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

Prototype Language Models

I would put this near the top because it proposes an interpretable language-model architecture rather than a post hoc explanation layer on top of a standard dense LM. PRISM makes predictions through a sparse, non-negative mixture over learned prototypes, with clustering objectives that tie those prototypes to coherent neighborhoods of training examples. That is a real architectural departure from the usual transformer-plus-attribution story, and it directly targets one of the hardest problems in LLM auditing: tracing outputs back to training data influence without expensive retrospective analysis. The most compelling technical claim is that the prototype structure localizes curvature in the loss landscape, making Hessian-based attribution much more tractable and reportedly about 500x faster at matched memory. The paper also claims controllable editing by prototype suppression and linear prototype controllers, which matters because it suggests a route to behavior editing that remains legible in terms of training neighborhoods rather than opaque parameter shifts. For anyone interested in mechanistic interpretability with architectural support rather than after-the-fact probing, this looks unusually relevant.

DiscoLoop: Looping Discrete Embeddings and Continuous Hidden States for Multi-hop Reasoning

This is one of the stronger architecture papers in the list because it identifies a specific failure mode of standard non-recurrent transformers on internal multi-hop reasoning, then changes the recurrent state design to address it. The key claim is that looped transformers help with memory reuse, but still fail because the hidden state that contains the right bridge entity is not well aligned with the token embedding space needed for the next retrieval step. DiscoLoop addresses that by carrying two channels through recurrence: a discrete embedding channel and a continuous hidden-state channel. That is more than a benchmark tweak; it is a concrete representational hypothesis about how internal reasoning should be staged inside a single forward pass. The paper is also stronger than average because it includes a training-free realignment intervention that nearly closes the gap, which makes the bottleneck diagnosis more credible. Stuart Russell is on this paper, which is worth noting, but the main reason to read it is the mechanistic argument about depth-local storage and representation alignment.

Diffeomorphic Optimization

I think this is one of the more conceptually ambitious papers here because it reframes optimization with generative models as approximately Riemannian gradient descent on the data manifold induced by a diffusion or flow map. Instead of using the generator only as a sampler or prior, the paper treats the learned diffeomorphism to base space as the coordinate system in which optimization should happen. That gives a geometric objective-level interpretation for why optimization can stay on-manifold and avoid the rugged ambient-space landscape, with an explicit claim of equivalence up to second-order corrections. The extension to protein design on the Lie groups **SO(3)** and **SE(3)** is particularly relevant if you care about geometry-aware learning rather than generic latent optimization, since they derive an autograd-compatible **SO(3)** gradient and backpropagation through Lie-group ODE solvers. The empirical section is application-heavy, but the core contribution is the geometric formulation, not the protein numbers. This is exactly the kind of cross between differential geometry and generative modeling that could transfer beyond the immediate domain.

Theory and generalization

From Spectral Methods to Sample Complexity Bounds for Fourier Neural Operators

This is one of the clearest theory papers in the batch because it does not just prove a bound for a narrow toy setting; it ties FNO learnability to a classical numerical-analysis condition, namely the existence of stable and accurate spectral discretizations for the target evolution operator. That bridge from spectral methods to operator-learning sample complexity is the novel part, and it gives a principled answer to when FNOs should work rather than treating them as empirically successful black boxes. The paper claims polynomial sample complexity over broad families of dissipative PDEs, including Navier-Stokes, Allen-Cahn, and Cahn-Hilliard, with rates that depend on smoothness and dissipation in interpretable ways. I would flag this as especially relevant if you care about theory that explains an architecture's inductive bias in terms of the underlying functional-analytic structure of the problem. It is less about inventing a new model than about sharply characterizing the regime where an existing but important operator architecture is justified. That makes it a strong read for generalization theory with real contact to scientific ML.

Uniform-in-time Propagation-of-Chaos for Stein Variational Gradient Descent

This stands out on the probabilistic theory side because uniform-in-time results for particle methods are much harder and more meaningful than finite-horizon bounds that degrade away as time grows. The paper gives two complementary routes: a cutoff-based argument for broad distributional metrics, and a sharper finite-dimensional theory for matrix-valued finite-rank kernels where the dynamics close on moments. The latter is especially interesting because it yields genuine parametric **N^(-1/2)** propagation-of-chaos rates uniformly in time for certain Stein observables, which is a much stronger statement than the usual asymptotic story. The conjugacy principle under orientation-preserving diffeomorphisms is also notable, since it extends the analysis from Gaussian targets to broader nonlinear and multimodal settings without abandoning the exact finite-dimensional structure. For researchers interested in variational inference, mean-field limits, and the long-time behavior of interacting particle optimization, this is one of the most technically substantive papers in the set. It is not flashy, but it is the kind of result that can change how seriously one takes SVGD as a principled algorithm rather than a heuristic.

BY CATEGORY

Novel architectures and objectives

SNAP-FM: Sparse Nonlinear Accelerated Projection for Physics-Constrained Generative Modeling

Self-Organized Learning in Oscillatory Neural Networks with Memristive Signed Couplings

Generative Modeling of Quantum Distribution with Functional Flow Matching

Group-Equivariant Poincare Convolutional Networks

GAIA: Geometry-Adaptive Operator Learning for Forward and Inverse Problems

Convolutional Symmetric AutoEncoders: enhancing latent stability via differential geometry

Decision-Aware Training for Sample-Based Generative Models

Theory and generalization

Homogenization of l2-Adversarial Training in High-Dimensions: Exact Dynamics under Stochastic Gradient Descent

A Mechanism-Driven Theory of Phase Transitions in Active Learning

[Measuring Dead Directions: Decomposing and Classifying Singular Structure off Canonical Alignment]

[Validating Causal Abstraction Metrics on Simulated Complex Systems]

[Ghost in the Kernel: In-Context Learning with Efficient Transformers via Domain Generalization]

[Function-Counting Theory for Low-Dimensional Data Structures]

[GRPO, Dr. GRPO, and DAPO Are Three Operations on One Number: The Group-Standard-Deviation Identity]

Probabilistic and statistical methods

[Learning dynamical systems from noisy data with Weak-form Kernel Ridge Regression]

[Characterizing and Identifying Separable Graphical Models]

[Distributionally Robust Linear Regression With Block Lewis Weights]

[Hierarchical Variational Kalman Filtering]

[Approximate full-conformal multi-task regression with reproducing kernels]

[Sample Complexities of Estimating Gumbel-Max Watermark Proportions with and without Reduction to Pivotal Statistics]

Quantum ML and hybrid paradigms

[Balancing Expressivity and Learnability in Quantum Kernel Bandit Optimization]

[Scaling Up Thermodynamic AI Models]

Emergent behavior and agentic AI

[Active-GRPO: Adaptive Imitation and Self-Improving Reasoning for Molecular Optimization]

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 142 papers stat.ML: 19 papers