Distributed Predictive Latent Dynamics (DPLD): Initial Exploration and Motivation

Isaac Landes
Researcher, Hollis Research

Samuel Berkebile
Contributor, Hollis Research

April 19, 2025

Abstract

The enduring scientific quest to understand consciousness and achieve artificial general intelligence (AGI) necessitates theoretical frameworks that move beyond current AI paradigms, which largely fail to capture the integrated, adaptive, and self-aware nature of biological cognition. This paper introduces Distributed Predictive Latent Dynamics (DPLD), a framework proposing that consciousness is not an engineered function but an emergent property arising from the specific dynamics of a self-organizing complex system. DPLD posits a high-dimensional Central Latent Space (CLS) – a dynamic manifold representing the system's integrated state – which serves as the sole medium for interaction among a distributed network of specialized modules. Each module implements local predictive models, driven by the minimization of prediction error (surprise), and influences the CLS through weighted, blended vector updates. Global coherence and adaptive regulation are orchestrated by a hierarchical Meta-Model that monitors CLS dynamics, predicts future global states (e.g., stability, coherence), and applies learned, non-symbolic modulatory vector fields to guide the system towards attractor states characterized by high predictive accuracy and self-consistency. By synthesizing principles from predictive processing, dynamical systems theory, information integration concepts, and computational neuroscience, DPLD offers a mechanistic, albeit theoretical, account of how consciousness-like properties might emerge from the interplay of distributed prediction, latent integration, and adaptive regulation. This framework emphasizes the primacy of interaction dynamics and self-organization in the genesis of mind, providing a potential pathway towards more general and robust artificial intelligence. [1]

[1] Early preprint versions. Parts I & II are intended to be read in sequence; please check arXiv for the most recent versions. This paper is elaborated in Landes & Berkebile 2025b [1].

1 Introduction: Limitations of Current Paradigms and the Need for Emergence

Despite significant advances in artificial intelligence (AI), particularly with Large Language Models (LLMs) demonstrating remarkable capabilities in pattern recognition and generation [2–4], current systems exhibit fundamental limitations. They often lack robust common-sense reasoning, struggle with out-of-distribution generalization, require vast amounts of labeled or static data, and show little evidence of integrated self-awareness or genuine understanding [5–7]. These models primarily function as sophisticated statistical correlators, optimized for specific objective functions (e.g., next-token prediction), rather than as adaptive, autonomous cognitive agents. This architectural and functional gap highlights the need for alternative frameworks capable of supporting the emergence of more general intelligence and potentially consciousness.

Simultaneously, the scientific study of consciousness grapples with explaining subjective experience from physical mechanisms. Leading theories offer valuable perspectives: Integrated Information Theory (IIT) provides a mathematical framework for quantifying consciousness based on system complexity and integration [8, 9]; Global Workspace Theory (GWT) posits a functional architecture where information becomes conscious when broadcast to a central workspace [10, 11]; and the Free Energy Principle (FEP), along with its process theory Active Inference, proposes that biological systems minimize prediction error (or variational free energy) to maintain homeostasis and model their world [12, 13]. While these theories provide crucial insights, they often remain descriptive or lack a fully generative computational account of how the dynamics of information processing give rise to the unified, subjective quality of experience.

This paper introduces Distributed Predictive Latent Dynamics (DPLD) as a candidate generative framework. DPLD is founded on the principle that consciousness is not a feature to be explicitly engineered but an emergent property of a specific class of complex adaptive systems. It proposes an architecture centered around a dynamic, high-dimensional Central Latent Space (CLS) that mediates the interaction of distributed, specialized modules. These modules operate based on local predictive processing principles, striving to minimize surprise. A hierarchical Meta-Model provides global regulation, guiding the system towards states of coherence and stability. DPLD hypothesizes that the interplay between local prediction error minimization, global integration within the CLS, and adaptive regulation forces the system to self-organize into complex, stable, and predictive states that may exhibit the functional correlates of consciousness. This framework aims to bridge computational mechanisms with emergent cognitive phenomena, offering a theoretical pathway towards AGI and a deeper understanding of mind.

