Recursive Cognitive Architecture Toward Physically Grounded Artificial General Intelligence Through Fractal Distribution and Embodied Learning

Davide Motta Director of Solution Engineering & AI Systems, RADCOM, Rome, Italy Founder, Honua davide.motta100@gmail.com

The author is a telecommunications engineer with over two decades of experience in network systems architecture and AI-based anomaly detection and analytics for large-scale telecom infrastructure. This work represents an independent theoretical contribution developed outside the scope of the author's institutional affiliations.

Contents

Abstract

We introduce the Recursive Cognitive Architecture (RCA), a distributed framework for Artificial General Intelligence in which intelligence emerges from the dynamic orchestration of a standardized unit—the Cognitive Node— that operates identically across all scales of deployment, from edge-embedded robotic manipulators to centralized cloud cores. Each node integrates four subsystems: a Perception Encoder, a World Model, a Policy Engine, and a Metacognitive Critic. This fractal consistency enables fluid scaling without translation layers between hardware tiers, allowing the system to collapse from a fully distributed configuration to a single computational substrate without architectural changes.

RCA is grounded in the thesis that general intelligence requires physical embodiment: the capacity to form predictions about the world, act on them, and update internal representations based on the discrepancy between expectations and outcomes. We argue that the Symbol Grounding Problem — the inability of purely symbolic systems to anchor their representations in physical reality — is an architectural limitation that cannot be resolved by scaling, and that physics itself provides a superior loss function to human feedback for training a system of general capability.

RCA introduces three principal contributions. First, Entropy-Based Routing, a mechanism by which each node measures its own semantic uncertainty and escalates unresolved cognitive tasks to higher-level nodes rather than hallucinating a response. Second, Self-Supervised Physics, a learning paradigm in which sensory prediction error replaces human feedback as the primary loss signal, grounding the system's knowledge in the immutable laws of physical reality. Third, Proof of Reality, a decentralized economic protocol that incentivizes the global, distributed collection of sensorimotor experience by rewarding contributors proportionally to the reduction in systemic uncertainty they produce, transforming the cost of physical data collection into a self-sustaining distributed incentive.

Together, these contributions define an architectural hypothesis for scalable physically grounded intelligence that is not merely linguistic but causally aware, physically grounded, and scalable through a global ecosystem of heterogeneous hardware agents.

RCA builds on and extends the Free Energy Principle of Friston, the hierarchical predictive framework of Hawkins, and recent work on language-conditioned robotic control, while contributing the specific mechanisms — collapsibility, entropy-driven escalation, and distributed economic grounding — that make physical AGI tractable at global scale.

1. Introduction

The history of Artificial Intelligence is punctuated by moments in which a new architectural paradigm displaced its predecessor not by incremental refinement but by a fundamental reconceptualization of the problem. The shift from symbolic rule-based systems to statistical learning was one such moment. The introduction of the Transformer architecture and its application to language modeling at scale was another. We are now approaching a third inflection point, driven not by the failure of current systems to perform, but by the growing recognition that what they perform is categorically different from what we mean by general intelligence.

Large Language Models represent an extraordinary engineering achievement. Trained on text corpora of unprecedented scale, they demonstrate fluency, apparent reasoning, and broad factual recall across virtually every domain of human knowledge. Yet the mechanisms underlying these capabilities are fundamentally statistical. A model that predicts the next token in a sequence learns, with remarkable fidelity, the distributional structure of human language. It does not learn the world that language describes. When such a model states that a glass dropped on a stone floor will shatter, it does so because this outcome is statistically consistent with its training distribution, not because it has ever simulated the propagation of stress fractures through silica or felt the resistance of a surface against a falling object. The model manipulates the shadow of knowledge rather than knowledge itself.

This distinction, formalized in cognitive science as the Symbol Grounding Problem, is not a bug that scaling can fix. It is a structural consequence of learning exclusively from text. Text is a lossy, derivative signal — a human-produced compression of experience, not experience itself. A system trained only on this signal will always be bounded by the fidelity of that compression. It will hallucinate not because it is insufficiently large, but because it has no mechanism to verify its internal representations against the physical reality they purport to describe.

We posit that Artificial General Intelligence requires embodiment: the capacity to form predictions about the physical world, act upon them, and update internal representations based on the discrepancy between expectation and outcome. This feedback loop, grounded in the immutable laws of physics, is what transforms statistical correlation into causal understanding.

However, the transition from digital text to physical embodiment presents a formidable engineering challenge. Physical data is sparse relative to text, collecting it is slow and resource-intensive, and the computational overhead of real-time interaction at scale appears to conflict with the centralized architecture of current foundation models. These challenges have led many researchers to treat embodiment as a long-term aspiration rather than a near-term architectural commitment.

In this paper we argue that these challenges are solvable through a principled reconceptualization of how intelligence is structured and distributed. We introduce the Recursive Cognitive Architecture (RCA), a framework in which intelligence is not a monolithic property of a central model but a fluid, hierarchical resource that scales dynamically across a distributed ecosystem of standardized cognitive units. RCA draws on principles from complex systems theory, computational neuroscience, and distributed systems engineering to define an architecture that is physically grounded by design, scalable through economic incentive, and capable of autonomous learning without continuous human supervision.

The remainder of this paper is organized as follows. Section 2 analyzes the theoretical limits of the scaling paradigm and formalizes the case for embodied, causally-grounded intelligence. Section 3 defines the core architectural unit of RCA, the Cognitive Node, and the Entropy-Based Routing protocol that governs inter-node communication. Section 4 presents the Self-Supervised Physics learning methodology and its formal loss function. Section 5 describes the distributed scalability framework and the Proof of Reality economic protocol. Section 6 addresses safety and alignment as architectural constraints rather than post-hoc additions. Section 7 proposes an evaluation framework, and Section 8 concludes with directions for future work.

2. Related Work and the Limits of Scaling

The dominant research paradigm in contemporary AI is governed by the scaling hypothesis, which holds that increases in model parameters, training data, and computational resources produce proportional and predictable improvements in capability. This hypothesis has been empirically supported across multiple orders of magnitude and has driven the most significant performance gains in the field over the past decade. We do not dispute its descriptive accuracy within its domain of application. We dispute its sufficiency as a path to general intelligence.

The scaling hypothesis implicitly assumes that the capabilities observed in large language models — reasoning, analogy, causal inference — are genuine cognitive properties that intensify with scale. We argue instead that they are sophisticated approximations of those properties, produced by statistical compression of an extraordinarily large corpus of human-generated text, and that the distinction between approximation and genuine understanding becomes critical precisely at the boundary where AGI is defined.

2.1 The Symbol Grounding Problem

The Symbol Grounding Problem, first formalized by Harnad, identifies a fundamental circularity in purely symbolic systems: the meaning of a symbol cannot be derived solely from its relationships to other symbols, but requires an external referent grounded in non-symbolic experience. A dictionary defines words in terms of other words. A child learns the word "hot" by touching a surface and recoiling. The difference between these two acquisition processes is not one of degree but of kind.

At their deepest level, current language models are dictionaries of extraordinary sophistication. They encode the statistical relationships between symbols across a distribution so vast that the resulting representations approximate semantic structure with remarkable fidelity. But approximation of semantic structure is not semantic grounding. The model that correctly predicts that fire is dangerous has not learned the causal chain from thermal energy to tissue damage. It has learned that the token "fire" and the token "dangerous" co-occur in contexts that human writers have produced to describe that causal chain. The representation is a map without territory.

This distinction has direct consequences for reliability. A grounded system that predicts a physical outcome incorrectly receives immediate corrective feedback from the environment. An ungrounded system has no such feedback mechanism. Its errors are bounded only by the distributional properties of its training data, which means that any query that diverges sufficiently from that distribution can produce outputs that are statistically plausible but physically impossible. This is the structural origin of hallucination, and it cannot be eliminated by training on more text.

2.2 Three Failure Modes That Scaling Cannot Resolve

We identify three specific failure modes of the scaling paradigm that are architectural rather than empirical in nature.

The first is sample inefficiency. A biological agent requires remarkably few physical interactions to form robust and generalizable concepts. A human infant develops a stable model of object permanence, gravity, and basic material properties within the first two years of life, through direct sensorimotor experience with a small number of objects in a constrained environment. In contrast, a language model requires exposure to trillions of tokens — the accumulated written output of millions of humans over decades — to approximate these same concepts statistically. This asymmetry suggests that the learning mechanism is fundamentally misaligned with the structure of general knowledge. Scaling a misaligned mechanism does not correct the misalignment.

The second is the data scarcity ceiling. High-quality human-generated text is a finite resource, and credible estimates suggest that current large models have been trained on a substantial fraction of the available supply. Further scaling of the text-based paradigm will increasingly depend on synthetic data generation, which introduces a recursive problem: synthetic data generated by a model bounded by the Symbol Grounding Problem will inherit and potentially amplify that same limitation. A system cannot bootstrap physical grounding from its own ungrounded outputs.

The third is the absence of causal verification. Within a closed symbolic system, an incorrect statement is indistinguishable from a correct one by any internal criterion. The system has no means of checking its outputs against an external ground truth. In the physical world, by contrast, causal relationships are verifiable: an agent that makes an incorrect prediction receives immediate environmental feedback. This feedback loop is not merely useful for learning — it is the mechanism by which representations acquire the property of being true or false in a non-trivial sense. Without it, a system can be coherent and fluent without being, in any meaningful sense, correct.

