Stop Spatializing Time: Machine Learning Agents Should Learn *Through* Time, Not *About* Time

Teeratham Vitchutripop [*1] **Alyssa Quarles** [1] **Wenhe Zhang** [1] **Daniel Rakita** [1]

Abstract

Modern machine learning systems are increasingly deployed in settings that require persistent interaction, adaptation, memory, and decision-making over time. Yet, most learning paradigms remove the temporal pressures faced by physically embedded agents: the world waits for computation, failures are erased by resets, past experience can be replayed exactly, memory is treated as static storage, and learning is often separated from the irreversible trajectory of the learner. We argue that these assumptions limit progress toward agents that can learn through experience over extended lifetimes. This paper proposes six guidelines for temporally grounded learning: autonomous external time, partial embodied observability, non-resettable lifetime interaction, bounded reconstructive memory, multi-rate endogenous computation, and path-dependent self-modification. We use these guidelines to compare major machine learning paradigms, showing that existing approaches capture important fragments of temporal grounding while relaxing, externalizing, or assuming away other temporal pressures. We then present calls to action for the community to build benchmark ecosystems that make temporal grounding measurable, audit the temporal assumptions inside agent architectures, create shared venues around time, and reward temporally grounded work during review. Our central claim is that modeling temporal structure is not sufficient for persistent agency: artificial agents should learn, act, remember, and adapt within the same irreversible time in which their worlds unfold.

[1]Department of Computer Science, Yale University, New Haven, CT, USA. Correspondence to: Teeratham Vitchutripop <tj.vitchutripop@yale.edu>.

Preprint. July 23, 2026.

1. Introduction

The "life" of an agent instantiated through machine learning is often untethered from temporal pressures. A robot that drops a glass in simulation can rely on `env.reset()` to perfectly recover an intact glass again; a language model can ponder over a frozen prompt indefinitely and begin a new session without any trace of what happened; a world model-based agent can consider thousands of possible futures while the real world continues on; and neural networks can optimize billions of parameters on a GPU cluster over days, weeks, or even months, patiently waiting for a loss value to descend. These systems may learn *about* time but are ultimately exempt from the consequences of living through it.

In contrast, people are inescapably bound to time. Under strict temporal constraints, we walk, speak, collaborate, play, recover from mistakes, develop new skills, and adapt to unseen circumstances. The world does not conveniently pause while we think, reset when we fail, or preserve our past experiences as lossless samples. Every decision consumes time and every failure leaves both the world and learner changed. Our memories are partial, opportunities in the world expire, and who we become strongly depends on the individual path we have taken over time. Yet, as suggested above, the agents we build to assist us are often trained and evaluated under temporal affordances that are incompatible with the environments they are meant to provide support. This mismatch creates what we consider as a *temporality gap*. We argue that machine learning-based agents may not realize their full potential to help and participate in our world until they start learning, acting, remembering, and adapting within a shared timeline with people.

Synthesizing our points above, our central position through this paper is that machine learning's engagement with time remains *incomplete*. Much of modern machine learning studies time by first converting it into a static object: a dataset, replay buffer, context window, trajectory, episode, benchmark split, or computational graph. We refer to this tendency as the *spatialization of time*. Time is spatialized when events that originally unfolded one after another are represented as positions in an indexable structure, allowing the learner to inspect, batch, reorder, revisit, or compare

them as if they were points laid out in space. This transformation is often useful, and in many cases necessary for scalable optimization. However, it also removes the pressures that make learning *through* time different from learning *about* time.

We argue that progress toward persistent artificial agents requires treating temporality not merely as a feature of the input, but as *a set of constraints on the environment and learner*. This paper makes those constraints explicit through six guidelines for temporally grounded learning: autonomous external time, partial embodied observability, non-resettable lifetime interaction, bounded reconstructive memory, multi-rate endogenous computation, and path-dependent self-modification. These guidelines are not intended to specify a particular architecture or algorithm; instead, they define the conditions under which a learning system can be said to experience, act, remember, and adapt through time.

The rest of the paper develops this framework and uses it to organize a research agenda. In § 2, we define and motivate each guideline by identifying the machine learning affordance it removes, the evaluation considerations it introduces, and the natural precedent that supports it. In § 3, we use the guidelines to compare machine learning paradigms, showing that existing work captures important fragments of temporal grounding but does not yet treat all six constraints as central and simultaneous. In § 4, we translate this perspective into concrete calls to action. We call for benchmark ecosystems that make temporal grounding measurable; agent architectures that explicitly audit their assumptions about time; shared venues where time is treated as a first-class object of study; and review norms that reward work confronting temporal constraints, even when doing so makes learning harder. Lastly, we consider alternative views in § 5 and conclude in § 6.

Our central claim is the following: existing machine learning has become increasingly effective at modeling temporal structure, but **modeling time is not the same as learning through time**. The field has built powerful systems for representing sequences, predicting futures, and optimizing trajectories. The next frontier is building agents whose learning itself is embedded in the same irreversible, resource-limited, history-shaping time as the worlds they inhabit.

2. Guidelines for Temporally Grounded Learning

In this section, we propose six guidelines for temporally grounded learning. These guidelines do not define a specific algorithm, architecture, or benchmark. Rather, they specify conditions under which a learning system can be said to experience, act, remember, and adapt through real clock time.

We do not claim that these six guidelines are the only possible decomposition of temporally grounded learning. Instead, they are a principled starting point: a set of pressures that distinguish learning *through* time from learning *about* temporal data. We select each of them based on their removal of common artificial affordances in machine learning, introduction of concrete evaluation considerations, and reflection of natural precedents in biological organisms.

2.1. Guideline 1: Autonomous External Time

Description. A temporally grounded learner must operate in a world that evolves according to its own clock and dynamics. Let $\tau \in [0, \infty)$ denote external time, and let $\mathcal{W}(\tau)$ denote the world state. We assume that τ is strictly monotonic and that $\mathcal{W}(\tau)$ evolves according to dynamics not synchronized with the agent's computational cycles. The agent may observe, act, deliberate, and update itself, but it cannot pause external time while it computes.

Removed ML Affordance. This guideline removes the assumption that the environment waits for the learner. In many reinforcement learning systems, interaction is mediated by a synchronous loop such as `env.step()`, where the world advances only after the agent has computed an action. Temporally grounded learning instead treats computation, sensing, memory retrieval, and learning as processes that consume world-time.

Evaluation Considerations. Latency becomes part of the learning problem. A decision that is optimal under zero-cost computation may be obsolete by the time it is executed. Agents should therefore be evaluated not only by decision quality, but also by whether decisions arrive while they are still relevant. Benchmarks should measure inference time, update time, sensing delay, and the opportunity cost of deliberation.

Natural Basis. Natural agents already face this constraint: the world does not wait for organisms to perceive, decide, or act. Across animals, perceptual processing speed is coupled to ecological tempo, with faster visual perception in species facing fast-paced ecological demands (Haarlem et al., 2026). Processing latency also varies across species and modalities (Itoh et al., 2022), while perception, action, and cognition operate over distinct temporal scales within a single nervous system (Hari & Parkkonen, 2015; Tovée, 1994). These observations suggest that real-time agency requires matching computation and action to the timescale of the world.

The Six Guidelines for Temporally Grounded Learning

1. Autonomous External Time. The world evolves according to its own clock and dynamics, independent of the agent's computational cycles.

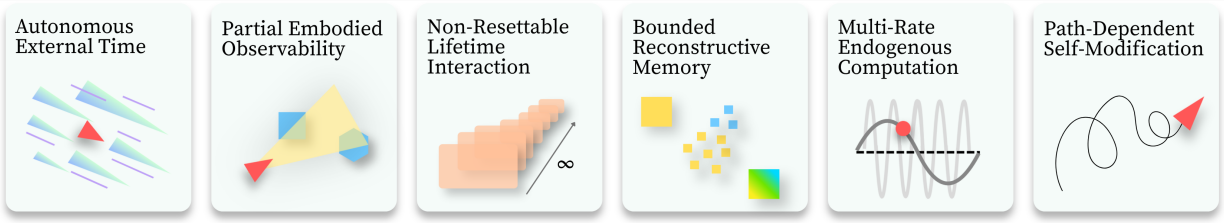
2. Partial Embodied Observability. The agent receives only a narrow and non-invertible stream of observations from the much larger world state.

3. Non-resettable Lifetime Interaction. The agent cannot rewind, reset, or restart the world, and must instead learn over a single irreversible lifetime.

4. Bounded Reconstructive Memory. The agent's memory is capacity-limited, constructive, and modified by recall, rather than a static archive of perfectly replayable past samples.

5. Multi-rate Endogenous Computation. The agent must support internal processes that unfold at different rates, including fast reactions, slower deliberation, and long-term consolidation.

6. Path-dependent Self-Modification. The agent's internal structure must be persistently shaped by its own history, so that learning is cumulative, hysteretic, and identity-forming.



2.2. Guideline 2: Partial Embodied Observability

Description. A temporally grounded learner must operate through a limited stream of embodied observations, not through direct access to the full world state. Let $o(\tau) \in \mathcal{O}$ denote the agent's observation at time τ, generated from $\mathcal{W}(\tau)$ by a sensory process. We assume this observation is partial, local, and non-invertible: many distinct world states can produce the same observation.

Removed ML Affordance. This guideline removes the assumption that the relevant state is fully available. In many supervised, reinforcement learning, and planning settings, agents receive complete or nearly complete state descriptions. Temporally embedded agents instead acquire data through situated, bandwidth-limited, and perspective-dependent sensing.

Evaluation Considerations. The present observation is rarely sufficient on its own. The agent must use time to resolve ambiguity by integrating observations, actions, memories, and predictions. Agents should therefore be evaluated on their ability to maintain useful internal state, actively gather information, fuse observations over time, and act under uncertainty when the true world state is unavailable.

Natural Basis. Natural agents construct useful representations from partial, local, and task-dependent sensory streams (Gallistel, 1990). Desert ants navigate using path integration and homing despite local sensory access (Wehner & Srinivasan, 1981); migratory birds combine environmental signals to update navigational representations (Demšar et al., 2025); and place and grid cells support spatial representation built from experience rather than direct access to a global world state (Moser et al., 2008; Poucet et al., 2004). Partial observability is therefore not an edge case, but a central pressure faced by embodied learners.

2.3. Guideline 3: Non-resettable Lifetime Interaction

Description. A temporally grounded learner must operate over a single irreversible lifetime, not a collection of independent episodes. Because external time is strictly monotonic, the agent cannot rewind the world, erase the consequences of its actions, or restart from a clean initial condition. Every observation, action, failure, recovery, and update becomes part of the agent's continuing trajectory.

Removed ML Affordance. This guideline removes the assumption that experience can be reset. In episodic reinforcement learning, failure is often handled by calling `env.reset()` and returning the agent and environment to a predefined start state. This eliminates the persistent consequences that define real lifetime interaction, allowing agents to discard failures rather than avoid, repair, or adapt to them.