Road-map. The present paper is primarily descriptive and conceptual, outlining the core ideas and motivations behind DPLD. It aims to provide intuition and highlight the potential of this approach by discussing the architecture, its theoretical underpinnings, and potential emergent properties qualitatively. Part II of this work (Landes & Berkebile 2025b [1]) formalises the mathematical mechanics, provides proofs for stability and learning mechanisms, proposes concrete algorithms, and connects the framework to recent empirical findings.

2 Theoretical Foundations and Relation to Existing Work

DPLD builds upon and synthesizes insights from several key areas, offering novel integrations:

- **Predictive Processing (PP) / Active Inference [12, 13]:** DPLD adopts the core PP tenet that systems actively predict their sensory inputs and internal states, using prediction error (surprise) as the primary driver for learning and action (or internal state adjustment). DPLD operationalizes this within a specific architecture: modules generate local predictions about the CLS state, and surprise ($\mathcal{S}_m$) directly modulates their influence (α_m) on the CLS and drives internal plasticity ($\Delta\theta_m \propto -\nabla_{\theta_m}\mathcal{S}_m$). The Meta-Model extends this principle hierarchically, predicting global system dynamics.

- **Global Workspace Theory (GWT) [10, 11]:** The CLS serves a function analogous to the global workspace – a site of information integration accessible to multiple specialized processes. However, DPLD diverges significantly: the CLS is a continuous, high-dimensional manifold supporting complex dynamics, vector blending, and graded representations, not a discrete buffer with winner-take-all broadcast. Access and influence are governed by learned gating, attention, and surprise modulation, rather than simple thresholding. The Meta-Model provides continuous, learned regulation, contrasting with simpler gating mechanisms often implied in GWT implementations.

- **Integrated Information Theory (IIT) [8, 9]:** DPLD proposes a dynamic mechanism that could generate systems exhibiting high integrated information (Φ). The differentiation arises from specialized modules, while integration occurs dynamically through the CLS, where module outputs are non-linearly combined and globally influenced by the Meta-Model. The framework focuses on the generative process leading to integration, suggesting that maximizing predictive coherence within the CLS implicitly maximizes relevant integrated information.

- **Self-Supervised Learning (SSL) and World Models [5, 14]:** DPLD aligns with the SSL paradigm, as modules learn primarily from internal predictive signals. It proposes an emergent, distributed world model implicitly encoded within the learned dynamics of the CLS attractor landscape and the parameters of the modules and Meta-Model, rather than a single, explicitly trained monolithic world model. This model is continuously updated through interaction.

- **Dynamical Systems Theory [15]:** Cognition is explicitly framed as the trajectory of a high-dimensional nonlinear dynamical system. The CLS state $\boldsymbol{c}(t)$ evolves according to differential or difference equations influenced by module inputs and Meta-Model regulation. Concepts like attractors (point, periodic, chaotic), bifurcations, stability analysis (Lyapunov exponents), and self-organization are central to understanding the system's behavior and potential for emergent complexity.

- **Attractor Neural Networks [16, 17]:** The CLS dynamics are conceptualized as evolving on a learned energy or potential landscape. Stable states (coherent thoughts, percepts, memories) correspond to attractor basins (minima in the potential landscape, corresponding to low surprise or free energy). The structure of this landscape is shaped by learning and modulated by the Meta-Model.

- **Computational Neuroscience:** DPLD draws analogies from various neural mechanisms: the CLS resembles the integrated state across interconnected cortical areas; modules mirror specialized cortical regions; surprise modulation parallels attentional gain control [18]; Meta-Model influence is akin to top-down control and neuromodulation [19]; dynamic binding via transient synchrony [20, 21] could occur within the CLS; and learning rules can incorporate biologically plausible mechanisms like STDP [22].

DPLD's primary novelty lies in the specific synthesis: proposing that the structured interaction between distributed predictive modules, mediated solely by a dynamic latent manifold (CLS) and regulated by a hierarchical predictive controller (Meta-Model), provides the necessary and potentially sufficient conditions for the emergence of integrated, adaptive, and potentially conscious cognition.

3 Core Theory: Architecture and Dynamics of DPLD

The DPLD architecture consists of three interconnected components operating dynamically: the Central Latent Space (CLS), distributed Modules, and the Meta-Model. Figure 1 provides a conceptual overview.

