual Review of Neuroscience*, 38:433–447, 2015. doi: 10.1146/annurev-neuro-071013-014030.

Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999. doi: 10.1038/4580.

Björn Rasch and Jan Born. About sleep's role in memory. *Physiological Reviews*, 93(2):681–766, 2013. doi: 10.1152/physrev.00032.2012.

Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc V. Le. Regularized evolution for image classifier architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4780–4789, 2019. doi: 10.1609/aaai.v33i01.33014780.

Jonathan Redshaw and Thomas Suddendorf. Children's capacity to imagine and prepare for alternative future possibilities. *Cognition*, 150:104–113, 2016. doi: 10.1016/j.cognition.2016.02.007.

Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gómez Colmenarejo, Alexander Novikov, Gabriel Barth-maron, Mai Giménez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, Tom Eccles, Jake Bruce, Ali Razavi, Ashley Edwards, Nicolas Heess, Yutian Chen, Raia Hadsell, Oriol Vinyals, Mahyar Bordbar, and Nando de Freitas. A generalist agent. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=1ikK0kHjvj.

Pengzhen Ren, Yun Xiao, Xiaojun Chang, Po-Yao Huang, Zhihui Li, Xiaojiang Chen, and Xin Wang. A survey of deep active learning. *ACM Computing Surveys (CSUR)*, 54(9):1–40, 2021. doi: 10.1145/3472291.

Mark B Ring. Child: A first step towards continual learning. *Machine Learning*, 28(1):77–104, 1997. doi: 10.1023/A:1007321720083.

Giacomo Rizzolatti and Laila Craighero. The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192, 2004. doi: 10.1146/annurev.neuro.27.070203.144230.

David A. Robinson. Control of eye movements. *Comprehensive Physiology*, pages 1275–1320, 1981.

Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, Hoboken, NJ, 4th edition, 2020.

Jenny R Saffran and Natasha Z Kirkham. Infant statistical learning. *Annual review of psychology*, 69(1):181–203, 2018. doi: 10.1146/annurev-psych-122216-011805.

Jenny R. Saffran, Richard N. Aslin, and Elissa L. Newport. Statistical learning by 8-month-old infants. *Science*, 274 (5294):1926–1928, 1996. doi: 10.1126/science.274.5294.1926.

Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A platform for embodied ai research. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 9339–9347, 2019.

Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. In *International Conference on Learning Representations (ICLR)*, 2016.

Benjamin Schmid, Katja Karg, Josef Perner, and Michael Tomasello. Great apes are sensitive to prior reliability of an informant in a gaze following task. *PLoS One*, 12(11):e0187451, 2017. doi: 10.1371/journal.pone.0187451.

Jürgen Schmidhuber. *Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-meta... hook*. PhD thesis, Technische Universität München, 1987.

Jürgen Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. In *Proceedings of the First International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, pages 222–227, Cambridge, MA, 1991. MIT Press.

Jürgen Schmidhuber. Formal theory of creativity, fun, and intrinsic motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development*, 2(3):230–247, 2010. doi: 10.1109/TAMD.2010.2056368.

Bernhard Schölkopf. Causality for machine learning. *Information Extraction: Towards Scalable, Adaptable and Distributed Processing*, pages 1–20, 2019. URL https://arxiv.org/abs/1911.10500. Also widely cited as arXiv preprint arXiv:1911.10500.

Brian J. Scholl and Patrice D. Tremoulet. Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8): 299–309, 2000.

Julian Schrittwieser, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, Edward Lockhart, Demis Hassabis, Thore Graepel, Timothy Lillicrap, and David Silver. Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588:604–609, 2020. doi: 10.1038/s41586-020-03051-4.

Wolfram Schultz, Peter Dayan, and P. Read Montague. A neural substrate of prediction and reward. *Science*, 275 (5306):1593–1599, 1997. doi: 10.1126/science.275.5306.1593.

Lars Schwabe and Oliver T. Wolf. Stress prompts habit behavior in humans. *The Journal of Neuroscience*, 29(22): 7191–7198, 2009. doi: 10.1523/JNEUROSCI.0979-09.2009.

Max Schwarzer, Ankesh Anand, Rishab Goel, R. Devon Hjelm, Aaron Courville, and Philip Bachman. Data-efficient reinforcement learning with self-predictive representations. In *Proceedings of the 9th International Conference on Learning Representations (ICLR)*, 2021.