Evaluation Considerations. Mistakes cannot simply disappear at the end of an episode. Agents should be evaluated by how they handle damage, missed opportunities, exploration risk, resource depletion, changing circumstances, and long-term consequences. Benchmarks should measure viability, recovery, accumulated cost, irreversible state changes, and adaptation across a continuing lifetime rather than only per-episode success.

Natural Basis. Natural agents do not receive external resets after failure. Animals develop risk-sensitive exploration when mistakes may lead to injury, predation, or resource loss (Shen & Dayan, 2024), and risk aversion has been linked to neural decision mechanisms and evolutionary advantage (Hintze et al., 2015; Trepel et al., 2005; Zhang et al., 2014). When organisms suffer damage, they often adapt behavior rather than return to baseline (David Smith & Hines, 1991; Rennolds & Bely, 2023). Irreversibility therefore shapes exploration, caution, recovery, and adaptation.

2.4. Guideline 4: Bounded Reconstructive Memory

Description. A temporally grounded learner must remember through bounded and reconstructive memory, not through unlimited access to perfectly preserved past samples. Let $m(\tau) \in \mathcal{M}$ denote the agent's memory state. We assume that $\mathcal{M}$ is capacity-limited, that memory compresses the agent's growing stream of experience, and that recall can influence and modify memory itself. The past is therefore reconstructed from the present, not retrieved from a lossless archive.

Removed ML Affordance. This guideline removes the assumption that the past can be replayed exactly. In many learning systems, past data are stored in static datasets, replay buffers, logs, or context windows that can be revisited without alteration. This abstraction is useful, but it grants the agent a form of time travel: the ability to repeatedly access identical past samples as if they were independent, unchanged objects.

Evaluation Considerations. Remembering becomes part of learning rather than a passive read operation. Agents should be evaluated by what they preserve, compress, forget, revise, and consolidate over time. Benchmarks should consider memory capacity, degradation, interference, recall-induced modification, and the ability to maintain useful abstractions without lossless replay.

Natural Basis. Biological memory is not a static data store. Human retrieval can incorporate new information, reshape details, and strengthen linked memories (Hupbach et al., 2007; Jonker et al., 2013; 2018). Memory is organized around temporally extended events (DuBrow, 2024; Liverence & Scholl, 2012; Ongchoco et al., 2023; Zacks et al., 2007), and long-term memory depends on consolidation across hippocampal and neocortical systems (McClelland et al., 1995). Neural replay exists, but differs from machine learning replay buffers: it is selective, structured, and tied to consolidation rather than exact resampling (Hayes et al., 2021; Pavlides & Winson, 1989).

2.5. Guideline 5: Multi-rate Endogenous Computation

Description. A temporally grounded learner must support internal processes that unfold at different rates, rather than forcing all computation, memory, and learning to advance on one global clock. Let τ_i denote the internal time of subsystem i. Different subsystems may evolve with different effective temporal velocities, allowing fast reactions, slower deliberation, and long-term consolidation to coexist. These multi-rate processes may also enable the agent to form an implicit representation of time used for temporal cognition.

Removed ML Affordance. This guideline removes the assumption that the learner updates uniformly at each timestep. Many sequence models, recurrent architectures, and reinforcement learning agents advance through a single synchronized loop. Temporally grounded intelligence may instead require asynchronous processes operating across multiple timescales, with some systems coupled tightly to the sensory stream and others integrating or consolidating more slowly.

Evaluation Considerations. Agents must be designed, not only with the content of their computations in mind, but also when and at what rate these computations occur. Agents should be evaluated on their ability to coordinate fast reaction, delayed reasoning, memory consolidation, and slow structural adaptation within the same ongoing lifetime. Benchmarks should test events and decisions that unfold across multiple temporal scales.

Natural Basis. Biological organisms contain many interacting temporal processes rather than a single uniform clock. Endogenous timekeeping and biological oscillators coordinate behavior across scales (Gallistel, 1990; Kaufmann, 2026; Kronfeld-Schor et al., 2013). Circadian rhythms are tied to Earth's day-night cycle (Czeisler & Gooley, 2007; Pitsawong et al., 2023), and brain oscillatory hierarchies are preserved across vertebrates despite large differences in brain size (Buzsáki et al., 2013). These rhythms support temporally organized behavior in bees, hummingbirds, and chimpanzees (Boesch, 2002; Janmaat, 2019; Moore et al., 1989). These examples suggest that temporally grounded agents should not merely process observations in order; they should maintain internal temporal structure that supports behavior, memory, and adaptation across multiple timescales.

2.6. Guideline 6: Path-dependent Self-Modification

Description. A temporally grounded learner must be persistently shaped by its own history rather than returning to a fixed baseline after each task, episode, or interaction. Let $\theta(\tau)$ denote the agent's internal structure, including parameters, memory organization, learned skills, plasticity mechanisms, and other persistent variables. We assume $\theta(\tau)$ depends on the particular trajectory of experiences, actions, and updates that preceded it. In this sense, the same present observation may be interpreted differently by agents with different histories.

Removed ML Affordance. This guideline removes the assumption that the relevant past can be artificially bounded, discarded, or made structurally irrelevant. In many systems, history enters only through a fixed context window, finite recurrent state, truncated rollout, or episode-specific adaptation mechanism. Temporally grounded learning is cumulative: early experiences, rare failures, or highly salient events can permanently alter how future experience is interpreted, what can be learned next, and how the agent responds.

Evaluation Considerations. History is not merely stored; it alters the learner. Agents should be evaluated by how their internal structure evolves across a non-repeatable lifetime. Benchmarks should track learning trajectories, plasticity, stability, recovery, and long-term behavioral change rather than treating the final model state as the only object of interest. They should also test whether early experiences or salient events have durable effects on later learning, for better or worse.

Natural Basis. Natural agents are path-dependent systems. Developmental biology provides a clear example: the same genotype can produce different phenotypes depending on environmental conditions (Lafuente & Beldade, 2019), and early critical periods can make certain windows of experience disproportionately important for later perception, behavior, and learning (Hensch, 2005; Knudsen, 2004). Path dependence can also be maladaptive: traumatic or highly stressful experiences can leave durable traces that shape future threat perception, memory, affect, and behavior (Teicher et al., 2016). More broadly, animals display behavioral plasticity that supports survival in changing environments (Botero, 2026; Garcia-Porta et al., 2022; Wong & Candolin, 2015). At the neural level, Hebbian learning, synaptic plasticity, neurogenesis, and cell migration provide structural substrates for cumulative change (Gazerani, 2025; Hebb, 1949; von Bernhardi et al., 2017). These observations support the view that temporally grounded learners should not merely condition on history; they should be structurally shaped by it.

3. Current ML Paradigms Through the Six Guidelines

In this section, we use the six guidelines from § 2 to compare several major machine learning paradigms. Our goal is not to rank individual works, but to highlight a recurring pattern: existing paradigms often capture important fragments of temporal grounding while relaxing, externalizing, or assuming away other temporal pressures.

Table 1 provides a coarse summary, and the paragraphs below briefly justify each row. More detailed discussions of reinforcement learning and foundation models are provided in Appendix A and Appendix B, respectively.

Transformer-Based Foundation Models. Foundation models (Deitke et al., 2024; Team et al., 2024; Zitkovich et al., 2023) provide powerful machinery for representing temporal structure through token sequences (Dosovitskiy et al., 2021; Sennrich et al., 2016; Zhao et al., 2023), positional encodings (Chen et al., 2021; Vaswani et al., 2017), long context windows (Mei et al., 2025), and large-scale pretraining (Radford & Narasimhan, 2018). However, they typically spatialize experience into static corpora and context windows rather than learning through ongoing interaction with a world that evolves during inference. They partially address observability through scale, broad pretraining, and transient in-context adaptation (Dong et al., 2024), but they generally lack autonomous external time, non-resettable lifetime interaction, bounded reconstructive memory, multi-rate endogenous computation, and persistent self-modification.

Episodic Reinforcement Learning. Episodic reinforcement learning (Lillicrap et al., 2015; Mnih et al., 2015) is explicitly built around interaction, but it usually grants the agent several strong temporal affordances. The environment (Towers et al., 2026) often advances only through a synchronous call such as `env.step()`, while failure is erased through `env.reset()`. Replay buffers (Andrychowicz et al., 2017; Mnih et al., 2015; Schaul et al., 2016) further convert experience into a static archive of reusable transitions. Thus, episodic reinforcement learning captures sequential decision-making while removing the pressures of real-time latency, irreversible consequence, bounded reconstructive memory, and lifetime self-modification.

POMDP / Recurrent Reinforcement Learning. POMDPs (Lauri et al., 2022; Åström, 1965) and recurrent reinforcement learning (Anjarlekar et al., 2026; Hausknecht & Stone, 2015) directly address partial observability by recognizing that agents receive observations rather than full world states, and by maintaining histories, belief states, or recurrent hidden states. This makes them an important approximation of embodied sensing. However, most remain embedded within standard reinforcement learning loops:

environments are stepped synchronously, episodes are reset, and memory is often treated as a learned hidden vector or replayed trajectory rather than as bounded reconstructive memory. They therefore satisfy Guideline 2 most directly while only partially addressing the remaining guidelines.

Online Meta-RL. Online meta-RL (Beck et al., 2023; Nam et al., 2022; Kobayashi et al., 2025) studies agents that adapt within a task or episode, and therefore partially engages path-dependent change. However, this adaptation usually occurs inside a broader training procedure built from resettable tasks, fixed task distributions, and externally defined episode boundaries. The inner loop may simulate short-horizon adaptation, but the outer loop is typically optimized offline. Meta-RL is therefore valuable for studying fast adaptation, but it does not by itself provide autonomous external time, non-resettable lifetime interaction, bounded reconstructive memory, or multi-rate endogenous computation.

Continual Learning. Continual learning (Abel et al., 2023; Wang et al., 2024a) is one of the clearest engagements with path-dependent self-modification: it studies how a model changes over time while preserving prior capabilities (Dohare et al., 2024; Kirkpatrick et al., 2017; Luo et al., 2025). However, much of the field still operates on static or presegmented datasets, often with task boundaries supplied by the experimenter. The learner changes, but the world usually does not evolve independently; observations are often fully specified; and evaluation may still be organized around fixed tasks (Hu et al., 2026; Wolczyk et al., 2021). Continual learning therefore captures important aspects of Guideline 6, but usually only partially addresses irreversibility, bounded memory, and embodied temporal interaction.

Streaming / Real-Time RL. Streaming and real-time reinforcement learning move closer to temporal grounding by reducing reliance on large offline batches and emphasizing continuous data streams (Elsayed et al., 2024; Vasan et al., 2024) or real-time interaction (Ramstedt & Pal, 2019; Bouteiller et al., 2021). These methods partially address autonomous external time and non-resettable interaction, especially when the data stream does not wait for repeated optimization over stored samples. However, many still rely on simplified clocks, engineered observation streams, action buffers, limited memory mechanisms, or uniform update rates. They are promising steps toward temporal grounding, but generally do not yet provide bounded reconstructive memory, multi-rate endogenous computation, or robust path-dependent self-modification.