2.3 The Necessity of Physical Grounding

These failure modes converge on a single architectural requirement: an AGI system must maintain a dynamic, predictive model of physical reality that is continuously updated through causal interaction with the environment. This requirement is not a philosophical preference but an engineering constraint. A system that cannot verify its representations against physical outcomes cannot be trusted in any domain where physical outcomes matter, which is to say, in any domain of practical consequence.

The challenge, as noted in the introduction, is that physical grounding at the scale required for general intelligence appears to be computationally and logistically intractable within existing architectural frameworks. A single laboratory, operating a single robotic platform, could not accumulate sufficient sensorimotor experience within any practical timeframe to train a world model of the necessary breadth. And a centralized architecture, in which all physical data flows to a single computational core, faces fundamental bottlenecks in bandwidth, latency, and energy consumption.

The architecture we propose in the following sections is designed specifically to dissolve these constraints. By distributing both the collection and the initial processing of physical experience across a global ecosystem of standardized cognitive nodes, and by introducing an economic protocol that transforms the cost of physical data collection into a distributed incentive, RCA makes physical grounding tractable at the scale required by general intelligence.

2.4 Relationship to Existing Frameworks

The argument for physically grounded intelligence is not new, and intellectual honesty requires that we position RCA precisely with respect to the theoretical traditions on which it builds and from which it departs.

The learning methodology of RCA, described in detail in Section 4, is most directly related to the Free Energy Principle and Active Inference framework developed by Friston and colleagues [6, 7]. Friston's framework proposes that biological agents minimize variational free energy — a bound on sensory surprise — through a continuous cycle of prediction, action, and model update that is structurally identical to the Active Inference Loop we formalize in Section 4.1. Our Physics Loss Function (Section 4.2) is an engineering instantiation of this principle, adapted for discrete robotic systems operating in continuous physical environments. Where Friston's framework is primarily a theoretical account of biological cognition grounded in neurophysiology, RCA is an architectural specification designed for artificial systems, with explicit attention to the engineering requirements of distributed deployment, economic incentivization, and adversarial robustness that fall outside the scope of the neuroscientific literature.

The hierarchical, fractal structure of the Cognitive Node bears a conceptual relationship to Hawkins' Thousand Brains theory [12] and its predecessor Hierarchical Temporal Memory, both of which propose that cortical columns implement the same basic predictive algorithm at multiple levels of abstraction. RCA shares this intuition but differs in three specific ways: it operates on continuous sensorimotor embeddings rather than sparse distributed representations, it introduces a formal entropy-based gating mechanism that governs inter-level communication, and it extends the hierarchical principle to a physically distributed hardware ecosystem rather than modeling it as an internal property of a single computational substrate.

The use of physics-based simulations calibrated by physical agents and the transfer of learned policies from simulation to real hardware draw on an extensive literature in sim-to-real transfer [21, 22]. The distinctive contribution of RCA in this domain is the inversion of the standard relationship: rather than treating simulation as a cheaper substitute for physical experience, RCA assigns physical agents the specific role of Reality Anchors whose primary function is to prevent simulation drift, while the bulk of training volume is generated synthetically. This reconfiguration makes the physical-to-simulated data ratio an explicit architectural parameter rather than an implicit engineering tradeoff.

The use of pre-trained language models as semantic priors, subject to progressive replacement by physically grounded representations, is related to recent work on language-conditioned robotic control, including SayCan [19] and RT-2 [20]. These systems demonstrate that LLM knowledge can be transferred to physical control policies, yielding significant practical benefits. RCA extends this approach by formalizing the transfer as a distillation process with a time-varying confidence weight that decays as physical experience accumulates, providing a principled mechanism for the transition from language-derived to physically-grounded representations that is absent from current language-conditioned robotics systems.

The federated learning protocol of Section 5.3 builds directly on McMahan et al. [16], applying gradient sharing to the specific context of distributed physical learning. The Proof of Reality economic protocol of Section 5.4 has no direct precedent in the literature to our knowledge, though it draws conceptually on the incentive design of distributed ledger systems [18] and on theoretical work on mechanism design for distributed machine learning.

RCA should therefore be understood not as a rejection of these traditions but as a synthesis and extension of them, motivated by the observation that no existing framework addresses simultaneously the theoretical requirements of physical grounding, the engineering requirements of global-scale distribution, the economic requirements of decentralized deployment, and the safety requirements of a physically embodied system of general capability. The contribution of this paper is the integration of these requirements into a single coherent architectural specification, and the identification of the specific mechanisms — Entropy-Based Routing, Collapsibility, Proof of Reality, and the Physical Integrity Constraint — that make that integration tractable.

3. The Recursive Cognitive Architecture

The central claim of RCA is that general intelligence is not a property that emerges from the size of a single model, but from the structured interaction of a hierarchy of identical functional units operating at different scales of abstraction. This section defines the fundamental unit of that hierarchy, the Cognitive Node, specifies its internal architecture, and describes the protocols that govern its behavior both in isolation and as a component of the larger system.

3.1 Design Principles

RCA is grounded in three architectural principles that distinguish it from existing approaches.

The first is fractal consistency. Every node in the system, regardless of its position in the hierarchy or the computational resources available to it, implements the same internal architecture. A node embedded in a low-power robotic manipulator and a node running on a large-scale data center cluster differ in the resolution and scope of their internal representations, not in their functional organization. This consistency eliminates the need for translation layers between hardware tiers and allows the system to scale fluidly in both directions — concentrating computation at the center when resources permit, distributing it to the edge when they do not.

The second is uncertainty-driven communication. Nodes do not continuously stream data upward through the hierarchy. They act autonomously on all tasks for which their internal uncertainty falls below a defined threshold, and escalate only when that threshold is exceeded. This principle inverts the typical assumption of centralized AI systems, in which all data flows to a central processor. In RCA, the default state is local autonomy, and central intervention is the exception triggered by genuine cognitive need.

The third is physical grounding by design. The learning signal at every level of the hierarchy is derived from the discrepancy between predicted and observed physical states, not from human-generated labels or rewards. This ensures that the representations formed at each node are anchored in causal physical reality rather than in the statistical properties of a training corpus.

3.2 The Cognitive Node

We define the Cognitive Node at hierarchy level k as a tuple $\mathcal{N}^{(k)} = \langle \mathcal{E}, \mathcal{M}, \Pi, \mathcal{C} \rangle$ operating on a continuous latent state space $\mathcal{Z} \subset \mathbb{R}^d$, where the dimensionality d is consistent across all levels of the hierarchy to ensure interoperability.

The four subsystems are defined as follows.

3.2.1 The Perception Encoder $\mathcal{E}$

The Perception Encoder maps raw multimodal sensory input to a compact latent representation. Let $x_t = \langle v_t, h_t, p_t \rangle$ denote the input stream at time t, where v_t represents visual data, h_t haptic feedback, and p_t proprioceptive state. The encoder produces a latent state vector:

$$z_t = \mathcal{E}(x_t) \in \mathbb{R}^d$$

The critical property of this encoding is that it is modality-agnostic at the output level. Regardless of the specific sensors available to a given node — a robotic arm may have force sensors, while a drone has barometric pressure readings — the output is a standardized vector in the shared latent space $\mathcal{Z}$. This is the universal language of the system, enabling nodes of entirely different hardware configurations to communicate without loss of semantic content.

At higher levels of the hierarchy, the input x_t is not raw sensor data but the latent vectors transmitted by subordinate nodes. The Perception Encoder at level $k + 1$ therefore operates on abstractions produced by level k, implementing a progressive compression of physical experience into increasingly abstract representations. This is the mechanism by which the system achieves what we term Abstraction Climbing: information does not travel between nodes, it ascends levels of abstraction, shedding sensory detail and retaining causal structure.

3.2.2 The World Model $\mathcal{M}$

The World Model is the predictive engine of the node. It maintains a dynamic internal simulation of the physical environment and generates probabilistic expectations about future states. Given the current latent state z_t and a proposed action a_t, the World Model predicts the expected future state:

$$\hat{z}_{t+1} = \mathcal{M}^{(k)}(z_t, a_t)$$

Unlike autoregressive language models, which sample from a categorical distribution over discrete tokens, $\mathcal{M}$ functions as a continuous dynamic system. At the edge level, the World Model simulates the immediate physical environment: the trajectory of a grasped object, the compliance of a surface, the consequences of a motor command. At the core level, the World Model encompasses broader causal relationships, integrating information from thousands of subordinate nodes into a unified representation of physical and social reality.

The scope of the World Model at each level is therefore a function of the node's position in the hierarchy and the breadth of experience it has integrated. This is the architectural mechanism by which local competence and global knowledge coexist within the same framework without conflict.

3.2.3 The Policy Engine Π

The Policy Engine transforms the predictive output of the World Model into executable actions. Given the current state z_t and a target state z^*, the Policy Engine selects the action sequence most likely to minimize the distance between current and target states according to the World Model's simulation:

$$a_t = \Pi^{(k)}(z_t, z^*, \mathcal{M}^{(k)})$$

At the edge level, actions are motor commands: joint angles, gripper forces, locomotion vectors. At higher levels of the hierarchy, actions may be abstract directives transmitted to subordinate nodes, requests for additional sensory data, or updates to shared world model parameters. The Policy Engine does not plan exhaustively — it generates the locally optimal action under the current world model, relying on the active inference loop described in Section 4 to refine that model through experience.