David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, and Dan Dennison. Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 28, pages 2503–2511, 2015.

Pierre Sermanet, Kelvin Xu, and Sergey Levine. Unsupervised perceptual rewards for imitation learning. In *Proceedings of Robotics: Science and Systems (RSS)*, 2017.

Burr Settles. Active learning literature survey. *Computer Sciences Technical Report 1648*, 2009. URL https://minds.wisconsin.edu/handle/1793/60660.

Reza Shadmehr and Ferdinando A. Mussa-Ivaldi. Adaptive representation of dynamics during learning of a motor task. *Journal of Neuroscience*, 14(5):3208–3224, 1994. doi: 10.1523/JNEUROSCI.14-05-03208.1994.

Amitai Shenhav, Sebastian Musslick, Falk Lieder, Wouter Kool, Thomas L. Griffiths, Jonathan D. Cohen, and Matthew M. Botvinick. Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40:99–124, 2017. doi: 10.1146/annurev-neuro-072116-031526.

Stephen V. Shepherd, Robert O. Deaner, and Michael L. Platt. Social status gates social attention in monkeys. *Current Biology*, 16(4):R119–R120, 2006. doi: 10.1016/j.cub.2006.02.013.

David Silver and Richard S Sutton. Welcome to the era of experience. *Google AI*, 1, 2025.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of go without human knowledge. *Nature*, 550: 354–359, 2017. doi: 10.1038/nature24270.

Francesca Simion, Lucia Regolin, and Hermann Bulf. A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, 105(2):809–813, 2008.

Ankur Sinha, Pekka Malo, and Kalyanmoy Deb. A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295, 2018. doi: 10.1109/TEVC.2017.2712906.

Linda B. Smith and Michael Gasser. The development of embodied cognition: Six lessons from babies. *Artificial Life*, 11(1-2):13–29, 2005. doi: 10.1162/1064546053278973.

Linda B. Smith, Swapna Jayaraman, Elizabeth Clerkin, and Chen Yu. The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, 22(4):325–336, 2018. doi: 10.1016/j.tics.2018.02.004.

Charlie Snell, Jaehoon Lee, Kelvin Xu, and Aviral Kumar. Scaling llm test-time compute optimally can be more effective than scaling model parameters. *arXiv preprint arXiv:2408.03314*, 2024.

Evgenii Nikolaevich Sokolov. *Higher nervous activity and the problem of perception*. Pergamon Press, Oxford, England, 1963.

Elizabeth S Spelke and Katherine D Kinzler. Core knowledge. *Developmental science*, 10(1):89–96, 2007. doi: 10.1111/j.1467-7687.2007.00569.x.

Elizabeth S Spelke, Karen Breinlinger, Janet Macomber, and Kristen Jacobson. Origins of knowledge. *Psychological review*, 99(4):605, 1992. doi: 10.1037/0033-295X.99.4.605.

Dan Sperber, Fabrice Clément, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson. Epistemic vigilance. *Mind & Language*, 25(4):359–393, 2010. doi: 10.1111/j.1468-0017.2010.01394.x.

Aimee E. Stahl and Lisa Feigenson. Observing the unexpected enhances infants' learning and exploration. *Science*, 348(6230):91–94, 2015. doi: 10.1126/science.aaa3799.

Kenneth O. Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. In *Evolutionary Computation*, volume 10, pages 99–127, 2002. doi: 10.1162/106365602320169811.

George F Striedter. *Principles of brain evolution*. Sinauer associates, 2005.

Francys Subiaul, Jessica F. Cantlon, Robert L. Holloway, and Herbert S. Terrace. Cognitive imitation in rhesus macaques. *Science*, 305(5682):407–410, 2004. doi: 10.1126/science.1099136.

Thomas Suddendorf and Michael C. Corballis. The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, 30(3):299–313, 2007. doi: 10.1017/s0140525x07001975.

Xin Sun, Zhongqi Chen, Qiang Liu, Shu Wu, Bowen Song, Weiqiang Wang, Zilei Wang, and Liang Wang. Predict the retrieval! test time adaptation for retrieval augmented generation. *arXiv preprint arXiv:2601.11443*, 2026.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 2nd edition, 2018.

Akiko Takaoka, Takaka Maeda, Yutaka Hori, and Kazuo Fujita. Do dogs follow behavioral cues from an unreliable human? *Animal Cognition*, 18(2):475–483, 2015. doi: 10.1007/s10071-014-0816-2.