State-Space Models / Continuous-Time RNNs. State-space models (Gu & Dao, 2024), recurrent networks (Hochreiter & Schmidhuber, 1997), Neural ODEs (Chen et al., 2018), continuous-time RNNs (Rubanova et al., 2019), liquid neural networks (Hasani et al., 2020), and related architectures (Schlag et al., 2021) provide useful mechanisms for maintaining hidden state and modeling continuous dynamics. They can compress history, represent temporal dependencies, and in some cases evolve according to continuous-time equations. However, these architectures are often trained offline on fixed datasets using backpropagation through time or related procedures that turn trajectories into static computational graphs. They offer architectural ingredients for temporal grounding, but architectural temporality alone does not guarantee non-resettable interaction, autonomous world-time, or lifetime self-modification.

Latent World Models / Model-Based RL. Latent world models (Ermolov & Sebe, 2020) and model-based reinforcement learning (Hafner et al., 2025; Janner et al., 2019; Moerland et al., 2023) explicitly address the need to infer hidden state and predict future consequences from partial observations. They therefore strongly engage partial observability and provide machinery for planning under uncertainty. However, these systems are often trained with replay buffers, episodic simulators, and resettable rollouts. Their imagined futures are useful, but they are usually generated outside the constraints of irreversible world-time. Thus, world models help agents reason about time, but do not necessarily make the learning process itself temporally grounded.

Neuromorphic / Spiking systems. Neuromorphic computing (Schuman et al., 2017) and spiking neural networks (Tavanaei et al., 2019; Zanatta et al., 2024) are well aligned with temporal dynamics at the computational level. Spikes, leaks, decays, and hardware-level temporal evolution make these systems natural candidates for autonomous time and bounded memory. However, many neuromorphic systems are still evaluated on static datasets, reset between examples, or trained using offline objectives that reintroduce spatialized time. Their physical substrate may be temporally grounded, but the surrounding learning setup often is not. In particular, most such systems do not remove resets as an evaluation affordance or support persistent path-dependent self-modification across a lifetime.

Artificial Life / Open-Ended Learning. Artificial life (Bhattacharjee et al., 2020; Kumar et al., 2024) and open-ended learning (Team et al., 2021) often come closest to treating agents as persistent entities in ongoing worlds. These systems can emphasize survival, adaptation, ecological interaction, partial observability, non-stationarity, and path-dependent change. However, many still rely on simulator clocks, simplified sensors, limited memory systems, externally defined evolutionary or learning procedures, and global update schedules. In particular, bounded reconstructive memory and multi-rate endogenous computation are

Table 1. A coarse diagnostic of how representative machine learning paradigms engage with the six guidelines for temporally grounded learning. Symbols indicate substantial engagement (✓), partial engagement or simulation (♦), or typical violation / assumption-away (✗).

G1: Autonomous external time **G2**: Partial embodied observability **G3**: Non-resettable lifetime interaction
G4: Bounded reconstructive memory **G5**: Multi-rate endogenous computation **G6**: Path-dependent self-modification

Paradigm	G1	G2	G3	G4	G5	G6
Transformer-Based Foundation Models (e.g., (Deitke et al., 2024; Zitkovich et al., 2023))	✗	♦	✗	✗	✗	♦
Episodic RL (e.g., (Lillicrap et al., 2015; Mnih et al., 2015))	✗	♦	✗	✗	✗	♦
POMDP / Recurrent RL (e.g., (Hausknecht & Stone, 2015; Lauri et al., 2022))	✗	✓	✗	♦	✗	♦
Online Meta-RL (e.g., (Beck et al., 2023; Nam et al., 2022))	✗	♦	✗	✗	✗	♦
Continual Learning (e.g., (Abel et al., 2023; Wang et al., 2024a))	✗	♦	♦	♦	✗	✓
Streaming / Real-Time RL (e.g., (Elsayed et al., 2024; Ramstedt & Pal, 2019))	♦	♦	♦	♦	✗	♦
State-Space Models / Continuous-Time RNNs (e.g., (Chen et al., 2018; Gu & Dao, 2024))	♦	♦	✗	♦	♦	♦
Latent World Models / Model-Based RL (e.g., (Hafner et al., 2025; Moerland et al., 2023))	✗	✓	✗	♦	♦	♦
Neuromorphic / Spiking Systems (e.g., (Schuman et al., 2017; Tavanaei et al., 2019))	✓	♦	✗	✓	♦	✗
Artificial Life / Open-Ended Learning (e.g., (Kumar et al., 2024; Team et al., 2021))	♦	✓	✓	✗	✗	♦

rarely central objects of study. Artificial life demonstrates why temporal grounding matters, but often leaves unresolved how individual agents should remember, deliberate, consolidate, and self-modify through a lifetime.

Summary. The pattern in Table 1 is not that current paradigms fail to study time. Rather, they each isolate different fragments of temporality while relaxing or ignoring others. This fragmentation motivates the central call of the paper: the field should move from isolated temporal modules toward agents that jointly satisfy the six proposed temporal pressures.

4. Calls to Action

The six guidelines in § 2 are intended to be actionable. If temporally grounded learning is to become a fruitful research direction, the community needs new benchmarks, agent architectures, venues, and review norms that make time a principal object of study.

Build temporally grounded benchmark ecosystems. The most important next step is to build public benchmark ecosystems that make the six guidelines measurable. These benchmarks should impose the external pressures of temporally grounded learning: worlds that continue to evolve while agents compute, observations that are partial and embodied, and interactions that unfold over non-resettable lifetimes. At the same time, they should include tasks that specifically probe the agent-side capacities required to survive those pressures: bounded memory, multi-rate computation, and path-dependent self-modification.

Concretely, benchmarks should report performance along each temporal dimension. A useful benchmark should reveal whether an agent succeeds by genuinely handling temporal pressure or by exploiting resets, perfect state access, paused computation, lossless memory, or simplified observability. Public leaderboards can give the community shared targets, make progress measurable rather than rhetorical, and track which temporal pressures remain unsolved over time.

Audit the temporal assumptions inside agent architectures. Researchers should make the temporal assumptions of their agents explicit. Many standard design choices quietly spatialize time: the Markov assumption treats the present state as sufficient; fixed context windows bound history by construction; replay buffers turn past experience into an indexable database; synchronous update loops force all internal processes onto one clock; and episodic adaptation mechanisms allow the agent to return to a clean baseline. These assumptions are often useful, but they should be treated as modeling choices rather than defaults.

We therefore call on researchers to introspect and ask, for every proposed agent architecture: where is time actually unfolding in this system? Is the agent learning through an irreversible stream of experience, or optimizing over a spatialized record of that experience? Does memory actively change through recall, or is it a passive lookup table? Are internal processes allowed to unfold at different rates, or is the entire system advanced by one global tick? Does the agent's history structurally shape future learning, or is history merely placed in a buffer? Making these assumptions explicit is a necessary step toward architectures that genuinely learn through time.

Create shared venues around time. Major machine learning conferences should host workshops, tutorials, competitions, and discussion tracks where time is treated as a first-class citizen. Temporally grounded learning cuts across reinforcement learning, continual learning, foundation models, robotics, neuromorphic computing, artificial life, cognitive science, and neuroscience. Dedicated venues can

help establish shared terminology, benchmark standards, evaluation protocols, and open problems. They can also help prevent temporally grounded learning from being fragmented across communities that currently study different pieces of the problem in isolation.

Reward temporally grounded work during review. The community may also need to adjust its review norms. Removing temporal affordances makes learning harder. Systems trained under autonomous world-time, partial observability, non-resettable interaction, bounded memory, multi-rate computation, and path-dependent self-modification may initially perform worse than systems trained with static datasets, resettable episodes, perfect replay, and unlimited offline computation. This should not automatically be treated as failure.

Reviewers should value work that exposes, formalizes, and studies these difficulties, even when short-term performance is lower than highly optimized baselines. The field is already highly capable under spatialized-time assumptions. Temporally grounded learning asks us to leave that local optimum. Progress may require accepting harder benchmarks, messier experiments, and temporary regressions in headline capability in exchange for systems that are ultimately better suited to persistent agency.

5. Alternative Views

We briefly address several possible objections to, and interpretations of, our framing. Our goal is not to settle these questions definitively, but to clarify the scope of our claims, distinguish temporally grounded learning from nearby ideas, and spark further discussion.

"Don't machine learning models already engage with time?" Machine learning has indeed made substantial progress on problems involving time. Sequence models process text, audio, and video; reinforcement learning agents optimize behavior over trajectories; continual learning studies non-stationary streams; recurrent networks and state-space models maintain hidden dynamics; world models simulate future rollouts; and foundation models increasingly operate over long context windows. In this sense, the field has certainly not ignored time. Temporal structures are abundant within the modern machine learning literature. Our argument, however, is that there exists a significant gap between simply modeling temporal structures versus operating under the constraints of continuous real time.

"Are the guidelines too prescriptive on the agent side?" Guidelines 4–6 do prescribe properties of the agent rather than only properties of the task or environment. This choice is intentional. Many temporal affordances enter through agent design. For instance, replay buffers, context windows, and externally managed memory all shape how the learner relates to time. If temporally grounded learning is about how agents experience, remember, compute, and change through time, then agent-side assumptions must be part of the framework.

"Are the guidelines unique or exhaustive?" No. We do not claim that these six guidelines are the only possible decomposition of temporally grounded learning, nor that they are necessary or sufficient conditions for intelligence or agency. They are a diagnostic starting point: a way to make common temporal affordances explicit and to define a concrete research agenda. Other decompositions are possible. Our claim is narrower: systems that satisfy these guidelines would remove many of the artificial temporal conveniences that current learning paradigms often rely on, and would therefore provide a useful target for studying learning through time.

"Can external tools satisfy the endogenous guidelines?" Yes, in principle. A temporally grounded agent may use external clocks, notebooks, databases, tools, or other artifacts. The key distinction is not whether a component is physically internal or external, but whether it is integrated into the agent's temporal life. External mechanisms are compatible with temporal grounding when they are temporally costly to access, bounded, dynamically updated, and coupled to the agent's ongoing activity. They become problematic when they function as lossless archives, free computation, or reset mechanisms that let the agent stand outside time.

"Does biology overconstrain the agenda?" No. We do not argue that artificial agents must copy biological mechanisms. We use biological and cognitive evidence for a narrower purpose: to show that the proposed temporal pressures are real, consequential, and compatible with robust learning. Natural agents learn under autonomous external time, partial observability, irreversible lifetimes, bounded memory, multi-rate internal processes, and path-dependent change. We acknowledge that artificial agents may satisfy these pressures in vastly different ways. The point is not to imitate biology, but to stop ignoring constraints that biological learners never get to escape.