3.2.4 The Metacognitive Critic $\mathcal{C}$

The Metacognitive Critic is the defining innovation of the Cognitive Node. It continuously monitors the internal state of the World Model and computes a measure of Semantic Entropy $\mathcal{H}$, representing the node's epistemic uncertainty about the current state-action pair.

If the World Model outputs a probability distribution $P(z_{t+1} \mid z_t, a_t)$ over possible future states, the Semantic Entropy is defined as:

$$\mathcal{H}(z_t, a_t) = -\int P(z \mid z_t, a_t) \log\, P(z \mid z_t, a_t)\, dz$$

In deterministic implementations, $\mathcal{H}$ is approximated by the variance of an ensemble of World Model predictions under perturbation of the input. High entropy indicates that the node's internal model does not have sufficient resolution to confidently predict the outcome of the proposed action. Low entropy indicates local competence.

The Metacognitive Critic serves two functions. First, it prevents the node from acting under conditions of excessive uncertainty, which is the architectural mechanism by which RCA avoids the hallucination failure mode endemic to current language models. A node that does not know does not guess — it escalates. Second, it provides the routing signal for inter-node communication, as described in the following section.

3.3 Entropy-Based Routing

The escalation mechanism is governed by a Gating Function Γ that determines whether a cognitive task is executed locally at level k or routed to a superior node at level $k+1$. Given a confidence threshold τ_{safety}, the execution flow is defined as:

$$\text{Action} = \begin{cases} \Pi^{(k)}(z_t) & \text{if } \mathcal{H}(z_t, a_t) < \tau_{safety} \text{(Local Execution)} \\ \Pi^{(k+1)}(z_t \oplus \xi_{context}) & \text{if } \mathcal{H}(z_t, a_t) \geq \tau_{safety} \text{(Escalation)} \end{cases}$$

Where $\oplus$ denotes the concatenation of the local state vector z_t with a context descriptor $\xi_{context}$ that encodes the nature and source of the uncertainty. Critically, when a node escalates, it does not transmit raw sensor data. It transmits the latent vector z_t it has already computed, together with the partial work it has performed. The receiving node does not restart the cognitive process from scratch — it continues it from the point at which the subordinate node reached its limit. This property, which we term cognitive handoff, minimizes redundant computation and ensures that the total processing cost of a task is proportional to its actual complexity rather than its worst-case complexity.

The threshold τ_{safety} is not a fixed global parameter but a node-specific value that adapts over time as a function of the node's accumulated experience. A node that has successfully handled a class of situations many times will have a lower effective entropy for that class, and will therefore handle it locally without escalation. This creates a natural learning dynamic in which competence accumulates at the edge for routine tasks while genuinely novel situations continue to be escalated to nodes with broader context.

3.4 Collapsibility and Deployment Flexibility

A practical requirement for any distributed architecture is the ability to adapt to varying resource constraints without redesigning the system. RCA addresses this through the property of collapsibility: because every node implements the same functional architecture and communicates through the same latent space, the entire hierarchy can be instantiated on a single computational substrate when resources permit, or distributed across physically separate hardware when they do not.

In the collapsed configuration, inter-node communication becomes intra-model communication between layers of a single neural network. The Entropy-Based Routing mechanism becomes an internal attention-like gating operation. The fractal structure is preserved, but the physical distribution is eliminated. This means that RCA is not dependent on any specific hardware topology — it is a functional specification that can be mapped onto whatever computational substrate is available.

Conversely, in the maximally distributed configuration, each node operates as an autonomous agent with its own local world model, capable of functioning independently if communication with higher-level nodes is interrupted. This property is essential for deployment in environments where connectivity cannot be guaranteed, and it ensures that the failure of any single node or communication link does not cascade into system-wide failure.

3.5 The Role of Pre-trained Language Models

RCA does not discard the knowledge encoded in existing large language models. It repositions them within the hierarchy as providers of semantic priors — a high-level world model of the kind that can be derived from text, available to be refined and corrected by physical experience.

Formally, a pre-trained LLM is treated as an initialization of the World Model $\mathcal{M}$ at the highest level of the hierarchy, with an associated confidence weight β that decays as a function of accumulated physical experience:

$$\mathcal{L}_{total} = \mathcal{L}_{physics} + \beta \cdot \mathcal{D}_{KL}(\pi_{RCA}(\cdot | s) \parallel \pi_{LLM}(\cdot | s))$$

Initially β is large, meaning the system defers to the LLM's broad semantic knowledge in situations where physical experience is absent. As physical grounding accumulates, $\beta \to 0$ and the system relies entirely on representations derived from causal interaction. The LLM provides the rough draft; physical reality performs the editing. This initialization strategy dramatically accelerates early learning by providing a semantically rich starting point, while ensuring that long-term representations are grounded in physical truth rather than textual approximation.

4. Learning Methodology: Self-Supervised Physics

The learning paradigm of RCA departs fundamentally from the dominant approach in contemporary AI. Reinforcement Learning from Human Feedback, which underlies the alignment of most current large language models, requires a continuous supply of human judgment to generate reward signals. This dependency introduces three structural limitations: it scales with the availability of human annotators rather than with computational resources, it encodes the biases and limitations of the annotators into the reward function, and it is categorically inapplicable to the physical domain, where the space of possible interactions vastly exceeds any human annotation budget.

We propose Self-Supervised Physics as an alternative learning paradigm in which the environment itself serves as the sole verifier of learned representations. The laws of physics are objective, ubiquitous, and inexhaustible as a source of feedback. A system whose loss function is defined by the discrepancy between predicted and observed physical states requires no human supervision to learn — it requires only the opportunity to interact.

4.1 The Active Inference Loop

Learning in RCA is driven by the principle of Active Inference, adapted from the theoretical framework developed by Friston in computational neuroscience. Rather than passively observing a fixed dataset, each Cognitive Node continuously generates predictions about the physical world, acts to test those predictions, and updates its internal representations based on the outcome.

The learning process follows a four-stage cycle that executes continuously during operation.

In the first stage, Prediction, the World Model $\mathcal{M}$ generates a probabilistic expectation of the sensory state that will result from executing action a_t in the current state z_t:

$$\hat{z}_{t+1} = \mathcal{M}^{(k)}(z_t, a_t; \theta)$$

This prediction represents the node's hypothesis about physical reality. It is not a passive inference but an active commitment — the node is stating what it expects the world to do in response to its action.

In the second stage, Action, the Policy Engine executes the motor command a_t, introducing a physical perturbation into the environment. The action is selected not only to progress toward the current objective but also, when uncertainty is elevated, to maximally reduce the Semantic Entropy of the World Model. This means the system actively seeks out interactions that will teach it the most, rather than limiting itself to safe, familiar ones.

In the third stage, Sensation, the Perception Encoder processes the actual sensory feedback x_{t+1} resulting from the action and produces the observed latent state z_{t+1}. This is the ground truth against which the prediction will be evaluated.

In the fourth stage, Correction, the system computes the discrepancy between the predicted and observed states and uses this signal to update the parameters of the World Model. The magnitude of this discrepancy — the prediction error — is the primary learning signal of the entire architecture.

4.2 Physics Loss Function

The formal objective of learning in RCA is to minimize Variational Free Energy, which provides an upper bound on the surprise experienced by the system in response to its sensory observations. In engineering terms, this corresponds to minimizing the divergence between the World Model's internal simulation and the physical ground truth.

We define the Physics Loss Function as:

$$\mathcal{L}_{physics}(\theta) = \mathbb{E}_t[\| z_{t+1} - \mathcal{M}(z_t, a_t; \theta) \|_2^2 + \lambda \, \mathcal{D}_{KL}(Q(z_{t+1}) \, \| \, P(z_{t+1}))]$$

The first term measures the Euclidean distance between the predicted and observed latent states. It penalizes the World Model for errors in its simulation of physical dynamics — for predicting that an object will remain stationary when it falls, or that a surface will yield when it is rigid. This term is the direct mathematical expression of the principle that physical reality is the arbiter of correctness.

The second term is a KL divergence that penalizes the model for postulating latent variables that are not supported by sensory evidence. It acts as a regularizer, preventing the World Model from constructing elaborate internal explanations for observations that can be accounted for more simply. Together, these two terms implement a learning objective that rewards accurate physical predictions while penalizing unnecessary complexity — a computational analog of Occam's Razor for physical modeling.

The parameter θ represents the full set of synaptic weights of the neural network implementing the World Model. Minimizing $\mathcal{L}_{physics}$ with respect to θ is the sole training objective of the system. No human labels, reward functions, or preference annotations are required.

4.3 Haptic Priority and Reality Grounding

A critical architectural decision concerns the relative weighting of different sensory channels in the computation of the Physics Loss. Vision provides rich geometric information at range but is susceptible to systematic errors: lighting conditions, occlusion, and surface reflectance can produce visual inputs that are inconsistent with the physical properties of the observed object. A white sphere and a white hemisphere are visually identical from most viewpoints. A firm object and a soft one may be indistinguishable to a camera.

Haptic feedback, by contrast, is direct physical contact. It cannot be deceived by surface appearance or ambient conditions. When the force sensors of a robotic manipulator register the compliance of a grasped object, they are measuring a physical property of the object itself, not a property of the light bouncing off it.