Chen Tang, Ben Abbatematteo, Jiaheng Hu, Rohan Chandra, Roberto Martín-Martín, and Peter Stone. Deep reinforcement learning for robotics: A survey of real-world successes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 28694–28698, 2025.

Alex H. Taylor, Douglas M. Elliffe, Gavin R. Hunt, Nathan J. Emery, Nicola S. Clayton, and Russell D. Gray. New caledonian crows learn the functional properties of novel tool types. *PLoS ONE*, 7(12):e49374, 2012.

Annalisa T Taylor, Thomas A Berrueta, and Todd D Murphey. Active learning in robotics: A review of control principles. *Mechatronics*, 77:102576, 2021. doi: 10.1016/j.mechatronics.2021.102576.

Bahey Tharwat, Yara Nasser, Ali Abouzeid, and Ian Reid. Latent action pretraining through world modeling. *arXiv preprint arXiv:2509.18428*, 2025.

Esther Thelen and Linda B. Smith. *A Dynamic Systems Approach to the Development of Cognition and Action*. MIT Press, Cambridge, MA, 1994.

Erik D. Thiessen, Emily A. Hill, and Jenny R. Saffran. Infant-directed speech facilitates word segmentation. *Infancy*, 7 (1):53–71, 2005. doi: 10.1207/s15327078in0701_5.

Sebastian Thrun. Lifelong learning algorithms. In *Learning to learn*, pages 181–209. Springer, 1998. doi: 10.1007/978-1-4615-5529-2_8.

Sebastian Thrun and Lorien Pratt. Learning to learn: Introduction and overview. In *Learning to Learn*, pages 3–17. Springer, 1998.

Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Jürgen Schmidhuber, and Phillip Isola. What makes for good views for contrastive learning? *Advances in Neural Information Processing Systems (NeurIPS)*, 33:6827–6839, 2020.

Michael Tomasello. *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge, MA, 1999.

Michael Tomasello. *Why We Cooperate*. MIT Press, Cambridge, MA, 2009.

Michael Tomasello, Josep Call, and Brian Hare. Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, 55(4):1063–1069, 1998.

Giulio Tononi and Chiara Cirelli. Sleep and the price of plasticity: from synaptic and cellular homeostasis to memory consolidation and integration. *Neuron*, 81(1):12–34, 2014. doi: 10.1016/j.neuron.2013.12.025.

John Tooby and Leda Cosmides. The psychological foundations of culture. In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, pages 19–136. Oxford University Press, 1992.

Endel Tulving. Episodic memory and autonoesis: Uniquely human? In Herbert S. Terrace and Janet Metcalfe, editors, *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*, pages 3–56. Oxford University Press, New York, NY, 2005.

Alan M. Turing. Computing machinery and intelligence. *Mind*, 59(236):433–460, 1950.

Gerry Turkewitz and Patrick A. Kenny. Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology*, 15(4):357–368, 1982. doi: 10.1002/dev.420150409.

Sherry Turkle. *Alone Together: Why We Expect More from Technology and Less from Each Other*. Basic Books, 2011.

Michael T Turvey. Coordination. *American psychologist*, 45(8):938, 1990.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. URL https://arxiv.org/abs/1807.03748.

Athena Vouloumanos and Janet F. Werker. Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, 10(2):159–164, 2007. doi: 10.1111/j.1467-7687.2007.00549.x.

Lev Semenovich Vygotsky and Michael Cole. *Mind in society: Development of higher psychological processes*. Harvard university press, 1978.

Edgar Welte, Yitian Shi, Rosa Wolf, Maximillian Gilles, and Rania Rayyes. Flowcorrect: Efficient interactive correction of generative flow policies for robotic manipulation. *arXiv preprint arXiv:2602.22056*, 2026.

Janet F Werker and Richard C Tees. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63, 1984. doi: 10.1016/S0163-6383(84)80022-3.

Andrew Whiten and Rebecca Ham. On the nature and evolution of imitation in the animal kingdom: Reappraisal of a century of research. In *Advances in the Study of Behavior*, volume 21, pages 239–283. Elsevier, 1992.

Andrew Whiten, Victoria Horner, and Frans B. M. de Waal. Conformity to cultural norms of tool use in chimpanzees. *Nature*, 437(7059):737–740, 2005. doi: 10.1038/nature04047.