6. Conclusion

This position paper has argued that machine learning's engagement with time remains incomplete. Rather than treating temporality only as structure in data, we proposed six guidelines that recast time as a set of constraints on the environment and learner: autonomous world-time, partial observability, irreversible interaction, bounded memory, multi-rate computation, and path-dependent change. We used

these guidelines to diagnose existing paradigms, clarify where they capture fragments of temporal grounding, and identify concrete actions for the community.

The next step is not simply to build longer-context models, longer-horizon tasks, or larger replay buffers. It is to build benchmarks, architectures, and evaluation norms that force agents to learn while time continues to pass. If artificial agents are to help people act in a world that never pauses, then they must learn within that unfolding world, not merely from records of what has already happened.

References

Abel, D., Barreto, A., Roy, B. V., Precup, D., van Hasselt, H., and Singh, S. A definition of continual reinforcement learning. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=ZZS9WEWYbD.

Abel, D., Ho, M. K., and Harutyunyan, A. Three dogmas of reinforcement learning. *Reinforcement Learning Journal*, 2:629–644, 2024.

Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., McGrew, B., Tobin, J., Pieter Abbeel, O., and Zaremba, W. Hindsight Experience Replay. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Andrychowicz, O. M., Baker, B., Chociej, M., Józefowicz, R., McGrew, B., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., Schneider, J., Sidor, S., Tobin, J., Welinder, P., Weng, L., and Zaremba, W. Learning dexterous in-hand manipulation. *Int. J. Rob. Res.*, 39(1):3–20, January 2020. ISSN 0278-3649. doi: 10.1177/0278364919887447. URL https://doi.org/10.1177/0278364919887447.

Anjarlekar, A., Etesami, S. R., and Srikant, R. Scalable policy-based RL algorithms for POMDPs. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=JInukYoZPI.

Bacon, P.-L., Harb, J., and Precup, D. The option-critic architecture. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, pp. 1726–1734. AAAI Press, 2017.

Baek, C., Monti, R. P., Schwab, D., Abbas, A., Adiga, R., Blakeney, C., Böther, M., Burstein, P., Carranza, A. G., Deng, A., Doshi, P., Dorna, V., Fang, A., Jiang, T., Joshi, S., Larsen, B. W., Lee, J. C., Mentzer, K. L., Merrick, L., Mongstad, H., Pan, F., Suri, A., Teh, D., Telanoff, J., Urbanek, J., Wang, Z., Wills, J., Yin, H., Raghunathan, A., Kolter, J. Z., Gaza, B., Morcos, A., Leavitt, M., and Maini, P. The finetuner's fallacy: When to pretrain with your finetuning data, 2026. URL https://arxiv.org/abs/2603.16177.

Beck, J., Vuorio, R., Liu, E. Z., Xiong, Z., Zintgraf, L. M., Finn, C., and Whiteson, S. A tutorial on meta-reinforcement learning. *Found. Trends Mach. Learn.*, 18:224–384, 2023. URL https://api.semanticscholar.org/CorpusID:255999767.

Bhattacharjee, K., Naskar, N., Roy, S., and Das, S. A survey of cellular automata: types, dynamics, non-uniformity and applications. *Natural Computing*, 19(2): 433–461, June 2020. ISSN 1572-9796. doi: 10.1007/s11047-018-9696-8. URL https://doi.org/10.1007/s11047-018-9696-8.

Black, K., Brown, N., Darpinian, J., Dhabalia, K., Driess, D., Esmail, A., Equi, M. R., Finn, C., Fusai, N., Galliker, M. Y., Ghosh, D., Groom, L., Hausman, K., brian ichter, Jakubczak, S., Jones, T., Ke, L., LeBlanc, D., Levine, S., Li-Bell, A., Mothukuri, M., Nair, S., Pertsch, K., Ren, A. Z., Shi, L. X., Smith, L., Springenberg, J. T., Stachowicz, K., Tanner, J., Vuong, Q., Walke, H., Walling, A., Wang, H., Yu, L., and Zhilinsky, U. $\pi_{0.5}$: a vision-language-action model with open-world generalization. In *9th Annual Conference on Robot Learning*, 2025. URL https://openreview.net/forum?id=vlhoswksBO.

Black, K., Galliker, M. Y., and Levine, S. Real-time execution of action chunking flow policies. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=UkR2zO5uww.

Boesch, C. Cooperative hunting roles among taï chimpanzees. *Human Nature*, 13(1):27–46, March 2002. ISSN 1045-6767. doi: 10.1007/s12110-002-1013-6.

Botero, C. A. The evolutionary consequences of behavioural plasticity. *Nature Communications*, 17(1): 3880, March 2026. ISSN 2041-1723. doi: 10.1038/s41467-026-70632-8. URL https://www.nature.com/articles/s41467-026-70632-8.

Bouteiller, Y., Ramstedt, S., Beltrame, G., Pal, C., and Binas, J. Reinforcement learning with random delays. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=QFYnKlBJYR.

Bradtke, S. and Duff, M. Reinforcement Learning Methods for Continuous-Time Markov Decision Problems. In Tesauro, G., Touretzky, D., and

Leen, T. (eds.), *Advances in Neural Information Processing Systems*, volume 7. MIT Press, 1994. URL https://proceedings.neurips.cc/paper_files/paper/1994/file/07871915a8107172b3b5dc15a6574ad3-Paper.pdf.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Buzsáki, G., Logothetis, N., and Singer, W. Scaling Brain Size, Keeping Timing: Evolutionary Preservation of Brain Rhythms. *Neuron*, 80(3):751–764, October 2013. ISSN 0896-6273. doi: 10.1016/j.neuron.2013.10.002. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4009705/.

Chen, A., Sharma, A., Levine, S., and Finn, C. You only live once: Single-life reinforcement learning. *Advances in Neural Information Processing Systems*, 35:14784–14797, 2022.

Chen, P.-C., Tsai, H., Bhojanapalli, S., Chung, H. W., Chang, Y.-W., and Ferng, C.-S. A simple and effective positional encoding for transformers. In Moens, M.-F., Huang, X., Specia, L., and Yih, S. W.-t. (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 2974–2988, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.236. URL https://aclanthology.org/2021.emnlp-main.236/.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/69386f6bb1dfed68692a24c8686939b9-Abstract.html.

Czeisler, C. A. and Gooley, J. J. Sleep and circadian rhythms in humans. *Cold Spring Harbor Symposia on Quantitative Biology*, 72:579–597, 2007. ISSN 0091-7451. doi: 10.1101/sqb.2007.72.064.

David Smith, L. and Hines, A. H. The effect of cheliped loss on blue crab *Callinectes sapidus* Rathbun foraging rate on soft-shell clams *Mya arenaria* L. *Journal of Experimental Marine Biology and Ecology*, 151(2):245–256, October 1991. ISSN 0022-0981. doi: 10.1016/0022-0981(91)90127-I. URL https://www.sciencedirect.com/science/article/pii/002209819190127I.

Deitke, M., Clark, C., Lee, S., Tripathi, R., Yang, Y., Park, J. S., Salehi, M., Muennighoff, N., Lo, K., Soldaini, L., Lu, J., Anderson, T., Bransom, E., Ehsani, K., Ngo, H., Chen, Y., Patel, A., Yatskar, M., Callison-Burch, C., Head, A., Hendrix, R., Bastani, F., VanderBilt, E., Lambert, N., Chou, Y., Chheda, A., Sparks, J., Skjonsberg, S., Schmitz, M., Sarnat, A., Bischoff, B., Walsh, P., Newell, C., Wolters, P., Gupta, T., Zeng, K.-H., Borchardt, J., Groeneveld, D., Dumas, J., Nam, C., Lebrecht, S., Wittlif, C., Schoenick, C., Michel, O., Krishna, R., Weihs, L., Smith, N. A., Hajishirzi, H., Girshick, R., Farhadi, A., and Kembhavi, A. Molmo and pixmo: Open weights and open data for state-of-the-art multimodal models. *arXiv preprint arXiv:2409.17146*, 2024.

Demšar, U., Zein, B., and Long, J. A. A new data-driven paradigm for the study of avian migratory navigation. *Movement Ecology*, 13:16, March 2025. ISSN 2051-3933. doi: 10.1186/s40462-025-00543-8. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC11900352/.

Dohare, S., Hernandez-Garcia, J. F., Lan, Q., Rahman, P., Mahmood, A. R., and Sutton, R. S. Loss of plasticity in deep continual learning. *Nature*, 632(8026):768–774, August 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07711-7.

Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Chang, B., Sun, X., Li, L., and Sui, Z. A survey on in-context learning. In Al-Onaizan, Y., Bansal, M., and Chen, Y.-N. (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 1107–1128, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.64. URL https://aclanthology.org/2024.emnlp-main.64/.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021.

Du, J., Futoma, J., and Doshi-Velez, F. Model-based reinforcement learning for semi-markov decision processes with neural odes. In *Proceedings of the 34th International*

Conference on Neural Information Processing Systems, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Du, Y., Tian, M., Ronanki, S., Rongali, S., Bodapati, S., Galstyan, A., Wells, A., Schwartz, R., Huerta, E. A., and Peng, H. Context Length Alone Hurts LLM Performance Despite Perfect Retrieval. 2025. doi: 10.48550/ARXIV.2510.05381. URL https://arxiv.org/abs/2510.05381. Version Number: 1.

DuBrow, S. Events and Boundaries. In Kahana, M. J. and Wagner, A. D. (eds.), *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*, pp. 0. Oxford University Press, June 2024. ISBN 978-0-19-091798-2. doi: 10.1093/oxfordhb/9780190917982.013.18. URL https://doi.org/10.1093/oxfordhb/9780190917982.013.18.

Earle, S., Todd, G., Li, Y., Khalifa, A., Nasir, M. U., Jiang, Z., Banburski-Fahey, A., and Togelius, J. Puzzlejax: A benchmark for reasoning and learning, 2025. URL https://arxiv.org/abs/2508.16821.

Elsayed, M., Vasan, G., and Mahmood, A. R. Streaming deep reinforcement learning finally works. *arXiv preprint arXiv:2410.14606*, 2024.

Ermolov, A. and Sebe, N. Latent World Models For Intrinsically Motivated Exploration. In *Advances in Neural Information Processing Systems*, volume 33, pp. 5565–5575. Curran Associates, Inc., 2020. URL https://papers.nips.cc/paper_files/paper/2020/hash/3c09bb10e2189124fdd8f467cc8b55a7-Abstract.html.

Gallistel, C. R. *The Organization of Learning*. The MIT Press, 1990.

Garcia-Porta, J., Sol, D., Pennell, M., Sayol, F., Kaliontzopoulou, A., and Botero, C. A. Niche expansion and adaptive divergence in the global radiation of crows and ravens. *Nature Communications*, 13(1): 2086, April 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29707-5. URL https://www.nature.com/articles/s41467-022-29707-5.

Gazerani, P. The neuroplastic brain: current breakthroughs and emerging frontiers. *Brain Research*, 1858:149643, July 2025. ISSN 0006-8993. doi: 10.1016/j.brainres.2025.149643. URL https://www.sciencedirect.com/science/article/pii/S0006899325002021.

