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

The Token Is a Group Element: On Lie-Algebra Attention over Matrix Lie Groups

This is the clearest architectural departure in today's list. The core move is to make each token itself a bare element of a matrix Lie group, rather than a feature vector that is later pushed through an equivariant mechanism. Attention scores are then defined in closed form from the Lie algebra norm of the relative pose, using `log(g_i^{-1} g_j)`, so the inductive bias comes directly from group geometry rather than a learned kernel. That matters because the construction reaches non-compact affine groups with scale and shear, which the abstract explicitly claims are excluded by irrep-based and surjective-exp approaches. If you care about geometric ML and computational primitives that are not just another Transformer variant with equivariance bolted on, this is exactly the kind of paper worth reading closely. The empirical result is also conceptually aligned with the thesis: a fixed geometric score matches or beats a learned MLP kernel while using far fewer parameters and preserving invariance much more faithfully.

QMaxCal: Path-Space Regularization for Open Quantum Control via Girsanov's Theorem

This one stands out because the regularizer is defined on trajectory distributions in path space, not on control amplitudes or heuristic smoothness penalties. The key idea is to use Girsanov's theorem to get a differentiable closed-form estimator of KL divergence between monitored open-system trajectories, yielding regularizers that penalize the observable consequences of decoherence rather than the raw control signal. That is a genuinely different objective design, and it sits right at the intersection of probabilistic methods, statistical physics style reasoning, and quantum ML. Max Welling is on this paper, which is worth noting given the profile's preference for high-signal authors, even though he is not one of the explicitly starred names. The reported robustness gains under noise mismatch make the objective more than a mathematical curiosity: it seems to encode the right physical bias for control under uncertainty. If you are interested in hybrid paradigms or objectives derived from principled stochastic-process arguments, this looks unusually strong.

Sensorimotor World Models: Perception for Action via Inverse Dynamics

The novelty here is not just "world models plus another auxiliary loss", but a specific claim that inverse-dynamics regularization simultaneously prevents latent collapse and aligns representation learning with controllable factors. That gives the model a more action-grounded inductive bias than JEPA-style predictive objectives alone, and the abstract frames this as a way to avoid frozen encoders, EMA targets, and other stabilization machinery. Bernhard Schoelkopf is a notable author here, and the paper is directly relevant to interests around agentic systems and transferable ideas from control and perception. What makes it worth attention is the conceptual simplicity: preserve information about the action underlying a transition, and the latent state is forced to encode what matters for intervention rather than just what is visually predictable. If the empirical claims hold, this could be a useful design principle for offline world-model learning beyond the specific benchmarks. I would read this as a representation-learning paper with a strong causal-action flavor, not merely a control application.

Theory and Generalization

Fisher-Geometric Sharpness and the Implicit Bias of SGD toward Flat Minima

This is one of the more theoretically substantive papers in the batch. It tackles the long-running flat-minima story by replacing Euclidean sharpness with a Fisher-information-based Riemannian notion that is invariant under smooth function-preserving reparameterizations, directly addressing a standard criticism of flatness arguments. The paper then connects SGD noise, stationary distributions of the induced SDE, and concentration near Riemannian-flat minima, before tying the story to a PAC-Bayes bound. That combination of information geometry, optimization dynamics, and generalization theory is very much in scope for your interests. The abstract is careful about the gap between the true FIM and practical diagonal estimators, which is a good sign that the authors are not overselling the invariance claim. If you want papers that try to repair a shaky but influential intuition with a more principled geometric formulation, this is a strong candidate.

Recurrent neural networks approximate continuous functions

This paper asks a refreshingly nonstandard approximation-theory question: can one fixed network buy accuracy by running longer, instead of growing architecture size with target precision? The answer, according to the abstract, is yes for every continuous function on `[-1,1]`, via a single fixed ReLU RNN whose time evolution performs the approximation. The conceptual bridge is a new intermediate model, the Turing machine with neural units, which sounds like the real technical contribution because it makes the algorithmic viewpoint compatible with explicit neural simulation bounds. This is relevant because it reframes expressivity in terms of runtime as a computational resource, rather than width or depth alone. That is the kind of result that can shift how people think about recurrent computation, universality, and the boundary between neural approximation and algorithmic execution. The minimax lower bounds also make it more than a pure existence theorem.

BY CATEGORY

Probabilistic and Statistical Methods

Score Approximation for Diffusion Models on Arbitrary Low-Dimensional Structures

Global Convergence of Gradient Descent for Score Matching in Gaussian Mixtures via Reverse Fisher Divergence

An Information Theoretic Framework for Graph Novelty Generation via Latent Mixture Modeling

Theory and Generalization

Compositionality Emerges in a Narrow Depth-Connectivity Regime: Architecture Constraints and Solution Manifolds

What Makes Effective Supervision in Latent Chain-of-Thought: An Information-Theoretic Analysis

Quantum ML and Hybrid Paradigms

Effective Dimension Governs Generalization in Quantum Kernel Vision Models

Quantum ring all-reduce: communication and privacy advantages for distributed learning

Emergent Behavior and Agentic AI

Manifold Bandits: Bayesian Curriculum Learning over the Latent Geometry of Large Language Models

Interpreting Neural Combinatorial Optimization via Evolving Programmatic Bottlenecks

Transferable Methods from Adjacent Fields

Kolmogorov-Arnold Reservoir Computing

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 125 papers