Hi all,

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

Here's what you folks might find interesting today.

TOP PICKS

Theory and Generalization

Information from coincidences

This one looks especially aligned with your interest in information-theoretic learning theory. The core claim is a single algebraic "mixed coincidence" identity that unifies a surprisingly broad set of variational results: Sanov-style concentration, Chernoff information, Donsker-Varadhan, PAC-Bayes, and several rare-event laws all fall out as special cases. What makes it stand out is that this is not just another bound paper, but an attempt to expose a common primitive underneath multiple classical objectives and generalization tools. The multi-prior extension is particularly notable: the paper derives an exact multi-prior PAC-Bayes penalty with an explicit "coincidence bonus", which could matter if you care about richer posterior-prior geometries than the usual single-prior setup. I would read this as a potentially useful conceptual bridge between information geometry, variational principles, and practical ML objectives. It feels much more foundational than benchmark-driven.

Minimax PAC Bounds for Learning in Exogenous Contextual MDPs

This is a strong theory pick if you care about formal sample complexity and RL generalization. The paper studies contextual MDPs where contexts are exogenous and uncontrollable, then gives minimax-optimal or near-optimal PAC rates in several oracle regimes. The most interesting part is that some of the rates are independent of the context-space size $|Z|$, which is exactly the kind of structural result that changes how one thinks about the problem class. That makes it more than a routine PAC analysis: it identifies when contextual complexity is actually irrelevant to learning difficulty. The variance-reduced algorithm and matching lower bounds also suggest the authors are pinning down the right statistical object, not just proving an upper bound. If your interests include principled RL abstractions and sharp guarantees, this is worth attention.

Compositional Behavioral Semantics for State Abstraction in Reinforcement Learning

This paper is appealing because it tries to give a general language for what state abstractions preserve, rather than introducing yet another abstraction criterion. The key novelty is a compositional framework where behavioral semantics are specified locally through one-step system descriptions, then transferred between abstract and concrete systems with guarantees. That is a meaningful conceptual move: it treats value functions, bisimulation, invariants, and behavioral metrics as instances of a broader semantics-preservation story. The quantitative-metric construction from logical semantics is also a nice bridge between symbolic and metric views of behavior. For researchers interested in formal structure in RL, this looks like the kind of reusable theory that can outlast a single benchmark setting. I would flag it as a foundations paper rather than an algorithmic one.

Probabilistic and Statistical Methods

Variational Inference via Entropic Transport Descent

This is one of the more conceptually interesting VI papers in the list. Instead of kernel repulsion as in SVGD-style particle VI, it formulates each update as an entropy-regularized optimal transport problem derived from a JKO proximal perspective, with Sinkhorn solving the inner step. The important part is the claimed global transport structure: particles are coordinated through a transport plan rather than only local repulsion, which directly targets variance collapse and mode collapse in high dimensions and multimodal targets. I also like that the method can be fully score-free and only needs pointwise evaluations of the unnormalized density, which broadens where it can be used. This sits squarely in your priority zone around optimal transport, variational objectives, and probabilistic inference on hard distributions. If the empirical gains hold up, the bigger contribution may be the transport-based reframing of ParVI itself.

Is Variational Monte Carlo Robust? Sharp Moment Thresholds and Heavy-tailed Stochastic Optimization

This is a very strong cross-disciplinary pick because it connects optimization pathology to geometry of the wavefunction nodal set. The main theoretical result is that the local energy and gradient estimators in VMC are generically heavy-tailed for broad ansatz classes, with precise low-moment thresholds determined by nodal structure. That is exactly the kind of mechanistic, non-empirical explanation that can reshape how people think about training neural quantum states like FermiNet. The proposed PS-Clip-VMC method is then not an arbitrary robustification trick, but something designed to operate in the weak-moment regime the theory identifies. I would prioritize this if you like learning theory imported from physics or statistical mechanics, especially when it reveals why standard stochastic optimization assumptions fail. It is also one of the clearest examples here of geometry driving objective behavior.

BY CATEGORY

Novel Architectures and Objectives

Improved Large Language Diffusion Models

Two-dimensional Hyperbolic RNN Neural Quantum State

Laplace--Fisher Gate Identities for Optimal Matrix-Gated Blended Score Estimation

Communicability-Inspired Positional Encoding (CIPE)

A Framework for Directed Hypergraph Signal Processing via tensor t-SVD

Theory and Generalization

Black-Box Assisted Regression: Phase Transitions and Minimax Optimality

A functional central limit theorem for kernel gradient flow and infinitesimal gradient boosting

Gaussian Mean Field Variational Inference can Overestimate Predictive Variance

When Does Synthetic Data Augmentation Improve Score-Based Imbalanced Classification?

Stabilizing black-box algorithms through task-oriented randomization

Probabilistic and Statistical Methods

An Analysis of Posterior Collapse, Parameterization and Initialization in Variational Deep Gaussian Processes

Efficient Adaptive Data Acquisition via Pretrained Belief Representations

Hierarchical Partial-Order Models for Ranking

Deviance-style normalization for jointly overdispersed counts

Emergent Behavior and Agentic AI

LLM-ACES: Closed-Loop Discovery of Dynamical Systems with LLM-Guided Adaptive Search

Optimization and Training Dynamics

Bridging Spherical Black-Box Optimizers

Training for the Model You Return: Improving Optimization for Iterate-Averaged Language Models

Reliability and Validation

Statistically Valid Hyperparameter Selection: From Tuning to Guarantees

Silent Failures in Physics-Informed Neural Networks: Parameter Poisoning and the Limits of Loss-Based Validation

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 117 papers stat.ML: 13 papers