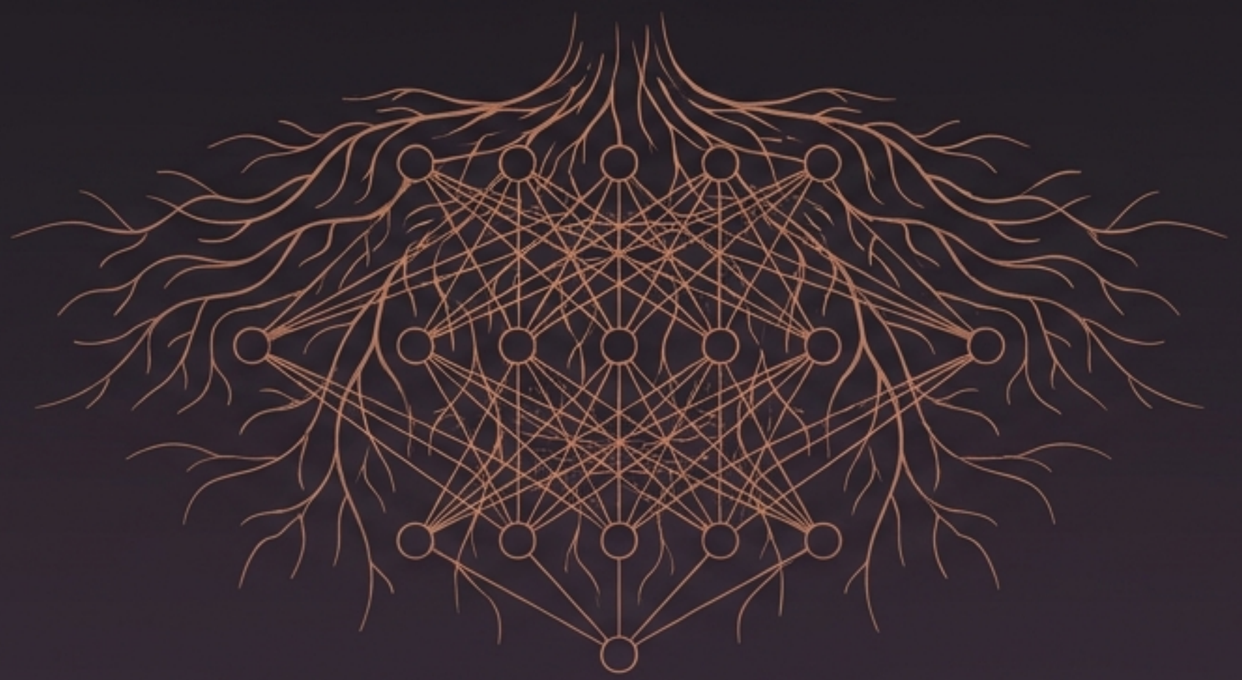
There Will Be a Scientific Theory of Deep Learning

The emergence of Learning Mechanics.

Based on arXiv:2604.21691 [stat.ML] | A 41-page seminal synthesis.