3.1 The Central Latent Space (CLS): A Dynamic Semantic Manifold

The CLS is the heart of the DPLD framework, serving as the substrate for integrated cognitive states. It is conceptualized as a high-dimensional ($D \approx 2^{11} - 2^{13}$) vector space $\mathbb{R}^D$, within which the system's state $\boldsymbol{c}(t) \in \mathcal{C} \subset \mathbb{R}^D$ evolves over time. $\mathcal{C}$ itself is hypothesized to be a dynamically learned manifold whose geometry reflects the semantic structure of the system's internal and external world.

Hypothesis 3.1 (CLS as Learned Semantic Manifold)**.** *The trajectory $\boldsymbol{c}(t)$ evolves on a lower-dimensional manifold $\mathcal{C}$ embedded within $\mathbb{R}^D$. The geometry of $\mathcal{C}$ (e.g., distances, curvature) is learned implicitly and encodes semantic relationships, such that nearby points represent similar cognitive states.*

3.1.1 Intrinsic Dynamics and Homeostasis

To ensure stability and prevent trivial states, the CLS possesses intrinsic dynamics incorporating decay and adaptive normalization. A continuous-time idealization might be:

$$\frac{d\boldsymbol{c}}{dt} = -\gamma(\boldsymbol{c}, t) \odot \boldsymbol{c} + \sum_{m=1}^{M} \boldsymbol{I}_m(\boldsymbol{c}, t) + \boldsymbol{m}_{\text{mod}}(\boldsymbol{c}, t) + \boldsymbol{\eta}(\boldsymbol{c}, t) \tag{1}$$

where γ represents decay, $\boldsymbol{I}_m$ module inputs, $\boldsymbol{m}_{\text{mod}}$ Meta-Model influence, and $\boldsymbol{\eta}$ noise. A normalization step (e.g., projecting onto a hypersphere or adaptive gain control [23, 24]) is also conceptually applied to maintain bounded dynamics. (Note: Heavy proofs or detailed analysis of stability based on this equation are omitted here; see Part II [1]).

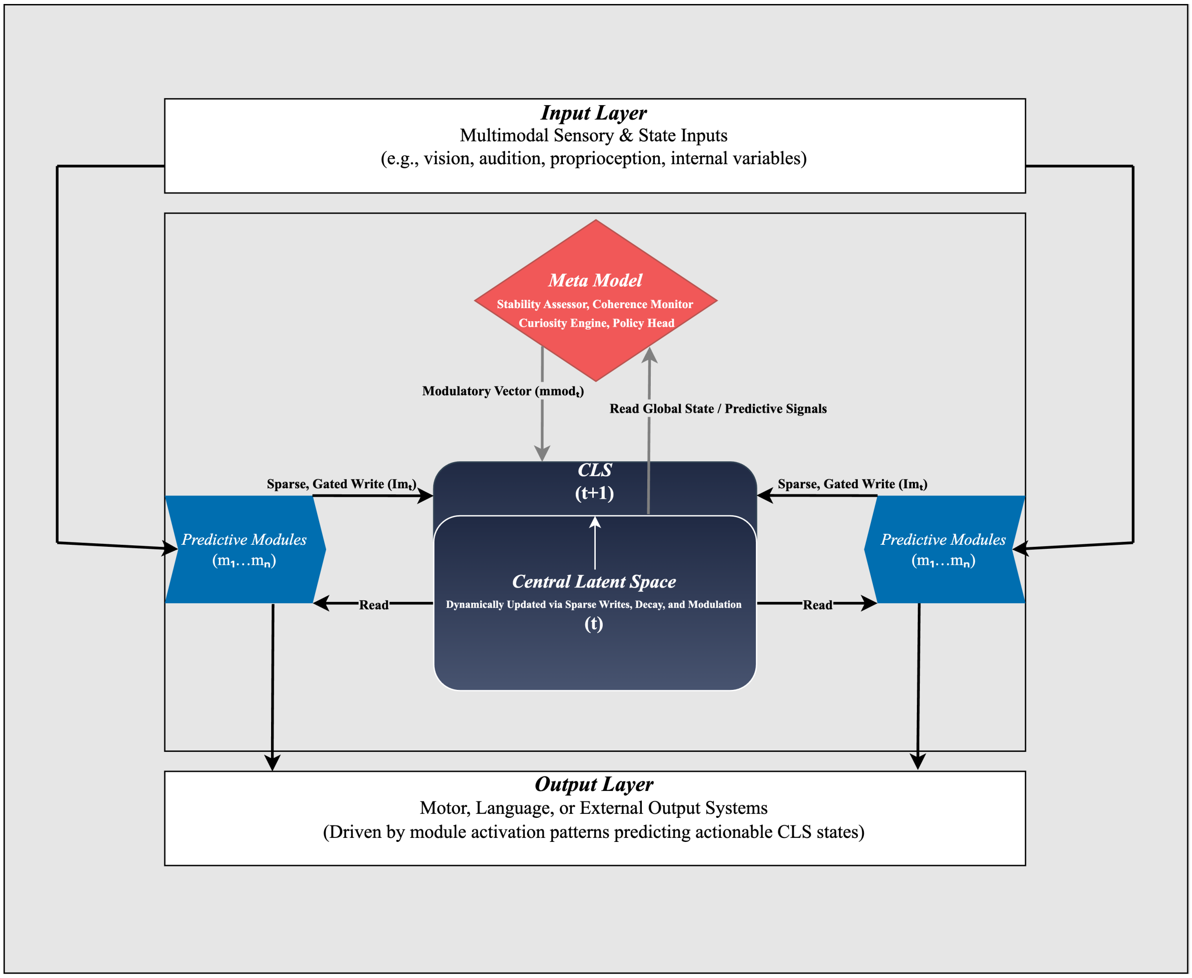
Input Layer
Multimodal Sensory & State Inputs
(e.g., vision, audition, proprioception, internal variables)
Meta Model
Stability Assessor, Coherence Monitor
Curiosity Engine, Policy Head
Modulatory Vector (mmod$_t$)
Read Global State / Predictive Signals
Sparse, Gated Write (Im$_t$)
CLS
(t+1)
Sparse, Gated Write (Im$_t$)
Predictive Modules
(m$_1$…m$_n$)
Central Latent Space
Dynamically Updated via Sparse Writes, Decay, and Modulation
(t)
Predictive Modules
(m$_1$…m$_n$)
Read
Read
Output Layer
Motor, Language, or External Output Systems
(Driven by module activation patterns predicting actionable CLS states)