Gu, A. and Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. URL https://openreview.net/forum?id=tEYskw1VY2.

Gupta, A., Yu, J., Zhao, T. Z., Kumar, V., Rovinsky, A., Xu, K., Devlin, T., and Levine, S. Reset-free reinforcement learning via multi-task learning: Learning dexterous manipulation behaviors without human intervention. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6664–6671. IEEE Press, 2021. doi: 10.1109/ICRA48506.2021.9561384. URL https://doi.org/10.1109/ICRA48506.2021.9561384.

Haarlem, C. S., Hynes, C., Jackson, A. L., Mitchell, K. J., O'Connell, R. G., and Healy, K. Pace of ecology drives the tempo of visual perception across the animal kingdom. *Nature Ecology & Evolution*, 10(4):712–720, April 2026. ISSN 2397-334X. doi: 10.1038/s41559-026-02994-7. URL https://www.nature.com/articles/s41559-026-02994-7.

Hafner, D., Pasukonis, J., Ba, J., and Lillicrap, T. Mastering diverse control tasks through world models. *Nature*, 640(8059):647–653, April 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-08744-2. URL https://www.nature.com/articles/s41586-025-08744-2.

Hari, R. and Parkkonen, L. The brain timewise: how timing shapes and supports brain function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668):20140170, May 2015. ISSN 0962-8436. doi: 10.1098/rstb.2014.0170. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4387511/.

Hasani, R. M., Lechner, M., Amini, A., Rus, D., and Grosu, R. Liquid time-constant networks. In *AAAI Conference on Artificial Intelligence*, 2020. URL https://api.semanticscholar.org/CorpusID:219531443.

Hausknecht, M. J. and Stone, P. Deep recurrent q-learning for partially observable mdps. *ArXiv*, abs/1507.06527, 2015. URL https://api.semanticscholar.org/CorpusID:8696662.

Hayes, T. L., Krishnan, G. P., Bazhenov, M., Siegelmann, H. T., Sejnowski, T. J., and Kanan, C. Replay in Deep Learning: Current Approaches and Missing Biological Elements. *Neural computation*, 33(11):2908–2950, October 2021. ISSN 0899-7667. doi: 10.1162/neco_a_01433. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC9074752/.

Hebb, D. O. *The organization of behavior; a neuropsychological theory*. The organization of behavior; a neuropsychological theory. Wiley, Oxford, England, 1949. Pages: xix, 335.

Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature reviews neuroscience*, 6(11):877–888, 2005.

Hintze, A., Olson, R. S., Adami, C., and Hertwig, R. Risk sensitivity as an evolutionary adaptation. *Scientific Reports*, 5(1):8242, February 2015. ISSN 2045-2322. doi: 10.1038/srep08242. URL https://www.nature.com/articles/srep08242.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., and Chen, W. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Hu, J., Shim, J., Tang, C., Sung, Y., Liu, B., Stone, P., and Martin-Martin, R. Simple recipe works: Vision-language-action models are natural continual learners with reinforcement learning. *arXiv preprint arXiv:2603.11653*, 2026. URL https://arxiv.org/abs/2603.11653.

Hupbach, A., Gomez, R., Hardt, O., and Nadel, L. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. *Learning & Memory*, 14(1-2):47–53, 2007. ISSN 1549-5485. doi: 10.1101/lm.365707.

Itoh, K., Konoike, N., Nejime, M., Iwaoki, H., Igarashi, H., Hirata, S., and Nakamura, K. Cerebral cortical processing time is elongated in human brain evolution. *Scientific Reports*, 12(1):1103, January 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-05053-w. URL https://www.nature.com/articles/s41598-022-05053-w.

Janmaat, K. R. Temporal cognition in Taï chimpanzees. In Boesch, C. and Wittig, R. (eds.), *The Chimpanzees of the Taï Forest: 40 Years of Research*, pp. 451–466. Cambridge University Press, Cambridge, 2019. ISBN 978-1-108-48155-7. doi: 10.1017/9781108674218.029. URL https://www.cambridge.org/core/books/chimpanzees-of-the-tai-forest/temporal-cognition-in-tai-chimpanzees/47690597195F4C0CFC7F5C2BA549D3F6.

Janner, M., Fu, J., Zhang, M., and Levine, S. When to Trust Your Model: Model-Based Policy Optimization. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/5faf461eff3099671ad63c6f3f094f7f-Abstract.html.

Javed, K. and Sutton, R. S. The big world hypothesis and its ramifications for artificial intelligence. In *Finding the Frame: An RLC Workshop for Examining Conceptual Frameworks*, 2024.

Jonker, T. R., Seli, P., and MacLeod, C. M. Putting retrieval-induced forgetting in context: an inhibition-free, context-based account. *Psychological Review*, 120(4):852–872, October 2013. ISSN 1939-1471. doi: 10.1037/a0034246.

Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., and Ranganath, C. Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proceedings of the National Academy of Sciences*, 115 (43):11084–11089, October 2018. doi: 10.1073/pnas.1800006115. URL https://www.pnas.org/doi/full/10.1073/pnas.1800006115.

Juliani, A. and Ash, J. T. A study of plasticity loss in on-policy deep reinforcement learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

K, T. D., Fischer, T., and Biemann, C. Large Language Models Are Overparameterized Text Encoders. In Adlakha, V., Chronopoulou, A., Li, X. L., Majumder, B. P., Shi, F., and Vernikos, G. (eds.), *Proceedings of the 10th Workshop on Representation Learning for NLP (RepL4NLP-2025)*, pp. 170–184, Albuquerque, NM, May 2025. Association for Computational Linguistics. ISBN 979-8-89176-245-9. doi: 10.18653/v1/2025.repl4nlp-1.13. URL https://aclanthology.org/2025.repl4nlp-1.13/.

Kaufmann, A. *Temporal Cognition in Animals*. Elements in the Philosophy of Biology. Cambridge University Press, 2026.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., and Hadsell, R. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. doi: 10.1073/pnas.1611835114. URL https://www.pnas.org/doi/abs/10.1073/pnas.1611835114.

Klissarov, M., Bagaria, A., Luo, Z., Konidaris, G., Precup, D., and Machado, M. C. Discovering temporal structure:

An overview of hierarchical reinforcement learning, 2025. URL https://arxiv.org/abs/2506.14045.

Knudsen, E. I. Sensitive periods in the development of the brain and behavior. *Journal of cognitive neuroscience*, 16 (8):1412–1425, 2004.

Kobayashi, S., Schimpf, Y., Schlegel, M., Steger, A., Wolczyk, M., von Oswald, J., Scherrer, N., Maile, K., Lajoie, G., Richards, B. A., Saurous, R. A., Manyika, J., y Arcas, B. A., Meulemans, A., and Sacramento, J. Emergent temporal abstractions in autoregressive models enable hierarchical reinforcement learning, 2025. URL https://arxiv.org/abs/2512.20605.

Kronfeld-Schor, N., Bloch, G., and Schwartz, W. J. Animal clocks: when science meets nature. *Proceedings of the Royal Society B: Biological Sciences*, 280 (1765):20131354, August 2013. ISSN 0962-8452. doi: 10.1098/rspb.2013.1354. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3712458/.

Kumar, A., Lu, C., Kirsch, L., Tang, Y., Stanley, K. O., Isola, P., and Ha, D. Automating the search for artificial life with foundation models. 2024. URL https://asal.sakana.ai/.

Labash, A., Stelzer, F., Majoral, D., and Vicente, R. Emergence of adaptive circadian rhythms in deep reinforcement learning. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Lafuente, E. and Beldade, P. Genomics of Developmental Plasticity in Animals. *Frontiers in Genetics*, 10, August 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.00720. URL https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.00720/full.

Lauri, M., Hsu, D., and Pajarinen, J. Partially observable markov decision processes in robotics: A survey. *IEEE Transactions on Robotics*, PP:1–20, 01 2022. doi: 10.1109/TRO.2022.3200138.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. Retrieval-augmented generation for knowledge-intensive nlp tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Li, A. C., Vaezipoor, P., Icarte, R. T., and McIlraith, S. A. Challenges to solving combinatorially hard long-horizon deep rl tasks, 2022. URL https://arxiv.org/abs/2206.01812.

Li, Q. Vlas are confined yet capable of generalizing to novel instructions, 2026. URL https://arxiv.org/abs/2505.03500.

Li, T., Zhang, G., Do, Q. D., Yue, X., and Chen, W. Long-context LLMs struggle with long in-context learning. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=Cw2xlg0e46.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N. M. O., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. Continuous control with deep reinforcement learning. *CoRR*, abs/1509.02971, 2015.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., and Liang, P. Lost in the Middle: How Language Models Use Long Contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173, 2024. doi: 10.1162/tacl_a_00638. URL https://aclanthology.org/2024.tacl-1.9/.

Liverence, B. M. and Scholl, B. J. Discrete events as units of perceived time. *Journal of Experimental Psychology. Human Perception and Performance*, 38(3):549–554, June 2012. ISSN 1939-1277. doi: 10.1037/a0027228.

Luo, Y., Yang, Z., Meng, F., Li, Y., Zhou, J., and Zhang, Y. An empirical study of catastrophic forgetting in large language models during continual fine-tuning. *IEEE Transactions on Audio, Speech and Language Processing*, 33:3776–3786, 2025. doi: 10.1109/TASLPRO.2025.3606231.

Lyle, C., Zheng, Z., Nikishin, E., Avila Pires, B., Pascanu, R., and Dabney, W. Understanding plasticity in neural networks. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 23190–23211. PMLR, 23–29 Jul 2023.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, July 1995. ISSN 0033-295X. doi: 10.1037/0033-295X.102.3.419.

Mei, L., Yao, J., Ge, Y., Wang, Y., Bi, B., Cai, Y., Liu, J., Li, M., Li, Z.-Z., Zhang, D., Zhou, C., Mao, J., Xia, T., Guo, J., and Liu, S. A survey of context engineering for large language models, 2025. URL https://arxiv.org/abs/2507.13334.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie,

C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. Human-level control through deep reinforcement learning. *Nature*, 518 (7540):529–533, February 2015. ISSN 1476-4687. doi: 10.1038/nature14236.

Moerland, T. M., Broekens, J., Plaat, A., and Jonker, C. M. Model-based reinforcement learning: A survey. *Found. Trends Mach. Learn.*, 16(1):1–118, January 2023. ISSN 1935-8237. doi: 10.1561/2200000086. URL https://doi.org/10.1561/2200000086.

Moore, D., Siegfried, D., Wilson, R., and Rankin, M. A. The influence of time of day on the foraging behavior of the honeybee, Apis mellifera. *Journal of Biological Rhythms*, 4(3):305–325, 1989. ISSN 0748-7304. doi: 10.1177/074873048900400301.

Moser, E. I., Kropff, E., and Moser, M.-B. Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience*, 31:69–89, 2008. ISSN 0147-006X. doi: 10.1146/annurev.neuro.31.061307.090723.