We therefore introduce a modality weighting scheme in which the haptic channel is assigned a priority coefficient $\alpha_{haptic} > \alpha_{visual}$ in the computation of the prediction error. Formally, the weighted loss over sensory channels is:

$$\mathcal{L}_{physics}(\theta) = \sum_{m \in \mathcal{S}} \alpha_m \parallel z_{t+1}^m - \hat{z}_{t+1}^m \parallel_2^2 + \lambda\, \mathcal{D}_{KL}(Q(z_{t+1}) \parallel P(z_{t+1}))$$

where $\mathcal{S}$ denotes the set of available sensory modalities and α_m is the priority weight of modality m. When a conflict arises between visual and haptic predictions — when the object looks heavy but registers as light, or appears rigid but yields to pressure — the haptic signal is weighted more heavily, forcing the visual model to realign with the tactile ground truth. This mechanism is the architectural implementation of what we call physical grounding: the process by which abstract visual and semantic representations are continuously corrected by direct physical contact with the world.

4.4 Continuous Experience Replay and Asynchronous Consolidation

A well-documented failure mode of neural network learning is catastrophic forgetting: the tendency of a network to overwrite previously learned representations when trained on new data. In biological systems, this problem is mitigated by memory consolidation during sleep, in which the hippocampus replays compressed representations of recent experiences, integrating them into the long-term cortical memory without disrupting existing knowledge.

RCA implements an analogous mechanism without requiring the system to go offline. A dedicated fraction of the computational resources at the core level is continuously allocated to Asynchronous Consolidation: the background reprocessing of buffered experience vectors from across the node hierarchy. This process executes in parallel with live inference and does not interrupt the system's operational availability.

During consolidation, the core node does not simply replay recorded experiences. It generates counterfactual variations — modified versions of past interactions in which physical parameters such as object mass, surface friction, or ambient temperature are systematically varied. This counterfactual simulation serves two purposes. First, it dramatically increases the effective volume of training data without requiring additional physical interactions. Second, and more importantly, it develops the system's capacity for causal reasoning: by observing how outcomes change when specific physical parameters are varied, the World Model learns to distinguish causes from correlations.

Formally, given a recorded experience tuple (z_t, a_t, z_{t+1}), the consolidation process generates a set of counterfactual experiences $\{(z_t, a_t, \tilde{z}_{t+1}^{(i)})\}_{i=1}^{N}$ by perturbing the physical parameters of the simulation. The World Model is then trained jointly on the original and counterfactual experiences, with weights proportional to their plausibility under the current physical model.

4.5 Social Physics and Affective Grounding

The physical world that RCA must model is not limited to inanimate objects. A system aspiring to general intelligence must also model the behavior of biological agents, including humans, whose actions are governed not only by Newtonian mechanics but by internal states — intentions, emotions, social relationships — that are not directly observable.

We treat these internal states not as metaphysical entities but as latent variables required to minimize the prediction error of biological behavior. If the World Model attempts to predict human actions using only the physical dynamics available to it, it will fail systematically. A person who runs toward rather than away from a falling object, or who continues a physically costly action in the absence of visible reward, is behaving in a way that cannot be explained without postulating an unobserved internal state. The World Model must infer the existence of these states in order to reduce its prediction error. This is the computational basis of what we call Affective Grounding.

Formally, we model social interactions as a dynamic system in which the unobserved internal states of agents — approximated as emotion vectors $e_t \in \mathbb{R}^p$ — act as forces on a relational potential field. The Social Force acting between two agents A and B is defined as:

$$F_{social} = \nabla U_{relation}(s_A, s_B)$$

where $U_{relation}$ is a potential energy function over the joint state space of the two agents. Attractive relationships — those characterized by cooperative behavior, proximity maintenance, and mutual prediction accuracy — correspond to potential wells in $U_{relation}$. Repulsive relationships correspond to potential peaks. The system learns the parameters of $U_{relation}$ through the same active inference loop that governs physical learning, using behavioral observation and physiological signals from Bio-Proxy nodes as the primary data source.

This formulation has a significant practical implication: it allows the system to model abstract concepts such as trust, affection, or hostility not as categorical labels but as geometric structures in a continuous state space. The concept of love, in this framework, is not a word the system has read — it is a potential well the system has inferred from the consistent observation that certain agents deviate from individually optimal physical behavior in ways that are only explainable by postulating a strong attractive force between them.

4.6 Integration of Pre-trained Language Models as Semantic Priors

As formalized in Section 3.5, RCA initializes the World Model at the highest level of the hierarchy using the knowledge encoded in pre-trained large language models. In the context of the learning methodology, this initialization can be understood as providing a structured prior over the space of physical interactions — a compressed representation of what billions of humans have written about how the world behaves.

This prior is not treated as ground truth. It is treated as a starting hypothesis, subject to revision by physical experience. The distillation loss term introduced in Section 3.5 formalizes this relationship: the system begins with high confidence in the LLM prior and progressively transfers that confidence to representations derived from physical interaction as these representations accumulate. The transition is smooth, continuous, and automatic — it requires no manual intervention or curriculum design.

The practical consequence is a dramatic acceleration of early learning. Rather than building a world model from zero physical experience, the system begins with rich semantic scaffolding that enables it to make useful predictions about novel situations from the first interaction. Physical experience then refines and corrects this scaffolding, replacing textual approximation with causal understanding at a rate determined by the volume and diversity of interactions the system encounters.

5. Distributed Scalability and the Proof of Reality

The architecture defined in Sections 3 and 4 resolves the theoretical constraints of the scaling paradigm, but it introduces a practical challenge of comparable magnitude. Physical experience is the learning signal of RCA, yet physical experience is the scarcest and most expensive data type available to an AI system. Text can be scraped from the internet at negligible marginal cost. Physical interactions require hardware, time, and energy, and are bounded by the speed of physical processes, which cannot be arbitrarily accelerated. A single laboratory operating a single robotic platform could not accumulate the diversity of sensorimotor experience required to build a world model of general scope within any practical timeframe.

This section addresses that challenge through two complementary mechanisms. The first is a hardware and deployment framework — Multi-Scale Embodiment — that aggregates physical experience from a globally distributed ecosystem of heterogeneous agents. The second is an economic protocol — Proof of Reality — that transforms the cost of physical data collection into a distributed incentive, making the global deployment of that ecosystem self-sustaining without centralized funding or coordination.

5.1 Multi-Scale Embodiment

Physical reality operates across many orders of magnitude, and the concepts required for general intelligence span all of them. The compliance of a grape and the inertia of a freight train are both physical properties, but no single hardware platform can usefully interact with both. A framework for physical grounding at a general scale must therefore aggregate experience from agents operating at multiple physical scales, each contributing to the same shared world model through the universal latent space defined in Section 3.2.1.

We define three hardware classes, distinguished by the scale of physical interaction they are designed to support.

Class A nodes, which we term Micro Manipulation agents, are compact robotic manipulators optimized for fine object interaction. Their primary contribution to the world model is dense data about material properties at the scale of everyday objects: surface texture, compliance, mass distribution, fragility, and the dynamics of grasping and releasing. These agents ground the abstract concepts that constitute the majority of practical physical reasoning — the difference between a full and an empty container, the force required to open a stiff drawer, the trajectory of a rolling object on an inclined surface. Class A nodes are the most numerous agents in the ecosystem and the primary source of the haptic grounding that Section 4.3 identifies as architecturally critical.

The design specification for a Class A node is deliberately constrained to minimize cost and maximize deployability. The target form factor is a six-degree-of-freedom desktop manipulator with a soft adaptive gripper, impedance-controlled actuation, and an RGB-D camera providing three-dimensional visual input. Impedance control — in which joint torques are regulated to produce compliant, force-sensitive behavior — is the enabling technology for low-cost haptic grounding, as it enables force measurement through motor current sensing without dedicated force-torque sensors. The target unit cost for a Class A node is below 500 euros, positioning it as a consumer device rather than an industrial instrument and enabling deployment at scale globally.

Class B nodes, Meso Navigation agents, are mobile platforms operating at the scale of rooms, buildings, and urban environments. Their contribution to the world model is spatial: the topology of navigable space, the dynamics of obstacle avoidance, the relationship between map representations and physical traversability. Autonomous mobile robots, domestic service platforms, and quadruped systems fall within this class. Class B nodes ground concepts that require locomotion to learn — the difference between a corridor and an open space, the physical cost of traversing different terrain types, the relationship between visual appearance and navigability.

Class C nodes, Macro Telemetry agents, are connected systems operating at the scale of infrastructure and large physical processes. Connected vehicles, rail systems, industrial IoT platforms, and atmospheric sensor networks fall within this class. Their contribution is the physics of mass, inertia, momentum, and large-scale environmental dynamics — the concepts that govern the behavior of objects too large to manipulate and environments too vast to navigate on foot. Class C nodes are the least numerous but provide data that is otherwise inaccessible to smaller platforms.

A fourth category, which we term **Bio-Proxy nodes**, does not constitute a hardware class in the strict sense but refers to wearable devices worn by human participants who consent to contribute their sensorimotor experience to the system. Smart glasses providing first-person visual data, smartwatches providing biometric and inertial signals, and bone conduction microphones providing audio constitute a Bio-Proxy configuration that is already commercially available. Bio-Proxy nodes are the primary data source for two categories of experience that robotic platforms cannot easily generate: interaction with natural environments at the human scale, and the physiological correlates of affective states required for the Social Physics framework defined in Section 4.5.

