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

Spin-Weighted Spherical Harmonics Enable Complete and Scalable E(3)-Equivariant Networks

This paper looks like a real architectural contribution rather than another equivariant benchmark tweak. The core claim is that the recent Gaunt Tensor Product shortcut loses antisymmetric interaction paths, and that moving to spin-weighted spherical harmonics restores those missing degrees of freedom without giving up the favorable asymptotic scaling. That matters because it targets a concrete expressivity gap in scalable E(3)-equivariant models, especially for chiral and non-centrosymmetric systems where parity-odd structure is not optional. The abstract is unusually crisp about the mathematical mechanism: the gain is not "more capacity" in the vague sense, but a basis expansion with the right algebraic structure to recover the omitted tensor-product channels. If the construction holds up, this is exactly the kind of geometry-driven primitive that could outlast the current crop of engineering-heavy equivariant approximations. For anyone tracking physically grounded inductive biases and scalable symmetry-respecting networks, this is one of the strongest reads in the batch.

Dendritic In-Context Learning in a Single-Layer Spiking Neural Network

This is a sharp conceptual challenge to the current assumption that in-context learning needs transformer-style depth, attention, or explicit inference-time plasticity. The authors argue that a single dendritic compartment already implements an online learning rule structurally equivalent to leaky Widrow-Hoff LMS, and they build a single-layer spiking architecture around that claim. What makes it notable is that the paper does not just report task performance; it claims a mechanistic identification of the embedded algorithm, backed by a linear probe recovering the online-LMS trajectory from membrane dynamics with high fidelity. If true, that is a much stronger statement than "SNNs can also do ICL": it says the relevant computational substrate may be much simpler and more biologically plausible than the mainstream narrative suggests. The Garg-2022 framing also makes the result legible against a known ICL benchmark rather than a custom setup. This is exactly the sort of non-Transformer computational primitive that deserves attention.

Set Diffusion: Interpolating Token Orderings Between Autoregression and Diffusion for Fast and Flexible Decoding

This is one of the more interesting decoding-model papers here because it changes the factorization itself rather than merely tuning a sampler. The proposed likelihood factorizes over flexible-position, flexible-length token sets, and the architecture is set-causal in a way that still permits KV-cache updates after each inference step. That combination directly targets one of the main practical weaknesses of discrete diffusion language models: rigid generation order and poor compatibility with standard fast decoding infrastructure. The novelty is not just "hybrid AR plus diffusion", but a more general token-set view that allows arbitrary-order decoding, including sliding-window sets, which could matter for infilling and mixed sequential-parallel generation. The abstract also suggests a genuine tradeoff argument: better speed-quality behavior than prior diffusion LMs while retaining stronger infilling than block diffusion. If you care about alternatives to left-to-right generation that are still operationally plausible, this is worth reading closely.

Theory and generalization

Born Discrete, Made Smooth: Variational Formulation of Shallow Neural Networks

This is the most theory-forward paper in the list. The authors replace discrete shallow-network training with a continuum variational problem over parameter densities, and the striking claim is that the resulting objective is lambda-convex, well-posed, regular, and solvable via a single linear system rather than iterative nonconvex optimization. That is a substantial departure from both standard finite-network analysis and the usual mean-field or Wasserstein formulations, which often trade tractability for weak regularity. The paper is especially relevant because it tries to bridge NTK-style linearization and feature-learning regimes in one variational framework, instead of treating them as separate asymptotic stories. The explicit rates are also meaningful: generalization control in the regularization parameter and O(1/N) approximation from finite width to the continuum optimum. If the assumptions are not overly restrictive, this could become a useful conceptual template for rethinking what network training is approximating in the first place.

Aggregation with Exponential Weights is Optimal in Expectation

This is a clean theory result on a long-open question rather than a broad conceptual manifesto, but it is still highly relevant if you value rigorous learning-theoretic progress. The paper settles whether aggregation with exponential weights is minimax-rate optimal in expectation for sufficiently large constant temperature under random design, and does so without a Bernstein-type assumption. The interesting part is the phase-transition statement: AEW is known to fail below some constant temperature, and this work pins down a regime where it becomes optimal, giving a sharp boundary to a problem that has stayed unresolved for years. That kind of result matters because it clarifies when a very classical procedure is fundamentally sound, rather than leaving practice to folklore. It is not a new architecture, but it is exactly the sort of formal resolution that improves the theoretical map of statistical learning. For readers prioritizing principled guarantees over empirical novelty, this is a strong pick.

BY CATEGORY

Novel architectures and objectives

Stable Self-Modulating Quantum Fast-Weight Programmers with Bounded Memory Gates

ART for Diffusion Sampling: Continuous-Time Control and Actor-Critic Learning

Role-Aware Neural Convex Divergence Heads for Asymmetric Representation Learning

Theory and generalization

Geometric Signatures of Reasoning: A Spectral Perspective on Task Hardness

Conditional Co-Ablation: Recovering Self-Repair Backups in Transformer Circuits

An Optimisation Framework for the Well-Conditioned Training of Physics-Informed Neural Networks

Identifiability Limits of Physics-Informed Inference for Spatial Stochastic Dynamics from Static Snapshots

Unveiling the Non-Monotonic Effect of Privacy on Generalization under Byzantine Robustness

Sequential Structure-Sensitive Residual Diagnostics for PDE Inverse Problems

Cross-Audit Projection for Model Risk Prediction

How to Allocate Your Tokens? Scaling Laws with Training Steps and Batch Size

Statistical Properties of k-means Clustering for Data Missing Completely at Random

Probabilistic and statistical methods

Regularized Variational and Spectral Log-Density-Ratio Estimation in the Gaussian Location Model

Prediction Sets for Counterfactual Decisions: Coverage, Optimality, and Conformal Prediction

Full Bayesian Reinforcement Learning via LF-IBIS

Conformal Bayes for Two-Sided Censored Gaussian Regression under Label Shift

Quantum ML and hybrid paradigms

Mechanistic Interpretability and Causal Feature Steering of Neural Quantum States via Sparse Autoencoders

Optimal Stabilizer Testing and Learning with Limited Quantum Memory

Ravines in quantum cost landscapes: opportunities for improved VQA predictions

One More Time: Revisiting Neural Quantum States from a Reinforcement Learning Perspective

Emergent behavior and agentic AI

Certified World Models as Sensing Clocks: Drift-Aware Deadlines for Active Perception

The Dual Nature of LLM Persona: Aggregated Tendencies and Frame-Dependent Geometry

Transferable methods from adjacent fields

Learning Effective Soliton Dynamics from Scattering Data

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 163 papers stat.ML: 20 papers