Matthew A. Wilson and Bruce L. McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679, 1994. doi: 10.1126/science.8036517.

Robert C. Wilson, Andra Geana, John M. White, Elliot A. Ludvig, and Jonathan D. Cohen. Humans use directed and random exploration to solve the explore-exploit dilemma. *Journal of Experimental Psychology: General*, 143(6): 2074–2081, 2014. doi: 10.1037/a0038199.

David Wood, Jerome S. Bruner, and Gail Ross. The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2):89–100, 1976. doi: 10.1111/j.1469-7610.1976.tb00381.x.

Yijie Xu, Huizai Yao, Zhiyu Guo, Weiyu Guo, Pengteng Li, Aiwei Liu, Xuming Hu, and Hui Xiong. You only need 4 extra tokens: Synergistic test-time adaptation for llms. *arXiv preprint arXiv:2510.10223*, 2025.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

Denis Yarats, Amy Zhang, Ilya Kostrikov, Brandon Amos, Joelle Pineau, and Rob Fergus. CURL: Contrastive unsupervised representations for reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119, pages 10490–10500, 2020.

Melis Yilmaz and Markus Meister. Rapid innate defensive responses of mice to looming visual stimuli. *Current Biology*, 23(19):2011–2015, 2013.

Ge Yuan, Qiyuan Qiao, Jing Zhang, and Dong Xu. Adaworldpolicy: World-model-driven diffusion policy with online adaptive learning for robotic manipulation. *arXiv preprint arXiv:2602.20057*, 2026.

Anthony M. Zador. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nature Communications*, 10(1):3770, 2019. doi: 10.1038/s41467-019-11786-6.

Eric Zelikman, Eliana Lorch, Lester Mackey, and Noah D. Goodman. Quiet-star: Language models can teach themselves to think before speaking. *arXiv preprint arXiv:2403.09629*, 2024.

Tony Z. Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. In *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

Fang Zhengxin, Yuan Yi, Zhang Jingyu, Liu Yue, Mu Yuechen, Lu Qinghua, Xu Xiwei, Wang Jeff, Wang Chen, Zhang Shuai, and Chen Shiping. Mlops spanning whole machine learning life cycle: A survey, 2023. URL https://arxiv.org/abs/2304.07296.

Chunting Zhou, Lili Yu, Arun Babu, Kushal Tirumala, Michihiro Yasunaga, Leonid Shamis, Jacob Kahn, Xuezhe Ma, Luke Zettlemoyer, and Omer Levy. Transfusion: Predict the next token and diffuse images with one multi-modal model, 2024. URL https://arxiv.org/abs/2408.11039.

Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, Quan Vuong, Vincent Vanhoucke, Huong Tran, Radu Soricut, Anikait Singh, Jaspiar Singh, Pierre Sermanet, Pannag R. Sanketi, Grecia Salazar, Michael S. Ryoo, Krista Reymann, Kanishka Rao, Karl Pertsch, Igor Mordatch, Henryk Michalewski, Yao Lu, Sergey Levine, Lisa Lee, Tsang-Wei Edward Lee, Isabel Leal, Yuheng Kuang, Dmitry Kalashnikov, Ryan Julian, Nikhil J. Joshi, Alex Irpan, Brian Ichter, Jasmine Hsu, Alexander Herzog, Karol Hausman, Keerthana Gopalakrishnan, Chuyuan Fu, Pete Florence, Chelsea Finn, Kumar Avinava Dubey, Danny Driess, Tianli Ding, Krzysztof Marcin Choromanski, Xi Chen, Yevgen Chebotar, Justice Carbajal, Noah Brown, Anthony Brohan, Montserrat Gonzalez Arenas, and Kehang Han. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In Jie Tan, Marc Toussaint, and Kourosh Darvish, editors, *Proceedings of*

The 7th Conference on Robot Learning, volume 229 of *Proceedings of Machine Learning Research*, pages 2165–2183. PMLR, 06–09 Nov 2023. URL https://proceedings.mlr.press/v229/zitkovich23a.html.

Norbert Zmyj, David Buttelmann, Malinda Carpenter, and Moritz M. Daum. The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, 106(4):208–220, 2010. doi: 10.1016/j.jecp.2010.03.002.

Barret Zoph and Quoc V. Le. Neural architecture search with reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2017.