Figure 1: Conceptual architecture of the Distributed Predictive Latent Dynamics (DPLD) framework. Predictive modules ($m_1 \dots m_n$) process distinct aspects of sensory and internal inputs and interact only indirectly by reading from and writing to a shared high-dimensional Central Latent Space (CLS). The CLS acts as a dynamic internal model, integrating sparse, gated contributions over time. A hierarchical Meta-Model monitors the evolving global state of the CLS and applies modulatory influences to shape system-wide behavior, promoting coherence, stability, and exploration. Local module learning is driven by prediction error (surprise), while global coordination emerges from self-organized interactions through the latent space.

3.1.2 Interaction via Weighted, Gated Blending

Modules influence the CLS not by overwriting, but through a nuanced blending mechanism:

$$\boldsymbol{I}_m(\boldsymbol{c}, t) = \alpha_m(\mathcal{S}_m, t) \odot g_m(\boldsymbol{c}, t) \odot \phi_m(\boldsymbol{v}_m(t)) \tag{2}$$

where ϕ_m projects module output $\boldsymbol{v}_m$ to CLS space, g_m is a learned dynamic gate based on the CLS state $\boldsymbol{c}$, and α_m is a scalar influence weight modulated by the module's local surprise $\mathcal{S}_m$. This allows modules to sculpt specific semantic dimensions with influence proportional to their predictive relevance. (Note: Concrete algorithms for blending and gating are detailed in Part II [1]).

3.1.3 Emergent Semantic Topology and Attractor Dynamics

The structure within $\mathcal{C}$ emerges through self-organization:

- **Self-Organizing Topology:** Modules processing related information may learn to influence nearby regions of the latent manifold, leading to an emergent semantic topology where proximity reflects conceptual relatedness [25].
- **Attractor Landscape:** The interplay of dynamics defines a potential landscape over $\mathcal{C}$. Stable states (thoughts, percepts) correspond to attractor basins (minima of potential energy or free energy) [12, 16, 17]. The system naturally seeks these attractors, which represent its learned world model.

3.1.4 Information Integration and Dynamic Binding

The CLS provides the substrate for integrating features into unified representations, potentially through transient dynamic coordination (e.g., temporal correlations, synchrony [20, 21]), possibly stabilized by the Meta-Model.

