Hi all,

Here is your disruptive ML/AI digest for the day.

Here's what you folks might find interesting today.

TOP PICKS

Novel architectures and objectives

- **[Autoregressive Boltzmann Generators]**

This is one of the clearest conceptual departures in today's list: it replaces the usual normalizing-flow formulation of Boltzmann Generators with an autoregressive generative framework. That matters because the paper explicitly targets the topological and likelihood-computation bottlenecks that have limited flow-based BGs, rather than just tuning an existing sampler. The authors also emphasize sequential inference-time interventions and scaling behavior borrowed from architectures that work well in large language models, which suggests a genuinely different computational primitive for equilibrium sampling. For your interests in **probabilistic/statistical methods** and transferable ideas from physics, this is especially relevant because it ties exact-likelihood style statistical physics objectives to a more flexible generative family. Also worth noting: **Yoshua Bengio** is on the author list, which gives this extra weight.

- **[Generative Models on Analog Hardware with Dynamics]**

This paper is interesting because it does not just compress a standard neural generator onto hardware; it starts from the opposite direction and asks what generative modeling looks like when the dynamics are constrained by physical analog systems. The proposed Analog Interaction Systems framework treats hardware-imposed differential equations as the primary modeling object, then adds two hardware-compatible mechanisms, time-varying piecewise parameters and hidden physical states, to recover expressivity. That is exactly the kind of non-mainstream architectural move that could matter if you care about alternatives to standard digital deep learning stacks. The use of a Wasserstein GAN objective here is less important than the broader claim: generative modeling can be built around physical dynamics rather than software-defined layers. This also fits the "transferable methods from adjacent fields" priority very well, since the inductive bias comes directly from oscillator and Ising-machine style physics.

Theory and generalization

- **[Recovering Governing Equations from Solution Data: Identifiability Bounds for Linear and Nonlinear ODEs]**

This is a strong theory pick because it tackles a foundational question in scientific ML that is often hand-waved: when is an underlying ODE actually identifiable from observed trajectories? The key move is to use Hausdorff distance on solution sets as the comparison metric between differential equations, which is a more principled object than parameter-space distance for the inverse problem they study. On top of that, the paper derives both identifiability conditions and sample-complexity bounds across linear and nonlinear ODE classes, including Lipschitz and Holder vector fields. That combination of minimax-flavored formulation, metric entropy, and quantitative recovery guarantees makes it much more than an application paper. If you care about **theory/generalization** and physics-informed learning, this looks like one of the most substantial reads today.

- **[Algorithmic Foundations of Deep Learning: Complexity-Theoretic Rates and a Characterization of Universal Approximation]**

The central claim here is that neural-network expressivity should be analyzed through algorithmic complexity, not just smoothness or regularity classes. That is a meaningful shift because it reframes networks as models of computation that can emulate real-valued circuits with explicit complexity control, rather than as generic basis expansions. The universal approximation result is also unusually broad: any definable NN model satisfying a parallelization condition, including architectures with multivariate nonlinearities like attention or layer norm, is universal iff it contains a non-affine nonlinearity. I also like that the paper cashes this out with concrete consequences, such as logarithmic-error complexity for holomorphic functions and efficient compilation of shortest-path dynamic programs. For your interests, this is exactly the kind of **formal analysis of representation and approximation** that could reshape how we think about architecture classes.

- **A Generalization Theory for JEPA-Based World Models**

JEPA has been influential empirically, but the theory has lagged behind, so this paper stands out for trying to formalize what the objective is actually doing. The main conceptual step is to cast JEPA pretraining as conditional spectral graph learning and show equivalence to low-rank factorization of an action-conditioned co-occurrence matrix. That gives a concrete mathematical object behind the latent predictive objective, instead of treating JEPA as a vaguely intuitive representation learner. The paper then connects pretraining error to downstream planning regret and derives finite-sample generalization bounds, which is exactly the sort of bridge from representation learning to control performance that world-model papers often lack. For **theory/generalization** and **emergent/agentic AI**, this is likely one of the more useful theory papers in the batch.

BY CATEGORY

Novel architectures and objectives

- Symplectic Neural Networks for learning Generalized Hamiltonians
- Scalable Message-Passing Quantum Graph Neural Networks in the Weisfeiler-Leman Hierarchy

Theory and generalization

- Structure Before Collapse: Transient semantic geometry in next-token prediction
- Zero-Shot Size Transfer for Neural ODEs on Sparse Random Graphs: Graphon Limits and Adjoint Convergence
- Equivariance and Augmentation for Bayesian Neural Networks
- When are likely answers right? On Sequence Probability and Correctness in LLMs

Probabilistic and statistical methods

- All you need is log
- Learning Probabilistic Filters with Strictly Proper Scoring Rules
- Beyond Global Divergences: A Local-Mass Perspective on Bayesian Inference
- $lambda$-PSD: Scalable Approximate SNR-Optimised Polynomial Stein Discrepancies
- Fast algorithms for learning a Gaussian under halfspace truncation with optimal sample complexity
- Accelerated sampling using SamAdams variable timesteps and position-adaptive Langevin dynamics
- Ribbon: Scalable Approximation and Robust Uncertainty Quantification
- Decision-Aligned Evaluation of Uncertainty Quantification
- The Geometry of Updates: Fisher Alignment at Vocabulary Scale
- Effective Covariance Dynamics in Solvable High-Dimensional GANs
- XMSE-Aware Adaptive Empirical Bayes Estimation

Quantum ML and hybrid paradigms

- Scalable Message-Passing Quantum Graph Neural Networks in the Weisfeiler-Leman Hierarchy

Emergent behavior and agentic AI

- Structure Before Collapse: Transient semantic geometry in next-token prediction
- When are likely answers right? On Sequence Probability and Correctness in LLMs

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 139 papers stat.ML: 16 papers