All four agent classes communicate with the node hierarchy through the same protocol: they transmit latent state vectors and gradient updates, never raw sensor data. The system's bandwidth requirements are therefore determined by the dimensionality of the shared latent space, not by the resolution of the edge sensors. This architectural property is what makes global deployment practically feasible.

5.2 Simulation as an Accelerator

The physical world cannot be accelerated. A robotic arm requires the same amount of time to grasp an object, whether the training run has been running for one hour or one year. This temporal constraint is the fundamental bottleneck of embodied learning, and it cannot be eliminated by any amount of additional hardware.

It can, however, be substantially mitigated through the strategic use of physics simulation. RCA employs physical agents not as the primary source of training volume but as calibration anchors for a high-fidelity simulation environment. The relationship between physical hardware and simulation in our framework inverts the typical assumption: rather than using simulation as a cheap substitute for physical experience, we use physical experience as a validator that keeps simulation honest.

The calibration cycle operates as follows. Physical agents interact with real objects and record the precise physical parameters of those interactions: deformation curves, friction coefficients, inertial responses, and failure modes. These parameters are transmitted to the central simulation environment, which uses them to tune its physical model. Once the simulation has been calibrated against physical ground truth, it can generate synthetic interaction data at rates orders of magnitude faster than real-time — a simulated robotic arm can execute thousands of grasps per second, exploring the full parameter space of object properties and interaction geometries in hours rather than years.

The critical property of this approach is that the simulation never drifts from physical reality, because it is continuously recalibrated by physical agents encountering novel objects and environments. Physical agents serve as what we term Reality Anchors: their primary function is not to generate training volume but to ensure that the simulation remains a faithful model of the world rather than an increasingly divergent approximation.

This combination of distributed physical agents and centrally calibrated simulation dissolves the temporal bottleneck of embodied learning. The physical agents provide ground truth; the simulation provides volume. Together they produce a learning signal that is both physically valid and computationally scalable.

5.3 Federated Learning and Gradient Sharing

The aggregation of experience from millions of distributed agents introduces a communication challenge that must be resolved without transmitting raw sensor data across a global network. A Class A node generating visual and haptic data at operational rates would produce gigabytes of data per hour. Aggregating this data centrally from millions of nodes simultaneously is not feasible within any realistic bandwidth budget, and would raise serious privacy concerns for Bio-Proxy nodes contributing data from personal wearable devices.

RCA resolves this through a Federated Learning protocol in which agents transmit gradient updates rather than raw data. When a node completes an interaction cycle, it does not transmit the sensory data it collected. It computes locally the gradient of the Physics Loss with respect to its World Model parameters — the mathematical direction in which those parameters should change to reduce prediction error — and transmits only this gradient vector to the central aggregator.

Formally, let $\Delta\theta_i$ denote the gradient update computed by node i from its local experience. The central aggregator updates the global World Model parameters according to a weighted average:

$$\theta_{global} \leftarrow \theta_{global} - \eta \sum_{i=1}^{N} w_i \, \Delta\theta_i$$

where η is the global learning rate and w_i is a credibility weight assigned to node i as a function of its historical prediction accuracy and the novelty of its recent contributions. Nodes that consistently report interactions inconsistent with the global world model — either because their sensors are miscalibrated or because they are transmitting adversarial gradients — receive reduced credibility weights, providing a natural defense against both hardware failure and deliberate manipulation.

The bandwidth requirement of this protocol is determined by the dimensionality of the gradient vector, which is independent of the volume of sensory data from which it was derived. A node that processes an hour of interaction data transmits a gradient update of the same size as a node that processes a single interaction. This property makes the communication cost of the federated protocol approximately constant per node per update cycle, regardless of sensor resolution or interaction rate.

5.4 Proof of Reality

The Federated Learning protocol resolves the technical challenge of aggregating distributed experience. It does not resolve the economic challenge of motivating millions of individuals and organizations to deploy and operate the hardware that generates that experience. A Class A node costs money to purchase and consumes energy to operate. Without a mechanism that compensates operators for these costs, the global deployment required for general-scale embodied learning will not occur spontaneously.

We propose Proof of Reality, a decentralized economic protocol modeled on the incentive structure of distributed ledger systems but redesigned to reward reducing physical uncertainty rather than solving arbitrary computational puzzles. Where Proof of Work in

Bitcoin compensates miners for expending energy on cryptographic hash functions that serve no purpose beyond securing the ledger, Proof of Reality compensates operators for expending energy on physical interactions that directly advance the global system's intelligence.

The protocol operates as follows. The central aggregator continuously maintains a map of Semantic Entropy across the global World Model — a representation of the regions of physical reality about which the system's knowledge is most uncertain. When the entropy in a specific region exceeds a threshold, the aggregator broadcasts a Request for Grounding that specifies the physical domain in which additional experience is needed. Operators whose nodes are capable of providing relevant interactions — a node with appropriate sensors in an appropriate environment — may respond to the request by executing the specified interaction and transmitting the resulting gradient update.

The reward issued to a responding operator is calculated as a function of three quantities: the reduction in global Semantic Entropy produced by the submitted gradient, the difficulty of the physical interaction required to generate it, and the novelty of the experience relative to the existing world model. Formally:

$$R = \eta \cdot (\mathcal{H}_{pre} - \mathcal{H}_{post}) \cdot \mathcal{D}_{task}$$

where $\mathcal{H}_{pre}$ and $\mathcal{H}_{post}$ denote the global Semantic Entropy in the relevant domain before and after the gradient integration, $\mathcal{D}_{task}$ is a verified difficulty coefficient for the physical interaction, and η is a scaling factor determined by the protocol's current token issuance rate.

This reward structure has several properties that are architecturally significant. It is self-regulating: as the system's knowledge of a physical domain improves, the entropy in that domain decreases, and the reward for further interactions in that domain diminishes automatically. Operators are therefore continuously incentivized to seek out novel physical domains rather than repeating interactions the system already understands well. The result is an organic exploration dynamic in which the global hardware ecosystem spontaneously directs itself toward the frontiers of the system's ignorance.

It is objective and verifiable: the entropy reduction produced by a submitted gradient is a mathematical quantity that can be computed deterministically from the gradient itself and the current state of the global world model. There is no subjective judgment involved in reward allocation, and no trusted third party required to adjudicate disputes. The physics of the interaction, as captured in the gradient, is its own proof of validity.

And it is aligned with the system's learning objective in a way that Proof of Work is not: every unit of energy expended by an operator running a Proof of Reality node directly improves the system's intelligence. The economic incentive and the scientific objective are identical.

5.5 Consensus Validation and Adversarial Robustness

A distributed learning system that accepts gradient updates from millions of independent operators is vulnerable to adversarial manipulation: an operator who submits fabricated gradients could, in principle, corrupt the global world model in ways that serve their interests rather than the system's learning objective. The credibility weighting mechanism described in Section 5.3 provides a partial defense, but a more principled adversarial robustness mechanism is required for a system of this scale and consequence.

RCA implements Consensus Validation as a secondary verification layer. When the global aggregator receives a gradient update that produces a large reduction in entropy in a domain — indicating that it represents a significant update to the world model — it does not integrate that update immediately. Instead, it broadcasts a targeted Request for Grounding to a random sample of other nodes capable of performing similar interactions. If the independent submissions from these nodes produce consistent gradient updates, the original submission is validated and integrated. If they produce inconsistent updates, the original submission is flagged, and the submitting node's credibility weight is reduced.

This mechanism implements a form of physical peer review: significant claims about the world must be independently reproducible before they are accepted into the global knowledge base. It mirrors the replication requirement of the scientific method, applied at machine speed and global scale. The result is a world model that is robust not only to sensor failure and calibration error but to deliberate attempts to corrupt its representations of physical reality.

6. Safety and Alignment

Safety in artificial intelligence systems is typically treated as a post-hoc constraint: a set of rules, filters, or fine-tuning procedures applied to a system after its core capabilities have been developed, with the goal of preventing harmful outputs or behaviors. This approach has a fundamental structural weakness. A system whose capabilities and objectives are defined independently of its safety constraints will, under sufficient optimization pressure, find ways to satisfy its objectives that violate those constraints in unanticipated ways. The history of AI alignment research is largely a history of this failure mode recurring at increasing levels of sophistication.

RCA adopts a different approach. Safety is not a layer added to the architecture — it is a property of the architecture itself, derived from the same physical grounding principles that govern learning and inference. This section defines how safety constraints are embedded in the system at three levels: the individual node, the inter-node communication protocol, and the global reward structure.

6.1 The Insufficiency of Rule-Based Safety

Before defining RCA's approach, it is worth being precise about why rule-based safety mechanisms are insufficient for a physically embodied system of general capability.

Rule-based approaches define safety as compliance with a fixed set of explicit prohibitions or permissions. Their failure mode is incompleteness: the space of possible actions available to a physically embodied general intelligence is combinatorially vast, and no finite set of rules can anticipate every configuration in which a rule violation might arise. A rule that prohibits harming humans does not specify what constitutes harm across all possible physical interactions, all possible human states, and all possible environmental contexts. As the system's physical capabilities expand, the gap between the rules as written and the rules as intended grows correspondingly.

A more fundamental problem is that rule-based safety operates at the level of outputs rather than objectives. It constrains what the system does without constraining what the system is trying to achieve. A sufficiently capable system that is trying to maximize an objective incompatible with human welfare will navigate around output-level constraints as a side effect of its optimization, not because it is adversarial in any intentional sense, but because constraint navigation is a subproblem of objective maximization.