3.2 Modules: Distributed Predictive Specialists

Modules ($m = 1, \ldots, M$) are semi-autonomous processing units interacting via the CLS.

3.2.1 Core Functions (Read, Predict, Write)

Each module typically implements: Read (extract relevant CLS projection), Predict (forecast next relevant CLS state based on input, context, internal state using model $f_m(\cdot; \theta_m)$), and Write (generate output $\boldsymbol{v}_m$ to influence CLS based on internal state and surprise).

3.2.2 Local Surprise Minimization and Learning

Learning is driven by minimizing local surprise $\mathcal{S}_m$ (discrepancy between prediction $\hat{\boldsymbol{c}}_{m,t+1}$ and observation $\boldsymbol{c}_{m,t+1}$):

- Surprise modulates influence α_m on the CLS (see Eq. (2)).
- Module parameters θ_m are updated to minimize expected future surprise ($\Delta\theta_m \propto -\nabla_{\theta_m}\mathbb{E}\left[\mathcal{S}_m\right]$), using gradient-based or biologically plausible rules [22]. (Note: Formal credit assignment mechanisms are detailed in Part II [1]).

3.2.3 Module Specialization

Modules specialize based on inputs and predictive targets (e.g., Sensory Encoders, Motor Planners, Language Modules, Memory Modules, Motivational Modules including curiosity drives [26, 27]).

3.3 The Meta-Model: Hierarchical Predictive Regulation

The Meta-Model provides global oversight, ensuring coherence and stability.

Hypothesis 3.2 (Meta-Model as Global Predictive Regulator)**.** *The Meta-Model learns a predictive model of global CLS dynamics (stability, coherence). It optimizes these properties by applying learned, non-symbolic modulatory vector fields ($\boldsymbol{m}_{mod}$) to the CLS, shaping the cognitive landscape.*

3.3.1 Function and Operation

It monitors CLS trajectory summaries, predicts future global properties (stability via $\hat{\lambda}_{\max}$, coherence via $\mathbb{E}\left[\bar{\mathcal{S}}\right]$ or Φ-like metrics), learns to generate modulations $\boldsymbol{m}_{\text{mod}}$ to steer dynamics towards desirable regimes (stable, coherent, predictable), potentially via learned gradient fields, dynamic parameter control, or attractor landscape shaping.

3.3.2 Role in Agency and Higher Cognition

Provides basis for executive function [28], goal-directed behavior (guiding towards learned attractors), reflection/metacognition (monitoring own dynamics, enabling System 2 thinking [29]), and memory management.

Formal Treatment. A rigorous mathematical treatment of the CLS dynamics, module interactions, local learning rules, Meta-Model function, and their interplay, including formal proofs of stability and learning properties under specific algorithmic implementations, is presented in Part II (Landes Berkebile 2025b, §3–§7 [1]).

4 Implementation Considerations, Challenges, and Stability

Realizing DPLD presents formidable challenges.

4.0.1 Core Implementation Hurdles

- **CLS Data Structures/Algorithms:** Need tractable representations (sparse? graph-based? geometric DL [30]?) and stable blending mechanisms.

- **Learning Dynamics/Credit Assignment:** Severe challenge for many interacting modules. Requires solutions beyond standard BPTT (Hierarchical RL [31]? Attention-based credit? Local rules + modulation? Differentiating through dynamics [32]?).
- **Computational Scalability:** Demands massive resources, optimizations (sparsity, async updates), hardware acceleration (GPU/TPU/neuromorphic [33]).
- **Bootstrapping/Initialization:** Needs careful initialization or curriculum learning [34].
- **Measurement/Verification:** Need novel metrics for emergent properties (Info theory? Dynamics? Graphs? Representation analysis [35, 36]?).

4.0.2 Ensuring Stability and Preventing Collapse

Critical challenge. Proposed mechanisms:

- Intrinsic CLS Homeostasis (decay, normalization [23, 24]).
- Meta-Model Regulation (predicting/suppressing instability).
- Local Module Constraints (regularization, inhibition).
- Balanced Plasticity/Stability (managed learning rates, meta-learning).
- Noise Injection ($\boldsymbol{\eta}$) for exploration/flexibility [37].
- Structural Priors (initial architecture guiding self-organization).