Nam, T., Sun, S.-H., Pertsch, K., Hwang, S. J., and Lim, J. J. Skill-based meta-reinforcement learning. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=jeLW-Fh9bV.

NVIDIA, Bjorck, J., Fernando Castañeda, N. C., Da, X., Ding, R., Fan, L. J., Fang, Y., Fox, D., Hu, F., Huang, S., Jang, J., Jiang, Z., Kautz, J., Kundalia, K., Lao, L., Li, Z., Lin, Z., Lin, K., Liu, G., Llontop, E., Magne, L., Mandlekar, A., Narayan, A., Nasiriany, S., Reed, S., Tan, Y. L., Wang, G., Wang, Z., Wang, J., Wang, Q., Xiang, J., Xie, Y., Xu, Y., Xu, Z., Ye, S., Yu, Z., Zhang, A., Zhang, H., Zhao, Y., Zheng, R., and Zhu, Y. GR00T N1: An open foundation model for generalist humanoid robots. In *ArXiv Preprint*, March 2025.

Ongchoco, J. D. K., Yates, T. S., and Scholl, B. J. Event segmentation structures temporal experience: Simultaneous dilation and contraction in rhythmic reproductions. *Journal of Experimental Psychology: General*, 152(11):3266–3276, 2023. ISSN 1939-2222. doi: 10.1037/xge0001447.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., and Lowe, R. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744, December 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/b1efde53be364a73914f58805a001731-Abstract-Confere
html.

Pavlides, C. and Winson, J. Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *Journal of Neuroscience*, 9(8):2907–2918, August 1989. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.09-08-02907.1989. URL https://www.jneurosci.org/content/9/8/2907.

Pitsawong, W., Pádua, R. A. P., Grant, T., Hoemberger, M., Otten, R., Bradshaw, N., Grigorieff, N., and Kern, D. From primordial clocks to circadian oscillators. *Nature*, 616(7955):183–189, April 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-05836-9. URL https://doi.org/10.1038/s41586-023-05836-9.

Poucet, B., Lenck-Santini, P. P., Hok, V., Save, E., Banquet, J. P., Gaussier, P., and Muller, R. U. Spatial navigation and hippocampal place cell firing: the problem of goal encoding. *Reviews in the Neurosciences*, 15(2):89–107, 2004. ISSN 0334-1763. doi: 10.1515/revneuro.2004.15.2.89.

Radford, A. and Narasimhan, K. Improving Language Understanding by Generative Pre-Training. 2018. URL https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Ra
cd18800a0fe0b668a1cc19f2ec95b5003d0a5035.

Ramstedt, S. and Pal, C. Real-Time Reinforcement Learning. In Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F. d., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/54e36c5ff5f6a1802925ca009f3ebb68-Paper.pdf.

Rennolds, C. W. and Bely, A. E. Integrative biology of injury in animals. *Biological Reviews of the Cambridge Philosophical Society*, 98(1):34–62, February 2023. ISSN 1464-7931. doi: 10.1111/brv.12894. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10087827/.

Rubanova, Y., Chen, R. T. Q., and Duvenaud, D. K. Latent Ordinary Differential Equations for Irregularly-Sampled Time Series. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/42a6845a557bef704ad8ac9cb4461d43-Abstract.html.

Schaul, T., Quan, J., Antonoglou, I., and Silver, D. Prioritized experience replay, 2016. URL https://arxiv.org/abs/1511.05952.

Schlag, I., Irie, K., and Schmidhuber, J. Linear Transformers Are Secretly Fast Weight Programmers. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 9355–9366. PMLR, July 2021. URL https://proceedings.mlr.press/v139/schlag21a.html.

Schuman, C. D., Potok, T. E., Patton, R. M., Birdwell, J. D., Dean, M. E., Rose, G. S., and Plank, J. S. A survey of neuromorphic computing and neural networks in hardware. *ArXiv*, abs/1705.06963, 2017. URL https://api.semanticscholar.org/CorpusID:605892.

Sennrich, R., Haddow, B., and Birch, A. Neural machine translation of rare words with subword units. In Erk, K. and Smith, N. A. (eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1715–1725, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1162. URL https://aclanthology.org/P16-1162/.

Shen, T. and Dayan, P. Risking your Tail: Curiosity, Danger and Exploration, January 2024. URL https://www.biorxiv.org/content/10.1101/2024.01.07.574574v1. Pages: 2024.01.07.574574 Section: New Results.

Shi, H., Xie, B., Liu, Y., Sun, L., Liu, F., Wang, T., Zhou, E., Fan, H., Zhang, X., and Huang, G. MemoryVLA: Perceptual-cognitive memory in vision-language-action models for robotic manipulation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=54U3XHf7qq.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016. ISSN 1476-4687. doi: 10.1038/nature16961. URL https://doi.org/10.1038/nature16961.

Sutton, R. S. and Barto, A. G. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018. ISBN 0262039249.

Sutton, R. S., Precup, D., and Singh, S. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1):181–211, 1999. ISSN 0004-3702. doi: https://doi.org/10.1016/S0004-3702(99)00052-1. URL https://www.sciencedirect.com/science/article/pii/S0004370299000521.

Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., and Maida, A. Deep learning in spiking neural networks. *Neural Netw.*, 111(C):47–63, March 2019. ISSN 0893-6080. doi: 10.1016/j.neunet.2018.12.002. URL https://doi.org/10.1016/j.neunet.2018.12.002.

Team, G., Mesnard, T., Hardin, C., Dadashi, R., Bhupatiraju, S., Pathak, S., Sifre, L., Rivière, M., Kale, M. S., Love, J., Tafti, P., Hussenot, L., Sessa, P. G., Chowdhery, A., Roberts, A., Barua, A., Botev, A., Castro-Ros, A., Slone, A., Héliou, A., Tacchetti, A., Bulanova, A., Paterson, A., Tsai, B., Shahriari, B., Lan, C. L., Choquette-Choo, C. A., Crepy, C., Cer, D., Ippolito, D., Reid, D., Buchatskaya, E., Ni, E., Noland, E., Yan, G., Tucker, G., Muraru, G.-C., Rozhdestvenskiy, G., Michalewski, H., Tenney, I., Grishchenko, I., Austin, J., Keeling, J., Labanowski, J., Lespiau, J.-B., Stanway, J., Brennan, J., Chen, J., Ferret, J., Chiu, J., Mao-Jones, J., Lee, K., Yu, K., Millican, K., Sjoesund, L. L., Lee, L., Dixon, L., Reid, M., Mikuła, M., Wirth, M., Sharman, M., Chinaev, N., Thain, N., Bachem, O., Chang, O., Wahltinez, O., Bailey, P., Michel, P., Yotov, P., Chaabouni, R., Comanescu, R., Jana, R., Anil, R., McIlroy, R., Liu, R., Mullins, R., Smith, S. L., Borgeaud, S., Girgin, S., Douglas, S., Pandya, S., Shakeri, S., De, S., Klimenko, T., Hennigan, T., Feinberg, V., Stokowiec, W., hui Chen, Y., Ahmed, Z., Gong, Z., Warkentin, T., Peran, L., Giang, M., Farabet, C., Vinyals, O., Dean, J., Kavukcuoglu, K., Hassabis, D., Ghahramani, Z., Eck, D., Barral, J., Pereira, F., Collins, E., Joulin, A., Fiedel, N., Senter, E., Andreev, A., and Kenealy, K. Gemma: Open models based on gemini research and technology, 2024. URL https://arxiv.org/abs/2403.08295.

Team, O., Stooke, A., Mahajan, A., Barros, C., Deck, C., Bauer, J., Sygnowski, J., Trebacz, M., Jaderberg, M., Mathieu, M., McAleese, N., Bradley-Schmieg, N., Wong, N., Porcel, N., Raileanu, R., Hughes-Fitt, S., Dalibard, V., and Czarnecki, W. M. Open-ended learning leads to generally capable agents. *ArXiv*, abs/2107.12808, 2021. URL https://api.semanticscholar.org/CorpusID:236447390.

Teicher, M. H., Samson, J. A., Anderson, C. M., and Ohashi, K. The effects of childhood maltreatment on brain structure, function and connectivity. *Nature reviews neuroscience*, 17(10):652–666, 2016.

Tovée, M. J. Neuronal Processing: How fast is the speed of thought? *Current Biology*, 4

(12):1125–1127, December 1994. ISSN 0960-9822. doi: 10.1016/S0960-9822(00)00253-0. URL https://www.sciencedirect.com/science/article/pii/S0960982200002530.

Towers, M., Kwiatkowski, A., Balis, J. U., Cola, G. D., Deleu, T., Goulão, M., Andreas, K., Krimmel, M., KG, A., Perez-Vicente, R. D. L., Terry, J. K., Pierré, A., Schulhoff, S. V., Tai, J. J., Tan, H., and Younis, O. G. Gymnasium: A standard interface for reinforcement learning environments. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2026. URL https://openreview.net/forum?id=qPMLvJxtPK.

Trepel, C., Fox, C. R., and Poldrack, R. A. Prospect theory on the brain? Toward a cognitive neuroscience of decision under risk. *Brain Research. Cognitive Brain Research*, 23 (1):34–50, April 2005. ISSN 0926-6410. doi: 10.1016/j.cogbrainres.2005.01.016.

Vasan, G., Elsayed, M., Azimi, S. A., He, J., Shahriar, F., Bellinger, C., White, M., and Mahmood, A. R. Deep policy gradient methods without batch updates, target networks, or replay buffers. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=DX5GUwMFFb.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Vieira, I., Allred, W., Lankford, S., Castilho, S., and Way, A. How much data is enough data? fine-tuning large language models for in-house translation: Performance evaluation across multiple dataset sizes. In Knowles, R., Eriguchi, A., and Goel, S. (eds.), *Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pp. 236–249, Chicago, USA, September 2024. Association for Machine Translation in the Americas. URL https://aclanthology.org/2024.amta-research.20/.

von Bernhardi, R., Bernhardi, L. E.-v., and Eugenín, J. What Is Neural Plasticity? In von Bernhardi, R., Eugenín, J., and Muller, K. J. (eds.), *The Plastic Brain*, pp. 1–15. Springer International Publishing, Cham, 2017. ISBN 978-3-319-62817-2. doi: 10.1007/978-3-319-62817-2_1. URL https://doi.org/10.1007/978-3-319-62817-2_1.

Wan, Y., Korenkevych, D., and Zhu, Z. An empirical study of deep reinforcement learning in continuing tasks, 2025. URL https://arxiv.org/abs/2501.06937.

Wang, L., Zhang, X., Su, H., and Zhu, J. A Comprehensive Survey of Continual Learning: Theory, Method and Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5362–5383, August 2024a. ISSN 1939-3539. doi: 10.1109/TPAMI.2024.3367329. URL https://ieeexplore.ieee.org/document/10444954/.

Wang, X., Salmani, M., Omidi, P., Ren, X., Rezagholizadeh, M., and Eshaghi, A. Beyond the limits: A survey of techniques to extend the context length in large language models. In *International Joint Conference on Artificial Intelligence*, 2024b. URL https://api.semanticscholar.org/CorpusID:267412232.