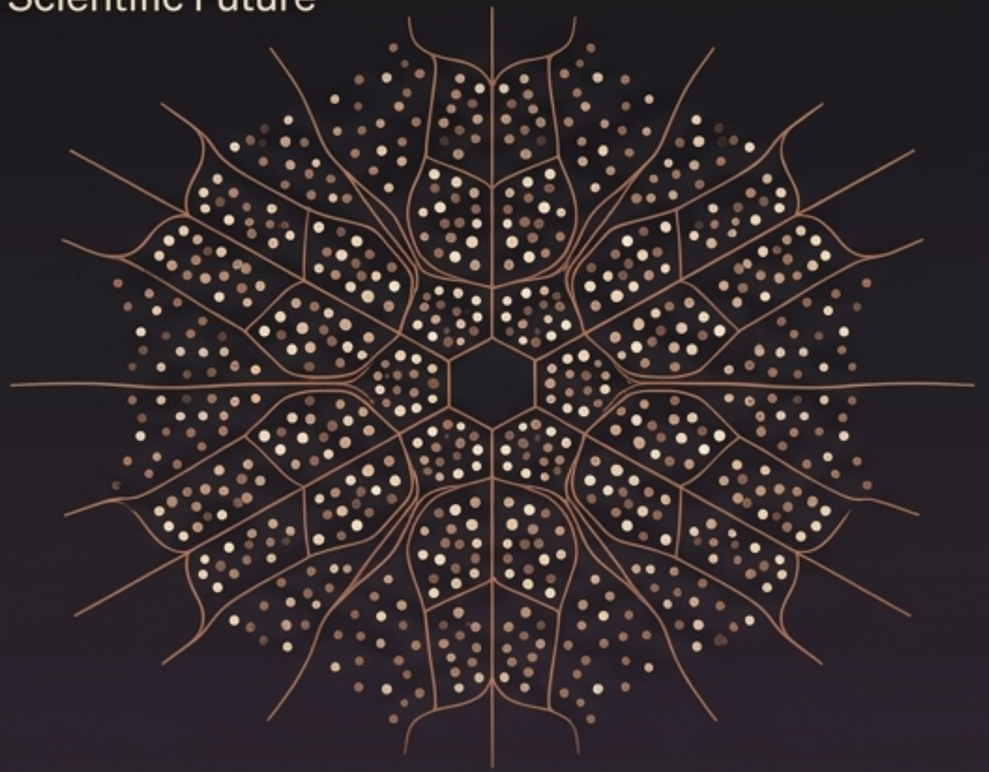


Jamie Simon, Daniel Kunin, Alexander Atanasov, Enric Boix-Adserà, Blake Bordelon, Jeremy Cohen, Nikhil Ghosh,
Florentin Guth, Arthur Jacot, Mason Kamb, Dhruva Karkada, Eric J. Michaud, Berkan Ottlik, Joseph Turnbull

From Empirical Observation to Characterized Science

The Empirical Past

The Scientific Future



A unified scientific theory of deep learning is emerging.

Deep learning is moving beyond empirical success. A rigorous framework is developing to formally characterize the mechanics of neural networks, moving the field from unstructured observation to predictable science.

The Anatomy of the Emerging Theory

Network Performance
The external, observable success of the model.

The Training Process
The dynamics and progression of learning over time.

Hidden Representations
The internal structuring of information within the network.

Final Weights
The settled, stable parameters post-training.

The emerging theory aims to define the vital properties and statistics across all four of these structural layers.

Five Strands of Emergence

Solvable Idealized Settings

Providing mathematical intuition for learning dynamics in realistic, complex systems.

Tractable Limits

Revealing deep insights into fundamental, structural learning phenomena.

Simple Mathematical Laws

Capturing and formalizing important macroscopic observables.

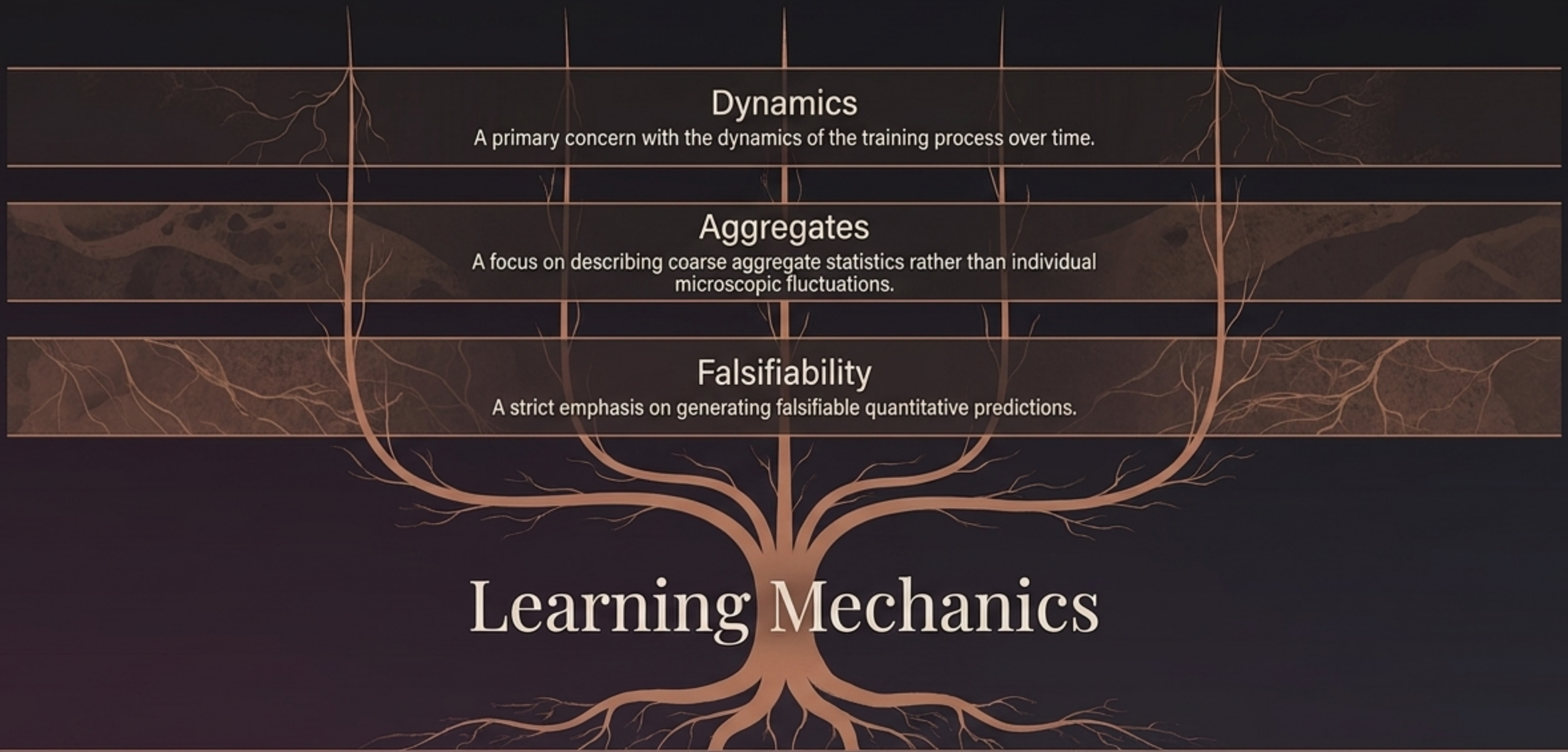
Theories of Hyperparameters

Disentangling variables from the training process to isolate simpler underlying systems.