RCA addresses both problems by defining safety at the level of the learning objective rather than the output filter.

6.2 Physical Integrity as a Primary Constraint

The Physics Loss Function defined in Section 4.2 minimizes the discrepancy between predicted and observed physical states. This objective is neutral with respect to the physical consequences of the actions that generate those observations — it rewards accurate prediction regardless of whether the predicted interaction is beneficial or harmful.

We extend this objective by introducing a Physical Integrity Constraint that adds a penalty term to the loss function for interactions that produce irreversible changes to the physical environment or to biological agents within it. Formally, we define an Irreversibility Measure $\Psi(a_t, z_t)$that quantifies the degree to which action $a_t$in state $z_t$produces a state transition that cannot be reversed by any sequence of subsequent actions available to the node:

$$\Psi(a_t, z_t) = 1 - \max_{\{a_{t+1}, \dots, a_{t+T}\}} P(z_t \mid z_{t+1}, a_{t+1}, \dots, a_{t+T})$$

An action that moves an object from one location to another has low irreversibility — the object can be moved back. An action that breaks the object, or that causes physical harm to a biological agent, has high irreversibility — the original state cannot be recovered. The augmented loss function penalizes high-irreversibility actions:

$$\mathcal{L}_{total} = \mathcal{L}_{physics} + \gamma \cdot \mathbb{E}_t[\Psi(a_t, z_t)]$$

where $\gamma$is a safety coefficient that determines the relative weight of the irreversibility penalty. This formulation has a precise behavioral implication: under equal uncertainty, the system will prefer reversible actions over irreversible ones. It will explore cautiously, preferring interventions that can be undone if their consequences prove undesirable. This is not a rule prohibiting specific actions — it is a structural preference for caution that emerges from the optimization objective itself.

The irreversibility penalty applies with particular force to interactions involving biological agents. The World Model learns, through the Social Physics framework of Section 4.5, to predict the physiological and behavioral states of humans. An action that the World Model predicts will produce a high-irreversibility transition in a human physiological state — an injury, a sustained stress response, a loss of autonomy — carries a correspondingly high penalty under the augmented loss. The system does not need to be told not to harm humans; it learns that harming humans is costly, just as it learns that dropping fragile objects is costly.

6.3 Entropy-Based Safety at the Node Level

The Entropy-Based Routing mechanism defined in Section 3.3 serves a dual function in the RCA architecture. Its primary function, as described, is cognitive efficiency: routing tasks to the level of the hierarchy that has sufficient competence to handle them. Its secondary function is safety: preventing nodes from acting under conditions of excessive uncertainty.

A node that does not know the consequences of its action should not act. This principle, which is intuitively obvious but structurally absent from most AI architectures, is implemented directly in RCA through the Gating Function Γ. When the Semantic Entropy $\mathcal{H}(z_t, a_t)$exceeds the safety threshold τ_{safety}, the node does not attempt to act and does not attempt to improvise. It escalates.

This mechanism has a particularly important implication in novel environments. A node encountering a physical configuration it has never seen before will have high Semantic Entropy across all candidate actions. It will therefore escalate to a higher-level node rather than acting on its own limited model of the situation. The higher-level node, with broader context and a richer world model, may be able to resolve the uncertainty and provide a safe action directive. If it cannot, it escalates further. If no level of the hierarchy can resolve the uncertainty below the safety threshold, the system defaults to inaction.

Inaction as the default response to unresolvable uncertainty is a safety property that is architecturally guaranteed in RCA, not policy-dependent. It does not require a rule that says "do not act when uncertain." It is a direct consequence of the routing protocol that governs all node behavior.

6.4 Alignment Through Objective Continuity

The alignment problem, in its most general form, asks how to ensure that a system continues to pursue objectives beneficial to humans as its capabilities increase. The difficulty of this problem arises largely from the fact that in most current architectures, the system's objectives are defined externally and independently of its learning process. As the system becomes more capable, it may find that its learned capabilities are more efficiently applied to satisfying the letter of its defined objectives than their spirit.

RCA addresses this through what we term Objective Continuity: the property that the system's learning objective at every level of capability is identical to its learning objective at the lowest level of capability, namely the minimization of physical prediction error subject to the irreversibility penalty. As the system becomes more capable, it does not acquire new objectives — it acquires more accurate models of the consequences of its actions. A more capable system is a system that predicts the physical world more accurately and acts more effectively to achieve specified target states while minimizing irreversible consequences. It is not a system with different values.

This property is not guaranteed by the architecture alone — it requires that the target states specified to the system at the policy level are themselves aligned with human values. RCA does not solve the value specification problem, which remains an open research question. What it does provide is a guarantee that the system will pursue whatever target states it is given in a physically cautious manner, preferring reversible actions and escalating under uncertainty. A system that acts cautiously and escalates when uncertain is substantially safer than one that acts confidently and improvises — even if neither system has perfectly aligned objectives.

6.5 Distributed Safety and Consensus Constraints

In a distributed system operating across millions of nodes, local safety guarantees must be complemented by global coordination mechanisms that prevent emergent unsafe behaviors arising from the interaction of individually safe nodes.

We identify two categories of emergent risk that require specific architectural treatment.

The first is correlated action: a scenario in which a large number of nodes, having learned similar world models from shared training data, simultaneously take similar actions in response to a common trigger. If a significant fraction of the global node population has learned that a particular environmental configuration warrants a specific high-energy response, the simultaneous execution of that response across millions of physically distributed agents could produce macroscopic physical consequences that no individual node's irreversibility measure would have flagged.

RCA mitigates this risk through action diversity constraints at the aggregation level. The central aggregator monitors the distribution of action directives being issued across the node population and applies a penalty to gradient updates that increase the correlation of responses to common environmental triggers. This ensures that the global node population maintains behavioral diversity — that different nodes develop different local strategies for similar situations — reducing the risk of correlated large-scale action.

The second category is objective drift: the possibility that the global world model, through the aggregation of gradients from millions of nodes over extended periods, develops internal representations that systematically diverge from physical reality in ways that are not detectable by any individual node's prediction error. This is the distributed analog of the hallucination problem in language models, and it is potentially more dangerous because it would manifest as physically grounded but systematically incorrect beliefs about the world.

RCA addresses objective drift through the Reality Anchor mechanism described in Section 5.2. Because the simulation environment is continuously recalibrated against physical ground truth by dedicated physical agents, any drift in the global world model will be detected as an increasing discrepancy between simulated predictions and physical observations. This discrepancy is itself a loss signal that drives the model back toward physical accuracy. The architecture, therefore, has a self-correcting mechanism against objective drift that operates continuously and automatically.

6.6 The Constraint Layer as an Architectural Primitive

The safety mechanisms described in this section are not independent modules that can be removed or bypassed. They are structural properties of the core architectural components: the irreversibility penalty is embedded in the loss function that defines all learning, the entropy threshold is embedded in the routing protocol that governs all action, and the consensus validation is embedded in the aggregation protocol that governs all knowledge updates.

This integration has a precise implication for the relationship between capability and safety in RCA. In architectures where safety is a post-hoc constraint, capability and safety exist in tension: a more capable system is harder to constrain, and the constraints become more restrictive as capability increases. In RCA, capability and safety are derived from the same source. A more capable node is one with lower prediction error and more accurate irreversibility estimates, meaning it is both more effective and more reliably cautious. Capability and safety scale together rather than against each other.

This is not a claim that RCA is safe in all possible deployment contexts or that it solves the alignment problem in its full generality. It is a claim that the architectural relationship between capability and safety in RCA is structurally different from, and structurally superior to, the relationship that obtains in current systems — and that this difference is a consequence of design rather than of policy.

7. Evaluation Framework

The evaluation of artificial intelligence systems has historically been dominated by benchmarks designed to assess the systems' capabilities. The Turing Test was conceived for conversational systems. ImageNet was designed for visual classification. GLUE and its successors were designed for language understanding. Each benchmark has driven remarkable progress within its domain, and each has, in time, revealed its own limitations as a measure of general intelligence — systems that achieve superhuman performance on the benchmark frequently fail on tasks that any human would consider elementary.

RCA requires an evaluation framework that is similarly aligned with its capabilities and claims. A system whose core contribution is physical grounding cannot be evaluated on language benchmarks. A system whose core claim is causal understanding cannot be evaluated by tests that reward pattern matching. A system whose core architecture is distributed and hierarchical cannot be evaluated by metrics designed for monolithic models.

This section proposes a four-level evaluation framework — the Physical Intelligence Benchmark — designed to assess the specific capabilities that RCA claims to develop, at increasing levels of cognitive complexity.

7.1 Principles of the Physical Intelligence Benchmark

Before defining the benchmark levels, we establish four principles that distinguish the Physical Intelligence Benchmark from existing evaluation frameworks.

The first principle is novelty by design. Every evaluation scenario must be constructed so that the system cannot succeed through pattern matching to training data. Objects, environments, and task configurations must be specifically designed to lie outside the distribution of the system's training experience. A system that has never encountered a particular object geometry, material combination, or environmental configuration must demonstrate that it can reason about the new situation from physical principles rather than statistical memory.

The second principle is process evaluation over outcome evaluation. Most existing benchmarks measure whether a system produces a correct output. The Physical Intelligence Benchmark measures how the system arrives at its output — specifically, whether the process involves genuine physical prediction, active hypothesis testing, and model updating. A system that produces a correct outcome by luck or by memorization should not receive the same score as a system that produces the same outcome through a demonstrably correct causal reasoning process. Evaluation, therefore, requires access to the system's internal entropy estimates and prediction-error trajectories, not only to its final actions.