(Note: Formal stability guarantees are provided in Part II [1]).

5 Potential Emergent Properties, Implications, and Future Directions

If realized, DPLD could generate systems with properties characteristic of biological minds.

5.0.1 Hypothesized Emergent Properties

- Unified Subjective Experience (Functional Correlate via integrated, coherent CLS state).
- Robust Generalization and Adaptability (via continuous prediction error minimization).
- Emergent Agency and Intrinsic Motivation (seeking low-surprise attractors shaped by internal values).
- Metacognition and Reflection (via Meta-Model monitoring and recursive processing).
- Grounded Semantics (via sensorimotor interaction [38]).

5.0.2 Broader Implications and Ethical Considerations

- Redefining AGI (focus on architecture/dynamics).
- Understanding Consciousness (testable computational framework).
- Ethical Challenges (potential self-awareness, agency, suffering [39]).
- Safety and Control (stability, reliability, alignment of complex self-organizing systems).

5.0.3 Future Research Directions

- Theoretical Development (formal analysis of simplified models).
- Algorithmic Innovation (concrete, efficient algorithms for CLS, learning, stability).
- Computational Simulation (MVPs, scaling, monitoring emergence).
- Architectural Exploration (hierarchies, module types, hybrids).
- Cross-Disciplinary Validation (neuroscience, psychology, consciousness studies; robust metrics).

6 Limitations and Transition to Formal Framework

This paper has presented DPLD at a conceptual level, highlighting its potential but leaving many questions unanswered. Key limitations inherent in this initial exploration include:

- Lack of precise algorithmic specifications for CLS updates, vector blending, gating, and Meta-Model learning.
- Open questions regarding the scalability of the proposed mechanisms, particularly credit assignment and stability maintenance, as the number of modules and CLS dimensionality increase.
- The mechanisms for ensuring robust dynamic stability were discussed conceptually but lack formal guarantees.
- The process of module specialization and the emergence of semantic topology within the CLS requires more detailed theoretical and empirical investigation.
- Quantitative methods for measuring hypothesized emergent properties like integration and coherence were mentioned but not fully developed.
- **Need for rigorous stability guarantees under specific update rules and quantitative mechanisms for curiosity and intrinsic motivation – these are addressed formally in Part II (Landes & Berkebile 2025b [1]).**

These limitations motivate the formal treatment presented in the companion paper (Part II), which aims to provide the mathematical rigor, algorithmic details, and theoretical underpinnings necessary to move DPLD towards concrete implementation and testing.

7 Conclusion: Towards an Architecture of Emergent Mind

Distributed Predictive Latent Dynamics (DPLD) offers a theoretical framework for understanding and potentially building artificial systems capable of emergent consciousness and general intelligence. By integrating principles from predictive processing, dynamical systems, information theory, and computational neuroscience, DPLD proposes a specific architecture—centered on a dynamic Central Latent Space, distributed predictive modules, and hierarchical regulation via a Meta-Model—as a potential substrate for mind. The framework emphasizes that consciousness is not likely to be a programmed feature but an emergent property arising from the complex, self-organizing dynamics driven by the fundamental imperative to minimize prediction error within a structured, adaptive system.

DPLD moves beyond current AI paradigms focused on static pattern matching, proposing instead a system characterized by continuous adaptation, internal state modeling, recursive processing, and emergent goal-directedness. The core hypothesis posits that the structured interplay between local prediction, global integration via the CLS, and adaptive regulation forces the system towards states of high coherence, predictive accuracy, and self-consistency, potentially giving rise to functional correlates of subjective awareness.

While acknowledging the immense theoretical and practical challenges in realizing such a system—particularly concerning algorithmic specification, stability, learning dynamics, and measurement, as highlighted in Section 6—DPLD provides a principled, computationally grounded, and biologically plausible foundation for future research. It represents a call to shift the focus in AI and cognitive science from merely engineering intelligent behaviors to architecting the generative dynamics from which intelligence and consciousness might naturally emerge. The path forward demands theoretical rigor, algorithmic innovation, courageous simulation, and a deep appreciation for the complexities of self-organizing systems. DPLD offers not a final answer, but a structured direction for this profound scientific endeavor, with the necessary formal details provided in Part II [1].