Universal Behaviors

Identifying shared phenomena across disparate systems and settings that demand fundamental explanation.

The Unifying Traits of a New Paradigm



Taken together, these shared traits transform isolated bodies of research into a cohesive mechanics of the learning process.

Carving Out the Territory

Dimension	Statistical Perspectives	Information-Theoretic Perspectives	Learning Mechanics
Primary Focus	Generalization & Risk	Data Compression & Flow	Training Dynamics & Trajectories
Scale of Analysis	Bounds & Limits	Mutual Information	Coarse Aggregate Statistics
Output Paradigm	Asymptotic Guarantees	Capacity Limits	Falsifiable Quantitative Predictions

While statistical and information-theoretic perspectives provide crucial bounds,
Learning Mechanics specifically models the physical trajectory of the learning process itself.

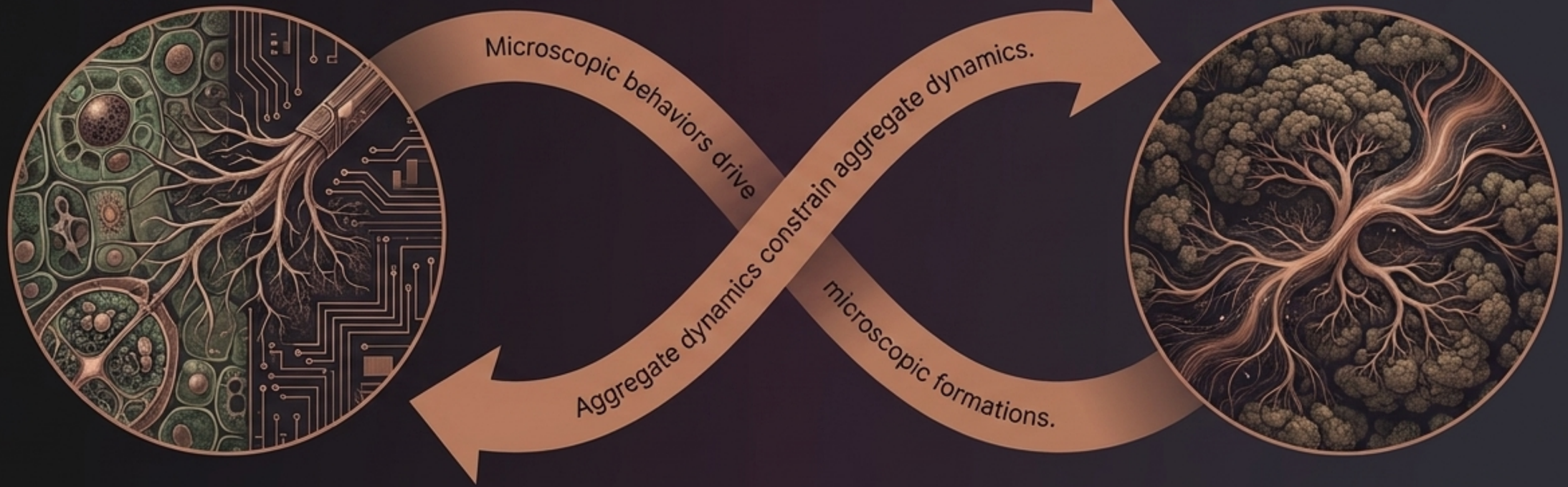
The Symbiosis of Scale

Mechanistic Interpretability

The Microscope. Focuses on the precise, microscopic behavior of individual components, circuits, and neurons.

Learning Mechanics

The Macro-Lens. Focuses on the coarse, macroscopic aggregate statistics and overarching dynamics of the system.



The paper anticipates a deeply symbiotic relationship. Neither perspective alone is sufficient; together, they provide a complete biological and ecological understanding of artificial neural networks.

The Road Ahead for Learning Mechanics

Addressing the Skeptics

A fundamental theory is both possible and imperative. The mechanics of learning are governed by discoverable, observable laws, countering arguments that deep learning is irreducibly opaque.

Important Open Directions

The territory is vast. Future work must bridge the remaining gaps between idealized tractable models and the universal behaviors of frontier, real-world networks.

Advice for Beginners

The field is highly accessible to new researchers willing to engage with falsifiable predictions and macroscopic observables.

Access further introductory materials, perspectives, and open questions via arXiv:2604.21691.