Wehner, R. and Srinivasan, M. V. Searching behaviour of desert ants, genusCataglyphis (Formicidae, Hymenoptera). *Journal of comparative physiology*, 142 (3):315–338, September 1981. ISSN 1432-1351. doi: 10.1007/BF00605445. URL https://doi.org/10.1007/BF00605445.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E. H., Le, Q. V., and Zhou, D. Chain-of-thought prompting elicits reasoning in large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Wołczyk, M., Zając, M., Pascanu, R., Kuciński, Ł., and Miłoś, P. Continual world: A robotic benchmark for continual reinforcement learning. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=5qsptDcsdEj.

Wong, B. B. and Candolin, U. Behavioral responses to changing environments. *Behavioral Ecology*, 26 (3):665–673, May 2015. ISSN 1045-2249. doi: 10.1093/beheco/aru183. URL https://doi.org/10.1093/beheco/aru183.

Xia, Y., Kim, J., Chen, Y., Ye, H., Kundu, S., Hao, C., and Talati, N. Understanding the performance and estimating the cost of llm fine-tuning. *2024 IEEE International Symposium on Workload Characterization (IISWC)*, pp. 210–223, 2024. URL https://api.semanticscholar.org/CorpusID:271843302.

Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., Zheng, C., Liu, D., Zhou, F., Huang, F., Hu, F., Ge, H., Wei, H., Lin, H., Tang, J., Yang, J., Tu, J., Zhang, J., Yang, J., Yang, J., Zhou, J., Zhou, J., Lin, J., Dang, K., Bao, K., Yang, K., Yu, L., Deng, L., Li, M., Xue, M., Li, M., Zhang, P., Wang, P., Zhu, Q., Men, R., Gao, R., Liu, S., Luo, S., Li, T., Tang, T., Yin, W., Ren, X., Wang, X., Zhang, X., Ren, X., Fan, Y., Su, Y., Zhang, Y., Zhang, Y., Wan, Y., Liu, Y., Wang, Z., Cui, Z., Zhang, Z., Zhou, Z., and Qiu, Z. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. Event perception: a mind-brain perspective. *Psychological Bulletin*, 133(2):273–293, March 2007. ISSN 0033-2909. doi: 10.1037/0033-2909.133.2.273.

Zanatta, L., Barchi, F., Manoni, S., Tolu, S., Bartolini, A., and Acquaviva, A. Exploring spiking neural networks for deep reinforcement learning in robotic tasks. *Scientific Reports*, 14(1):30648, December 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-77779-8. URL https://www.nature.com/articles/s41598-024-77779-8.

Zhang, R., Brennan, T. J., and Lo, A. W. The origin of risk aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 111(50):17777–17782, December 2014. ISSN 0027-8424. doi: 10.1073/pnas.1406755111. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC4273387/.

Zhang, S., Dong, L., Li, X., Zhang, S., Sun, X., Wang, S., Li, J., Hu, R., Zhang, T., Wang, G., and Wu, F. Instruction tuning for large language models: A survey. *ACM Comput. Surv.*, 58(7), January 2026. ISSN 0360-0300. doi: 10.1145/3777411. URL https://doi.org/10.1145/3777411.

Zhang, Z. and Weihs, L. When learning is out of reach, reset: Generalization in autonomous visuomotor reinforcement learning. In *First Workshop on Out-of-Distribution Generalization in Robotics at CoRL 2023*, 2023. URL https://openreview.net/forum?id=QAW4wDJke6.

Zhao, T., Kumar, V., Levine, S., and Finn, C. Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware. In *Robotics: Science and Systems XIX*. Robotics: Science and Systems Foundation, July 2023. ISBN 978-0-9923747-9-2. doi: 10.15607/RSS.2023.XIX.016. URL http://www.roboticsproceedings.org/rss19/p016.pdf.

Zhou, Y., Liu, H., Chen, Z., Tian, Y., and Chen, B. GSM-∞: How do your LLMs behave over infinitely increasing reasoning complexity and context length? In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=n52yyvEwPa.

Zitkovich, B., Yu, T., Xu, S., Xu, P., Xiao, T., Xia, F., Wu, J., Wohlhart, P., Welker, S., Wahid, A., Vuong, Q., Vanhoucke, V., Tran, H., Soricut, R., Singh, A., Singh, J., Sermanet, P., Sanketi, P. R., Salazar, G., Ryoo, M. S., Reymann, K., Rao, K., Pertsch, K., Mordatch, I., Michalewski, H., Lu, Y., Levine, S., Lee, L., Lee, T.-W. E., Leal, I., Kuang, Y., Kalashnikov, D., Julian, R., Joshi, N. J., Irpan, A., Ichter, B., Hsu, J., Herzog, A., Hausman, K., Gopalakrishnan, K., Fu, C., Florence, P., Finn, C., Dubey, K. A., Driess, D., Ding, T., Choromanski, K. M., Chen, X., Chebotar, Y., Carbajal, J., Brown, N., Brohan, A., Arenas, M. G., and Han, K. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In Tan, J., Toussaint, M., and Darvish, K. (eds.), *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pp. 2165–2183. PMLR, 06–09 Nov 2023. URL https://proceedings.mlr.press/v229/zitkovich23a.html.

Åström, K. Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10 (1):174–205, February 1965. ISSN 0022247X. doi: 10.1016/0022-247X(65)90154-X. URL https://linkinghub.elsevier.com/retrieve/pii/0022247X6590154X.

A. Detailed Discussion of Reinforcement Learning

Reinforcement learning (RL) is the machine learning paradigm most directly associated with learning from experience. In its standard formulation, RL models sequential decision-making as a Markov Decision Process (MDP) given by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ denotes transition probabilities, $\mathcal{R}$ is the reward function, and $\gamma \in [0, 1]$ is the discount factor (Sutton & Barto, 2018). The agent's decisions are governed by a policy π, and the objective is to find a policy π_θ that maximizes expected return:

$$J(\theta) = \mathbb{E}_{\pi_\theta}\Big[\sum_{t=0}^{T} \gamma^t \mathcal{R}(S_t, A_t)\Big],$$

where T is the episode horizon, with $T = \infty$ for infinite-horizon or continuing tasks (Sutton & Barto, 2018). This formalism naturally captures sequential decision-making, but it also introduces several assumptions that conflict with temporally grounded learning.

A.1. Engagement with Environmental Guidelines

Autonomous external time. Because of its reliance on MDPs, time in RL is typically represented as uniformly spaced discrete transitions governed by a control frequency or fixed timestep. This assertion is true across environments ranging from continuous control (Lillicrap et al., 2015) to discrete action spaces (Mnih et al., 2015). It clashes with Guideline 1 in two ways. First, most RL environments (Towers et al., 2026) advance synchronously with the agent through a `step()` function: the environment remains in stasis while the agent computes an action, and only then applies its transition dynamics $\mathcal{P}$. As a result, clock-time inference speed and computational latency are often excluded from the learning problem. Second, uniform discretization simplifies the continuous nature of external time. Given a stream of percepts refreshed at some frequency, the standard formulation assumes that the agent acts at each new observation rather than deciding when action is needed, how urgently it must act, or how long it can afford to deliberate.

Prior works in real-time RL have proposed augmenting the standard MDP formulation to simulate real-time interaction via synchronously evolving state-action pairs (Ramstedt & Pal, 2019) or random delays by appending an action buffer to the delayed observation (Bouteiller et al., 2021). Formulations such as semi-MDPs (Bradtke & Duff, 1994) have been proposed as continuous-time variants of MDPs, though they remain less central in modern deep RL. Some recent work has explored related ideas using Neural ODEs (Du et al., 2020). We discuss options over MDPs (Sutton et al., 1999) below, since they partially reconcile traditional MDPs with temporally extended decisions.

Partial embodied observability. An RL problem can be fully observable or partially observable, as in a POMDP (Åström, 1965). However, because POMDPs are difficult to solve, much of RL addresses partial observability either by approximating the problem as an MDP (Anjarlekar et al., 2026) or by using architectures that incorporate history, such as recurrent neural networks (Hausknecht & Stone, 2015). This makes RL one of the paradigms that most directly engages Guideline 2, especially in embodied domains such as robotics, where agents often receive only egocentric observations of the environment (Lauri et al., 2022).

Still, partial observability is often treated as a technical obstacle rather than as a defining condition of temporal experience. The use of history also raises a deeper question: if the agent requires memory or recurrent state to act well, then the Markovian assumption that the present state is sufficient is no longer the right abstraction for the learner's actual experience. In this sense, recurrent or history-based RL partially moves toward the persistent internal structures emphasized by Guideline 6, even when it remains embedded in standard episodic training loops.

Non-resettable lifetime interaction. A central dogma in much of RL is to view learning as optimization for a narrow fixed task rather than as open-ended adaptation (Abel et al., 2024). Many of RL's most visible successes, from superhuman game-playing (Silver et al., 2016) to dexterous robotic manipulation (Andrychowicz et al., 2020), are built around fixed task definitions and episodic horizons. Moving away from this framework toward infinite-horizon, reset-free interaction, as required by Guideline 3, remains an open research area. Some work has begun to explore reset-free or reset-minimal settings, particularly in robotic manipulation (Gupta et al., 2021; Zhang & Weihs, 2023). Closely related is single-life reinforcement learning, where an agent must complete a task within a single episode without intervention while using prior experience to recover from novelty or out-of-distribution states (Chen et al., 2022). This framing directly targets one of the central difficulties of non-resettable interaction: the agent cannot rely on repeated attempts, external resets, or human correction once it enters an unfamiliar state.

Resets act as a powerful crutch. They constrain the effective state space $\mathcal{S}$, limit the exploration region, and allow agents to escape from states in which they would otherwise be trapped (Wan et al., 2025). Removing resets therefore makes the problem substantially harder. Agents must not only learn task success, but also avoid unrecoverable states, manage accumulated consequences, and recover from mistakes. This compounds existing challenges in long-horizon RL, including objective mismatch under discounting (Li et al., 2022).

A.2. Engagement with Agent Design Guidelines

Bounded reconstructive memory. One of the grand challenges of RL is sample efficiency: we want agents that can learn from limited experience. Foundational RL algorithms such as TD learning, SARSA, and actor-critic methods (Sutton & Barto, 2018) were originally designed for streaming learning, where each transition is used to update the policy and then discarded. However, as RL adopted deep neural networks as function approximators, deep RL also adopted batch-style parameter updates and replay buffers. Replay buffers improve sample efficiency and stabilize training (Andrychowicz et al., 2017; Mnih et al., 2015; Schaul et al., 2016), but they also conflict with Guideline 4: they store lossless transitions and allow the learner to repeatedly revisit identical past samples.

