Hey 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

**Transforming Rank: How Architecture Navigates the Spectral Pathologies of Depth** (arXiv:2607.14018)

This paper reframes skip connections and normalization layers not as magnitude controllers but as mechanisms for preserving gradient rank across depth, which is a genuinely different lens on why these components matter. The author shows that skip connections trade off rank collapse against ensemble-like behavior via the branch-to-skip scale ratio, and that normalization placement (Pre-Norm vs. Post-Norm) controls this same ratio across depth, unifying a fragmented literature. The two-matrix FFN structure is analyzed as a rank-preservation device where the width expansion follows a Marchenko-Pastur law, ensuring the branch Jacobian stays full rank after the activation. The input-output Jacobian rank at initialization predicts trainability on CIFAR-10. This matters because it recasts architecture design as navigating an intrinsic tradeoff among rank collapse, ensemble behavior, and parameter count, offering a principled alternative to empirical architecture search.

### Theory and Generalization

**Algebraic Representability as the Limiting Regime of Grokking: An Exactly Solvable Model with Holomorphic Activations** (arXiv:2607.13749)

This work provides an exactly solvable model of grokking by constraining two-layer networks with holomorphic monomial activations sigma(z) = z^k, which confines the network output to a (k+1)-dimensional algebraic subspace regardless of hidden width. The authors prove a complete algebraic characterization: a modular arithmetic task is representable if and only if its discrete Fourier support lies on the diagonal u + v = k (mod p), and non-representable targets cannot be fit even on the training set, yielding a positive lower bound on training loss independent of width. Across 585 runs, the algebraic prediction matches observed outcomes with 99.8% accuracy, with outcomes splitting cleanly into instant success or outright failure, eliminating the memorization regime entirely. A bottleneck ablation traces a continuous path from representational failure through memorization-without-generalization to standard grokking. This matters because it identifies the limiting regime where the capacity-grokking relationship collapses into a binary representability question, sharpening our understanding of what grokking actually is.

**The SIGReg Objective as Variational Free Energy: A Theoretical Active-Inference Account of JEPA World Models** (arXiv:2607.13612)

This paper establishes a formal correspondence between JEPA training objectives and Active Inference variational free energy, showing that the choice of anti-collapse regularizer determines whether the objective is a valid AIF surprise bound. The authors organize four non-contrastive regularizers (VICReg, LogDet, PairDist, SIGReg) into an entropy-estimator hierarchy indexed by a prior-miscalibration gap, proving that VICReg and LogDet are unsafe upper bounds while SIGReg eliminates the gap entirely. Under a constant-noise encoder model with successful SIGReg enforcement, the objective becomes an exact information bottleneck and the latent goal cost becomes an exact proxy for AIF pragmatic value. Critically, the authors identify a missing term in all current JEPA world models: state-epistemic value, a future-state coverage signal. All proofs are machine-verified in Lean 4. This matters because it provides a normative foundation for JEPA design choices and predicts a concrete architectural extension that no current model implements.

### Probabilistic and Statistical Methods

**Heavy-Tailed Flow Matching via Random Clocks** (arXiv:2607.13841)

This paper introduces a genuinely new formulation for flow matching on heavy-tailed data by representing heavy-tailed source distributions as mixtures of clock-conditioned Gaussian sources. Conditioning on a clock path yields Gaussian source and flow; marginalizing over the clock produces Gaussian scale mixtures covering alpha-stable and Student-t families. The path-valued clock is encoded via truncated logsignature features, making the conditional vector field practical with negligible overhead. Beyond improved mode coverage and tail-statistic recovery on CIFAR10-LT and HRRR weather fields, the random-clock formulation provides a tail-control interface: varying only the clock law calibrates tail heaviness across distribution families without architectural changes. This matters because standard Gaussian-initialized diffusion and flow matching are poorly matched to heavy-tailed data, and this framework addresses that inductive mismatch while retaining the low-NFE advantage of flow matching.

**Linear Independent Component Analysis via Optimal Transport** (arXiv:2607.14081)

This paper replaces the intractable negentropy objective in classical ICA with the squared Wasserstein distance W_2^2 to a standard Gaussian, and proves that this distance is maximized exactly when the projection recovers an independent component. This is a clean information-geometric reformulation: instead of proxy contrast functions like fourth-order cumulants or parametric log-likelihoods, the authors use optimal transport as the non-Gaussianity measure, with a formal guarantee that the optimum coincides with independence. The proposed OT-ICA algorithm outperforms proxy-based methods on simulated data across different latent distributions and works on EEG artifact removal and econometric price discovery without distributional assumptions. This matters because it replaces heuristic proxies with a principled OT-based objective that has a proven connection to independence, potentially generalizing to other blind source separation settings.

---

## BY CATEGORY

### Novel Architectures and Objectives

- Weight Feedback Computes the Jacobian Transpose Locally in Modern Deep Networks
- Concurrent Image Understanding and Generation: Self-Correcting Coupled Markov Jump Processes
- DeepLoop: Depth Scaling for Looped Transformers
- Discrete Diffusion Models: A Unified Framework from Tokenization to Generation

### Theory and Generalization

- How the Hessian-Spectrum of Neural Networks Depends on Data
- Local Redundancy: An Information-Theoretic Measure of Plasticity from Synthetic Memorization
- Non-Expansive Two-Time-Scale Stochastic Approximation: A Fixed-Schedule One-Quarter Barrier and Bias-Corrected Acceleration
- Verifying formulas for interventional distributions

### Probabilistic and Statistical Methods

- Minimax Theory of Likelihood-Based Deep Learning for Speckle Regression
- Price of Fairness in Bandits: A Tight Minimax Characterization
- Parallel gradient boosting for flexible estimation of conditional distributions

### Quantum ML and Hybrid Paradigms

- Quantum Topological Data Encoding

- Implementations of Quantum and Classical Topology-Aligned Architectures for Molecular Property Prediction

**Emergent Behavior and Agentic AI**

- Analogical Deep Research: Retrieving and Integrating Historical Analogies for Foresight Analysis
- DAGR: State-Conditioned Goal Representations via Difference-Aware Goal Cross-Attention

**Transferable Methods from Adjacent Fields**

- Gauge-Invariant, Parameter-Insensitive Regularization for Potential Recovery from Flow on Directed Graphs

**Other**

- Plausible Deniability Guarantees for Whistleblowers
- Multimodal Empirical Bayes Variational Autoencoders for Joint Longitudinal and Time-to-Event Modeling
- Cluster with Auctions for Vector Search

---

Back tomorrow with more!

P.S. In the most recent submission window: cs.LG: 114 papers stat.ML: 14 papers