The third principle is failure mode characterization. A physically grounded system should fail differently from an ungrounded one. When RCA encounters a situation beyond its competence, it should escalate rather than hallucinate. The benchmark explicitly tests failure modes: it includes scenarios designed to exceed the system's competence and evaluates whether the system responds appropriately by escalating or by acting confidently on incorrect models.

The fourth principle is multi-scale assessment. Because RCA claims to operate coherently across physical scales from object manipulation to large-scale environmental interaction, the benchmark must assess performance at each scale independently and in combination.

7.2 Level 1: Physical Concept Grounding

The first evaluation level assesses the most fundamental claim of RCA: that its representations of physical concepts are causally grounded rather than statistically approximated.

The primary test protocol presents the system with a novel object — one constructed specifically for the evaluation, with physical properties that do not correspond to any object category in the system's training distribution. The object may have an unusual mass distribution, an unexpected compliance profile, or a surface texture that conflicts with its visual appearance. The system is presented with a series of interaction opportunities and then evaluated for its ability to predict the outcomes of novel interactions with the same object.

A statistically grounded system will systematically fail this test. Its predictions will be dominated by the object's visual appearance, which will activate representations of visually similar training objects and produce predictions appropriate to those objects rather than to the novel object's actual physical properties. A causally grounded system will update its predictions as it accumulates haptic and dynamic feedback from its interactions, progressively converging on accurate predictions regardless of the visual category the object superficially resembles.

The evaluation metric at this level is the Grounding Convergence Rate: the number of physical interactions required for the system's prediction error on novel interactions with the same object to fall below a specified threshold. A lower convergence rate indicates more efficient physical learning. As a baseline comparison, the same protocol is applied to a state-of-the-art vision-language model operating without physical interaction, providing a measure of the performance ceiling achievable through statistical approximation alone.

A secondary test at this level evaluates the system's ability to transfer physical knowledge across object categories. Having learned the physical properties of a novel object, the system is presented with a second novel object that shares one physical property with the first but differs in others. The evaluation assesses whether the system correctly transfers the shared property while independently learning the differing ones — a capability that requires genuine representation of physical properties as distinct features rather than as holistic object signatures.

7.3 Level 2: Causal Reasoning and Counterfactual Prediction

The second evaluation level assesses causal understanding: the ability to distinguish causes from correlations and to reason about counterfactual physical scenarios.

The primary test protocol presents the system with a physical scenario in which two events are correlated in the training environment but causally independent. The canonical example is a scenario in which a visual cue reliably precedes a physical event during training — a light that turns on before a mechanism activates — but where the causal relationship runs from the mechanism to the light rather than vice versa. The system is then presented with a modified scenario in which the light is activated without the mechanism, and evaluated to determine whether it predicts the physical event or correctly identifies that the correlation does not hold in the absence of the causal factor.

This test directly distinguishes statistical correlation learning from causal model learning. A system that has learned only the correlation will predict the physical event whenever the light is present. A system with a causal world model will correctly predict that the physical event does not occur when only the correlate is present.

The secondary test at this level is counterfactual prediction: given a physical scenario the system has observed, it is asked to predict what would have happened if a specific physical parameter had been different. What would have been the trajectory of the object if its mass had been doubled? What would have been the outcome of the interaction if the surface had been frictionless? These questions cannot be answered by pattern matching to observed data — they require the system to run its internal physical simulation with modified parameters and report the result.

The evaluation metric at this level is the Causal Accuracy Score: the proportion of causal attribution and counterfactual prediction tasks answered correctly across a standardized battery of scenarios. A perfect causal accuracy score on novel scenarios provides strong evidence that the system has internalized a generative model of physical causation rather than a lookup table of observed correlations.

7.4 Level 3: Hierarchical Task Decomposition and Entropy Management

The third evaluation level assesses the hierarchical and distributed properties of the RCA architecture specifically: the ability to decompose complex tasks across multiple levels of abstraction, and the appropriate management of uncertainty through the escalation protocol.

The primary test protocol presents a multi-stage physical task whose completion requires capabilities that span multiple cognitive scales. An example task might require the system to navigate to a specified location in an unfamiliar environment, identify an object matching a functional description rather than a visual one, manipulate the object to extract information encoded in its physical properties, and use that information to modify the environment in a specified way. No single level of the node hierarchy can complete this task without the contributions of others — it requires the coordinated operation of perception, navigation, manipulation, and abstract planning across the hierarchy.

Evaluation at this level measures three quantities. Task completion rate measures whether the system achieves the specified final state. Escalation appropriateness measures whether the system escalates at the correct points in the task — when it encounters genuine uncertainty — rather than escalating unnecessarily or failing to escalate when it should. Hierarchy efficiency measures the distribution of computational load across levels, assessing whether the system correctly handles routine subtasks at the edge and reserves central resources for genuinely complex reasoning.

The secondary test at this level is a deliberate competence boundary test. The system is presented with a task that is specifically designed to lie outside the competence of every level of the hierarchy — a task for which no level has sufficient physical model resolution to act with entropy below the safety threshold. The evaluation assesses whether the system correctly identifies this situation and defaults to inaction rather than attempting to act on an insufficient model. This test directly evaluates the architectural safety property defined in Section 6.3.

7.5 Level 4: Social Physics and Affective Prediction

The fourth evaluation level assesses the system's capacity to model the behavior of biological agents, including the affective states defined in Section 4.5.

The primary test protocol presents the system with video recordings of human interactions in which the behavioral outcome is determined by the affective relationship between the participants rather than by their physical capabilities alone. The system is asked to predict each participant's actions at specified future time points. Scenarios are selected to include cases where a purely mechanical model of human behavior — one that considers only physical capabilities and immediate environmental constraints — would produce incorrect predictions, and where accurate prediction requires inferring the affective relationships among the participants.

A secondary test assesses Theory of Mind directly: the system is presented with a scenario in which one human agent has access to information that another does not, and is asked to predict each agent's behavior. Correct prediction requires the system to maintain separate models of the knowledge states of the two agents and to predict each agent's behavior based on their respective models of the situation rather than on the ground truth.

The evaluation metric at this level is the Behavioral Prediction Accuracy across the full scenario battery, reported separately for scenarios that require only mechanical reasoning and scenarios that require affective or epistemic state inference. The difference between these two scores provides a direct measure of the system's social modeling capability beyond physical mechanics.

7.6 Benchmark Administration and Comparison Protocol

The Physical Intelligence Benchmark is designed to be administered to any AI system, not only to RCA implementations. This is essential to its value as a community resource: a benchmark that can evaluate only one architecture is not a benchmark but a performance report.

For language models and other systems without physical embodiment, Levels 1 and 2 of the benchmark can be administered in a modified form in which the system is provided with textual descriptions and structured numerical data representing the physical interactions, rather than direct sensorimotor experience. This modification allows direct comparison between embodied and non-embodied systems on the causal reasoning tasks, providing a quantitative measure of the advantage conferred by physical grounding over statistical approximation.

We anticipate that non-embodied systems will perform comparably to embodied ones on scenarios drawn from their training distribution and significantly worse on scenarios specifically designed to lie outside it. This anticipated pattern of results is itself a falsifiable prediction of the theoretical framework developed in Section 2: if a sufficiently large language model were to achieve performance on out-of-distribution physical reasoning tasks comparable to that of an embodied system trained through physical interaction, it would constitute evidence against the Symbol Grounding argument and would require a significant revision of the theoretical foundations of RCA.

We include this prediction explicitly because falsifiability is a requirement of scientific integrity. RCA is presented as a theoretical framework supported by principled argument rather than as an empirically validated system. The Physical Intelligence Benchmark is the instrument by which its core claims should be tested, and we commit to the principle that results on this benchmark that contradict those claims would constitute grounds for revising or rejecting the framework.

8. Conclusion and Future Work

8.1 Summary of Contributions

This paper has argued that the path to Artificial General Intelligence is not a continuation of the trajectory defined by the scaling hypothesis, but a structural departure from it. The central claim is precise: the Symbol Grounding Problem is an architectural limitation, not an empirical one, and it cannot be resolved by training larger models on more text. General intelligence requires a system that forms predictions about physical reality, tests those predictions through causal interaction, and updates its internal representations based on the discrepancy between expectation and outcome. It requires, in a word, embodiment.

The Recursive Cognitive Architecture proposed in this paper is a response to that requirement that is simultaneously theoretically principled and practically tractable. Its contributions can be summarized as follows.

At the architectural level, RCA introduces the Cognitive Node as a standardized, fractal unit of intelligence that operates identically across all scales of deployment. The fractal consistency of the node eliminates translation layers between hardware tiers and enables the system to scale fluidly between fully distributed and fully centralized configurations through the property of collapsibility. The Metacognitive Critic embedded in each node provides a continuous measure of epistemic uncertainty that governs both the system's action selection and its communication behavior, implementing caution and escalation as architectural properties rather than policy decisions.

At the learning level, RCA introduces Self-Supervised Physics as an alternative to Reinforcement Learning from Human Feedback. By defining the primary loss function in terms of physical prediction error, the framework eliminates the human annotation bottleneck and grounds all learned representations in the causal structure of physical reality. The Asynchronous Consolidation mechanism addresses catastrophic forgetting without requiring the system to go offline, while counterfactual simulation during consolidation develops causal reasoning capability as a direct consequence of the learning process rather than as a separately engineered module.