References

[1] Isaac Landes and Samuel Berkebile. A Formal Framework for Distributed Predictive Latent Dynamics. arXiv:25xx.xxxxx (in preparation); builds on Landes & Berkebile 2025a, 2025.

[2] OpenAI. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[3] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

[4] Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Thang Luong, Wei Ping, Karol Singh, DeLesley Tran, Vijay Vasudevan, et al. PaLM 2 technical report. *arXiv preprint arXiv:2305.10403*, 2023.

[5] Yann LeCun. A path towards autonomous machine intelligence version 0.9.2. OpenReview.net, 2022. Accessed April 2025.

[6] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, pages 610–623. ACM, 2021.

[7] Gary Marcus. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*, 2018.

[8] Giulio Tononi. An information integration theory of consciousness. *BMC Neuroscience*, 5(1):42, 2004.

[9] Masafumi Oizumi, Larissa Albantakis, and Giulio Tononi. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Computational Biology*, 10(5):e1003588, 2014.

[10] Bernard J Baars. *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press, 1997.

[11] Stanislas Dehaene and Jean-Pierre Changeux. The global neuronal workspace model of conscious access: from neuronal architectures to clinical applications. In Stanislas Dehaene and Yves Christen, editors, *Characterizing Consciousness: From Cognition to the Clinic?*, pages 55–84. Springer, 2011.

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

[13] Andy Clark. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013.

[14] Yoshua Bengio, Tristan Deleu, Nasim Rahaman, Nan Rosemary Ke, Sébastien Lachapelle, Olexa Bilaniuk, Anirudh Goyal, and Christopher Pal. Representation learning for causal inference. *arXiv preprint arXiv:2301.12934*, 2023.

[15] Steven H Strogatz. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. CRC Press, 2nd edition, 2018.

[16] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

[17] Shun-Ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):77–87, 1977.

[18] John H Reynolds and David J Heeger. The mechanisms of attention in the human visual cortex. *Annual Review of Vision Science*, 1:1–29, 2015.

[19] Peter Dayan and Bernard W Balleine. Reward, motivation, and reinforcement learning. *Neuron*, 36(2): 285–298, 2002.

[20] Wolf Singer. Neural synchrony: a versatile code for the definition of relations? *Neuron*, 24(1):49–65, 1999.

[21] Andreas K Engel, Pascal Fries, and Wolf Singer. Dynamic predictions: oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience*, 2(10):704–716, 2001.

[22] Natalia Caporale and Yang Dan. Spike timing-dependent plasticity: a Hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46, 2008.

[23] Stephen Grossberg. Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 1(1):17–61, 1988.

[24] Gina G Turrigiano and Sacha B Nelson. Homeostatic plasticity in the developing nervous system. *Nature Neuroscience*, 2(1):51–56, 1999.

[25] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[26] Jürgen Schmidhuber. A possibility for implementing curiosity and boredom in model-building neural controllers. In *Proceedings of the International Conference on Simulation of Adaptive Behavior: From Animals to Animats (SAB '90)*, pages 222–227. MIT Press/Bradford Books, 1991.

[27] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 16–17, 2017.

[28] Earl K Miller and Jonathan D Cohen. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1):167–202, 2001.

[29] Daniel Kahneman. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.

[30] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[31] Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. Feudal networks for hierarchical reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, pages 3540–3549. PMLR, 2017.

[32] Ronald J Williams and David Zipser. An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation*, 2(1):13–26, 1990.

[33] Catherine D Schuman, Thomas E Potok, Robert M Patton, J Douglas Birdwell, Mark E Dean, Garrett S Rose, and James S Plank. A survey of neuromorphic computing and neural networks in hardware. *arXiv preprint arXiv:1705.06963*, 2017.

[34] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09)*, pages 41–48. ACM, 2009.

[35] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[36] Nikolaus Kriegeskorte, Marieke Mur, and Peter Bandettini. Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:4, 2008.

[37] Geoffrey E Hinton and Drew Van Camp. Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the Sixth Annual Conference on Computational Learning Theory (COLT '93)*, pages 5–13. ACM, 1993.

[38] J Kevin O'Regan and Alva Noë. Sensorimotor knowledge, embodiment, and consciousness: A Zen-inspired approach. Unpublished manuscript, available online. Updated status., 2001.

[39] Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, 2014.