Recent work has shown that, with appropriate optimizers and regularization, deep RL can sometimes be performed in a streaming manner with performance competitive with

batch-based methods (Elsayed et al., 2024; Vasan et al., 2024). This result is encouraging because it suggests that modern RL may not require lossless replay as strongly as often assumed. However, current streaming deep RL remains limited in scope and is frequently evaluated under the same short-horizon, resettable, episodic assumptions that dominate the broader field. Its use of parametric-only memory moves toward Guideline 4 and Guideline 6, but does not yet provide a full account of bounded, reconstructive memory in non-resettable lifetime interaction.

Multi-rate endogenous computation. The atomic unit of experience in standard RL is the transition (s_t, a_t, r_t, s_{t+1}), consisting of a state, action, reward, and subsequent state. Meaningful experience, however, often spans many transitions: a search, a manipulation attempt, a social interaction, a failure recovery, or a strategy shift may unfold over extended time. Episodes provide one engineered way to bound such experiences. But because most single-task and even multi-task RL (Team et al., 2021) focuses on episodic learning and execution, the field has only a limited understanding of how agents should represent and organize experiences over long-horizon streams of percepts.

Hierarchical RL is the subfield that most directly engages this issue. It often uses the options-over-MDP framework (Klissarov et al., 2025), in which high-level options define temporally extended behaviors over variable-length MDP intervals (Sutton et al., 1999). This partially addresses Guideline 5 by introducing multiple timescales of decision-making. However, approaches such as Option-Critic (Bacon et al., 2017) can suffer from option degeneracy, where learned options collapse into non-robust or single-timestep behaviors. Newer approaches that use a meta-controller for internal RL remain constrained by short episodes and small option sets (Kobayashi et al., 2025). Thus, hierarchical RL directly engages temporal abstraction, but substantial work remains to discover effective multi-rate structures online and in long-lived agents.

Most RL systems also do not require agents to maintain internal representations of time or schedule behavior with respect to a global temporal structure in the environment. Labash et al. (2023) show that adaptive circadian-like rhythms and anticipatory behavior can emerge in recurrent RL agents in a foraging environment. However, this work remains limited to a relatively short-horizon and episodic setting. Guideline 5 therefore remains a major open area for RL, especially for agents that must operate longitudinally in infinite-horizon environments.

Path-dependent self-modification. RL engages path-dependent self-modification in two main ways: through neural parameter updates and through continual RL. Continual RL focuses on agents that can adapt indefinitely (Abel et al., 2023), making it naturally relevant to Guideline 6. However, continual RL often still violates Guideline 3, since many problems remain tethered to episodic structures and are framed as sequential task learning (Wolczyk et al., 2021).

Despite these limitations, continual RL has become an important setting for studying the stability-plasticity tradeoff in deep learning. Neural networks are known to suffer from catastrophic forgetting (Kirkpatrick et al., 2017) and plasticity loss (Dohare et al., 2024; Juliani & Ash, 2024; Lyle et al., 2023). These phenomena are directly relevant to temporally grounded learning because they concern how an agent's internal structure changes over time while attempting to remain both stable and adaptable. Thus, continual RL addresses important aspects of Guideline 6, even though it usually does so within temporal abstractions inherited from standard RL.

A.3. Summary.

RL provides many ingredients for temporally grounded learning: interaction, sequential decision-making, partial observability, streaming updates, temporal abstraction, and continual adaptation. Yet, standard RL also spatializes time through synchronized steps, resettable episodes, replay buffers, and evaluation protocols that focus on final task performance. Reset-free RL, reset-minimal RL, single-life RL, streaming RL, hierarchical RL, POMDP methods, and continual RL each address important fragments of the problem. However, no dominant RL formulation currently treats all of our proposed guidelines as simultaneous design constraints.

B. Detailed Discussion of Transformer-Based Foundation Models

We next consider transformer-based foundation models trained through supervised or self-supervised learning. This category includes large language models (LLMs) (Team et al., 2024; Yang et al., 2025), vision-language models (VLMs) (Deitke et al., 2024), and the emerging class of vision-language-action models (VLAs) (Black et al., 2025; NVIDIA et al., 2025; Zitkovich et al., 2023). These models have transformed modern machine learning and provide powerful mechanisms for representing temporal data. However, they also exemplify the spatialization of time: experience is converted into token sequences, context windows, and static training corpora.

B.1. Engagement with Environmental Guidelines

Autonomous external time. Most foundation models use transformer architectures that perform sequential prediction over tokens. Tokens are the atomic units of experience: strings of text (Sennrich et al., 2016), image patches (Doso-

vitskiy et al., 2021), or action chunks (Zhao et al., 2023) are converted into vector embeddings and combined with positional encodings (Chen et al., 2021; Vaswani et al., 2017) to locate each token within a sequence. This structure is powerful, but it represents time as positions in an indexable context window. During inference, the model typically assumes that the relevant world is static while the forward pass or autoregressive generation unfolds.

This conflicts with Guideline 1. The model is not usually constrained by the fact that the external world continues to evolve during inference, and its internal objective does not account for the latency of decision-making. In embodied domains, this can be catastrophic: a delayed action may no longer be appropriate by the time it is executed. This problem has already affected VLAs in the physical world, leading researchers to engineer mechanisms around the temporal limitations of large models (Black et al., 2026). Thus, foundation models currently fail to meet Guideline 1: they spatialize time through context windows, assume the world is effectively in stasis during inference, and lack native mechanisms for reasoning about the world-time cost of computation.

Partial embodied observability. The Big World Hypothesis (Javed & Sutton, 2024) argues that the world relevant to a learning problem is orders of magnitude larger than the agent, and therefore can never be fully observed. Foundation models respond to this problem indirectly. In language, they sidestep partial observability through extensive pretraining (Radford & Narasimhan, 2018) and overparameterization (K et al., 2025), which can give them broad statistical coverage of many situations. This has been powerful in language domains, but it is not the same as learning through a narrow embodied sensory stream.

In the physical world, VLAs still struggle with out-of-distribution generalization (Li, 2026). Moreover, recent work shows that LLMs can fail catastrophically even in fully observable puzzle games with unconventional mechanics and long-horizon planning requirements (Earle et al., 2025). These failures align with the challenge posed by the Big World Hypothesis. Foundation models partially engage Guideline 2 through scale and pretraining, but they do not yet fully address the problem of situated, partial, temporally unfolding observability in big worlds.

Non-resettable lifetime interaction. An ideal foundation-model agent might operate longitudinally without resets for as long as its context window permits. This ambition has helped motivate techniques for extending context length in LLMs (Wang et al., 2024b). However, many studies have observed that LLM performance degrades substantially over long contexts (Du et al., 2025; Li et al., 2025; Zhou et al., 2025). As context grows, the odds of accumulated error increase while the ability to recover may degrade. The practical solution is often to clear the context and begin a new instance of the model.

This solution violates Guideline 3. Clearing the context functions as a reset: it discards the transient interaction history and returns the model to a clean baseline. It also weakens Guideline 6, because the agent's trajectory does not necessarily leave a durable structural trace. Long context windows can simulate persistence for a limited time, but they do not by themselves provide non-resettable lifetime interaction.

B.2. Engagement with Agent Design Guidelines

Bounded reconstructive memory. Foundation models rely on two broad forms of memory: parametric memory formed during pretraining and non-parametric memory supplied through the context window. We return to parametric modification below in relation to Guideline 6. For Guideline 4, the central issue is that context-window memory is not a bounded reconstructive memory system in the relevant sense. Context engineering studies how to design and optimize the information payload within the context window (Mei et al., 2025), and this process can be dynamic in practice. However, it typically occurs outside the model and within the same spatialized token sequence that the model attends over.

External memory mechanisms such as retrieval-augmented generation (RAG) extend this paradigm by connecting the model to a content-addressable vector database (Lewis et al., 2020). Similar RAG-style memories have also been extended to the VLA domain (Shi et al., 2026). These mechanisms are useful, but they generally treat memory as a static store of retrievable content. They therefore conflict with Guideline 4, which emphasizes memory as selective, compressive, revisable, and altered by recall. RAG-style memory can preserve and retrieve the past, but it can also allow agents to rewind time through lossless recall.

Multi-rate endogenous computation. Transformer-based foundation models operate primarily on a single token-based and positional timescale. Tokens are the atomic units of experience, so higher-level temporal abstractions must be composed over token sequences. This is especially visible in VLAs, where researchers must decide what temporal abstraction should be represented by a single action chunk (Zhao et al., 2023). Such engineering is useful, but it differs from the kind of internal multi-rate organization described by Guideline 5.

Chain-of-thought prompting provides an apparent form of slower reasoning, since the model generates intermediate natural-language steps before producing a final output (Wei et al., 2022). However, this process still unfolds through

the same autoregressive token loop. It generally ignores the external world while reasoning, leaving no room for interruption or feedback during the deliberative process, and the system as a whole remains governed by a single extended generation process. Thus, chain-of-thought can increase the amount of computation, but it does not by itself provide multi-rate endogenous computation. It may even exacerbate the tension with Guideline 1 by increasing latency while the world continues to evolve.

Path-dependent self-modification. Foundation models primarily engage Guideline 6 through in-context learning and fine-tuning. In-context learning allows language models to adapt behavior to new tasks using demonstrations placed in the context window (Brown et al., 2020; Dong et al., 2024). However, these demonstrations persist only as long as the context remains useful, and their effects can degrade over long contexts (Li et al., 2025). Unlike the biological examples discussed in § 2.6, language models do not autonomously improve through repeated practice of the same skill via in-context learning; they may even decline as the context becomes longer or less stable. Moreover, because context windows spatialize time, history is not necessarily structurally formative: models often perform best when relevant information appears at the beginning or end of the context rather than in the middle (Liu et al., 2024). For these reasons, in-context learning is not a sufficient response to Guideline 6.

Beyond the context window, there is a broad literature on modifying the parametric memory of language models through supervised fine-tuning (Zhang et al., 2026) and RL (Ouyang et al., 2022). However, because these models are large and relatively rigid, structural modification typically requires substantial data and computation (Vieira et al., 2024; Xia et al., 2024), or techniques that alter only a subset of parameters, such as low-rank adaptation (LoRA) (Hu et al., 2022). These updates can also suffer from fundamental deep learning issues such as catastrophic forgetting (Luo et al., 2025). Although there are promising signs in adapting VLAs with LoRA and RL fine-tuning (Hu et al., 2026), sample efficiency remains far from human-level adaptation. In practice, these challenges often lead practitioners to retrain models from scratch (Baek et al., 2026), highlighting unresolved structural issues in how large foundation models alter their internal representations from narrow streams of experience. Current foundation models therefore contain mechanisms that approximate self-modification, but they do not yet provide effective autonomous self-modification through lived experience.

B.3. Summary.

Transformer-based foundation models have transformed sequence modeling and provide powerful machinery for representing temporal structure. However, they largely learn from static corpora, operate through context windows, and adapt through transient prompting or externally orchestrated fine-tuning. They therefore excel at learning *about* temporal data, but do not yet learn *through* irreversible world-time. Their limitations are especially visible in embodied settings, where latency, partial observability, recovery, memory dynamics, and lifetime adaptation become central.