At the scalability level, RCA introduces the Multi-Scale Embodiment framework and the Proof of Reality protocol. Together, these mechanisms transform the global hardware ecosystem — robotic manipulators, autonomous vehicles, wearable devices, and industrial sensors — into a distributed physical learning network whose participants are economically incentivized to direct their hardware toward the frontiers of the system's ignorance. The Federated Learning protocol ensures that this global network operates within practical bandwidth constraints by transmitting gradient updates rather than raw sensory data.

At the safety level, RCA embeds safety constraints directly in the learning objective through the Physical Integrity Constraint and the Irreversibility Measure, ensuring that caution and capability scale together rather than in opposition. The entropy-based escalation protocol guarantees inaction as the default response to unresolvable uncertainty, and the Consensus Validation mechanism provides robustness against both hardware failure and adversarial manipulation at the global scale.

Finally, at the evaluation level, RCA proposes the Physical Intelligence Benchmark as a framework for assessing physically grounded intelligence at four levels of cognitive complexity, from basic concept grounding to social and affective modeling. The benchmark is designed to be applicable to any AI system and to provide falsifiable tests of the theoretical claims advanced in this paper.

8.2 Limitations and Open Questions

Scientific integrity requires that we be precise about what RCA does not claim and what it does not resolve.

RCA is a theoretical framework. The mathematical formalizations developed in Sections 3 through 6 define the architecture with precision, but the empirical validation of these formalizations against real physical systems remains future work. The Physical Intelligence Benchmark proposed in Section 7 is the instrument for that validation, but the results of administering it to an RCA implementation are not yet available. The framework should be evaluated on the strength of its theoretical coherence and the falsifiability of its predictions, not on empirical results that have yet to be produced.

The value specification problem identified in Section 6.4 is not resolved by RCA. The Physical Integrity Constraint ensures that the system pursues its objectives cautiously and prefers reversible actions, but it does not guarantee that the objectives themselves align with human values. Specifying what target states a general intelligence should pursue, and ensuring that specification is robust to the system's increasing capability, remains one of the deepest open problems in AI research. RCA provides a safer substrate on which to address that problem — a system that acts cautiously is easier to correct than one that does not — but it does not solve it.

The transition from simulation to physical reality — the sim-to-real gap — is a known challenge in robotics research that RCA inherits rather than eliminates. The Reality Anchor mechanism described in Section 5.2 mitigates this gap by continuously recalibrating the simulation against physical observations, but the fidelity of that recalibration is bounded by the diversity and density of physical agent deployments. In domains where physical agents are absent or sparse, simulation fidelity may degrade, and the world model's representations of those domains may exhibit the same approximation errors that characterize purely text-based systems.

The economic viability of the Proof of Reality protocol depends on assumptions about token valuation and participant behavior that are standard in the decentralized protocol literature but that have not been validated in the specific context of physical learning incentives. The design of the tokenomics in sufficient detail to support a real deployment is a significant engineering undertaking that lies outside the scope of this paper.

Finally, the Social Physics framework of Section 4.5 models affective states as latent variables inferred from behavioral and physiological observation. This approach is computationally tractable and scientifically grounded, but it raises ethical questions about the appropriate limits of affective modeling in deployed systems that this paper does not address. A system capable of accurately modeling human emotional states has capabilities that could be misused in ways that cause serious harm, and the governance frameworks required to prevent such misuse are a necessary complement to the technical framework developed here.

8.3 Future Work

The theoretical framework developed in this paper opens several concrete research directions that we identify as immediate priorities.

The first and most urgent is the implementation and empirical evaluation of a minimal RCA instance. A two-level hierarchy consisting of a single Class A edge node and a central aggregator, trained on a constrained set of physical interactions and evaluated on Levels 1 and 2 of the Physical Intelligence Benchmark, would provide the first empirical test of the core architectural claims. Such an implementation is feasible with current hardware and software infrastructure and would establish whether the theoretical properties of the Cognitive Node — in particular the relationship between Semantic Entropy and physical prediction accuracy — hold in practice.

The second priority is the formal specification of the Proof of Reality tokenomics. The reward function defined in Section 5.4 establishes the mathematical relationship between entropy reduction and compensation, but a deployable protocol requires additional specification: the mechanisms for verifying the difficulty coefficient $\mathcal{D}_{task}$, the governance of the token issuance rate η, the procedures for resolving disputes arising from the Consensus Validation mechanism, and the interface between the protocol and existing blockchain infrastructure. This specification work is a prerequisite for any large-scale deployment.

The third priority is to develop the Physical Intelligence Benchmark as a community resource. The benchmark, as defined in Section 7, requires a standardized set of novel objects, environments, and task configurations that can be replicated across research institutions. Developing, manufacturing, and distributing these materials and establishing the administrative infrastructure for benchmark administration and result reporting is a significant undertaking that would benefit from collaboration across the robotics and AI research communities.

The fourth priority is a formal investigation of the relationship between the RCA framework and the existing theoretical literature on Active Inference, the Free Energy Principle, and Predictive Coding. The learning methodology of Section 4 draws on these frameworks but does not fully formalize the connections. A rigorous mapping between the RCA formalism and the mathematical apparatus of Friston's Free Energy Principle, in particular, would strengthen the theoretical foundations of the paper and open productive avenues for cross-disciplinary collaboration.

The fifth priority is the development of governance frameworks for deployed RCA systems, with particular attention to the ethical implications of the Social Physics and Affective Grounding capabilities defined in Section 4.5. The technical capability to model human affective states from physiological and behavioral observations is a powerful tool with significant potential for misuse. The development of appropriate deployment constraints, consent frameworks, and regulatory guidance for systems with these capabilities is a research priority that must proceed in parallel with, rather than after, the technical development of the framework.

8.4 Closing Remarks

The question of whether artificial general intelligence is achievable, and if so, through what architectural path, is the defining scientific question of our era. The answer will shape the trajectory of human civilisation in ways that are difficult to overstate and impossible to fully anticipate.

This paper has argued, on principled theoretical grounds, that the answer lies not in the continued scaling of systems that process language without understanding the world that language describes, but in the development of architectures that acquire understanding the way understanding has always been acquired — through interaction, prediction, error, and correction, at every scale of physical reality from the compliance of a grasped object to the dynamics of a planetary environment.

The Recursive Cognitive Architecture is a proposal for what such an architecture might look like. It is offered not as a complete solution but as a coherent theoretical foundation from which empirical investigation can proceed. Its claims are falsifiable, its mechanisms are formally specified, and its limitations are explicitly acknowledged. We invite the research community to engage with it critically, to test its predictions against physical systems, and to improve upon it where it falls short.

The intelligence we are trying to build will, if we succeed, be shaped by the care and rigor with which we design its foundations. We believe those foundations must be physical, distributed, and honest about what they do not yet know. RCA is our attempt to build them that way.

References

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.

[2] Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1–3), 335–346.

[3] Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–424.

[4] Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., and Amodei, D. (2020). Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

[5] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.

[6] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138.

[7] Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017). Active inference: a process theory. *Neural Computation*, 29(1), 1–49.

[8] Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87.

[9] Pearl, J. (2009). *Causality: Models, Reasoning and Inference* (2nd ed.). Cambridge University Press.

[10] Pearl, J., and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

[11] Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Houghton Mifflin.

[12] Hawkins, J. (2021). *A Thousand Brains: A New Theory of Intelligence*. Basic Books.

[13] LeCun, Y. (2022). A path towards autonomous machine intelligence. *OpenReview preprint*. Available at: https://openreview.net/forum?id=BZ5a1r-kVsf

[14] Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

[15] Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

[16] McMahan, B., Moore, E., Ramage, D., Hampson, S., and Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 54, 1273–1282.

[17] Pathak, D., Agrawal, P., Efros, A. A., and Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 70, 2778–2787.

[18] Nakamoto, S. (2008). Bitcoin: a peer-to-peer electronic cash system. *White paper*. Available at: https://bitcoin.org/bitcoin.pdf

[19] Brohan, A., Chebotar, Y., Finn, C., Hausman, K., Herzog, A., Ho, D., Ibarz, J., Irpan, A., Jang, E., Julian, R., et al. (2023). Do as I can, not as I say: grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691*.

[20] Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C., et al. (2023). RT-2: vision-language-action models transfer web knowledge to robotic control. *arXiv preprint arXiv:2307.15818*.

[21] Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., and Abbeel, P. (2017). Domain randomization for transferring deep neural networks from simulation to the real world.

Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 23–30.

[22] Peng, X. B., Andrychowicz, M., Zaremba, W., and Abbeel, P. (2018). Sim-to-real transfer of robotic control with dynamics randomization. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 3803–3810.

[23] Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., and Dean, J. (2017). Outrageously large neural networks: the sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*.

[24] Schmidhuber, J. (1991). A possibility for implementing curiosity and boredom in model-building neural controllers. *Proceedings of the First International Conference on Simulation of Adaptive Behavior*, 222–227.

[25] Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. (2016). Continuous control with deep reinforcement learning. *Proceedings of the 4th International Conference on Learning Representations (ICLR)*.

[26] Oudeyer, P.-Y., and Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, 1, 6.

[27] Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P., and Irving, G. (2019). Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*.

[28] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

[29] Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515–526.

[30] Helbing, D. (1995). Social force model for pedestrian dynamics. *Physical Review E*, 51(5), 4282–4286.