

## Research Statement

### Optimization-Based Generative Systems for Scientific Foundation Models

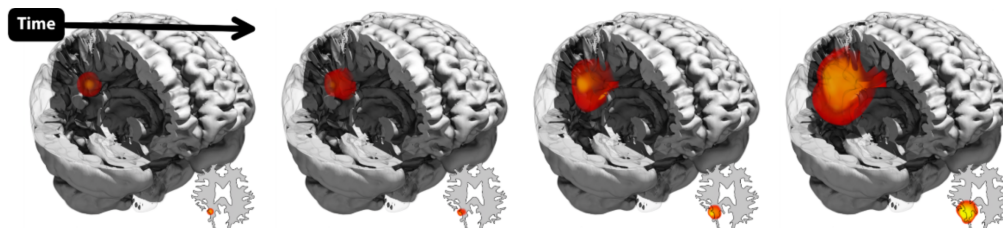
#### Research Overview

My research develops generative AI systems that combine scientific and learned priors within explicit inference-time objectives. Rather than treating generation as a single feed-forward prediction, I build systems in which outputs are produced by optimization, sampling, or search over objective landscapes that can incorporate physical laws, partial differential equation (PDE)-based simulations, measurement data, learned likelihoods, and task-specific constraints. This view is especially important in scientific domains, where useful predictions must be controllable, interpretable, and adaptable to new evidence without retraining.

My work has developed this idea along two complementary directions. First, I have built differentiable scientific priors for inverse problems, where physical equations, biomechanics, and patient or image data are combined through transparent optimization. Second, I have developed efficient-to-sample learnable priors, where the prior itself is learned as an energy landscape that can support classification, generation, and reasoning by being sampled, optimized, and composed with constraints at inference time. Across the four first-author papers that underpin this research statement, my contributions include theoretical formulation, efficient PyTorch/JAX implementations, large-scale evaluation, benchmark development, and open-source dissemination. My long-term aim is to build scientific foundation models that seamlessly integrate learned priors, simulators, tools, and external evidence within mathematically explicit inference loops.

#### Differentiable Scientific Priors

My work on scientific inverse problems began in settings where useful priors are partly known through physical laws or mechanistic models. In *GliODIL* (NeurIPS 2024, Nature Communications 2025), I developed a physics-informed method for individualized glioma radiotherapy planning and recurrence prediction. The model represents tumor growth with a Fisher–Kolmogorov-type physics model and optimizes a discretized loss that softly combines patient imaging data, PDE constraints, and expected tumor growth patterns. By embedding the growth model inside a differentiable optimization framework, GliODIL produces individualized predictions while keeping the assumptions and constraints behind those predictions explicit. I further extended this direction to physics-regularized multimodal image assimilation with biomechanical modeling, including tissue elasticity. This line of work addresses settings in which observations from multiple modalities must be reconciled while remaining physically consistent. Together, these projects shaped my view of scientific AI: in high-stakes domains, models should make the priors, constraints, and uncertainties behind their outputs visible, rather than hiding them inside a black-box predictor.



Physics-informed optimization assimilates patient data into a tumor-growth model for individualized prediction.

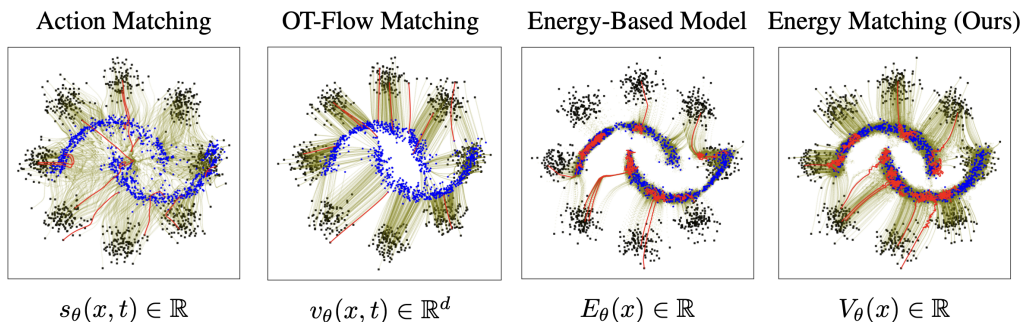
My research is implementation-driven and full-stack. Across these projects, I led PyTorch/JAX implementations, training and evaluation on GPU/HPC infrastructure, evaluation suites, and reproducible open-source releases. This systems practice is central because questions about sampling efficiency, prediction-time compute, robustness of guidance, and constraint satisfaction only become clear when the full generative system is running.

#### Efficient-to-Sample Learnable Priors

The scientific-prior setting raises a complementary problem: many domains do not have complete differentiable simulators or easily specified physical priors. In these cases, I aim to learn priors that remain usable as inference-time objectives. In *Energy Matching* (NeurIPS 2025), I developed a generative modeling framework that gives flow-based approaches the flexibility of energy-based models (EBMs). Many state-of-the-art

generative models match flows or scores along intermediate noise-to-data states, whereas partial observations, priors, and rewards are defined on samples at the data distribution, making test-time integration difficult. Energy Matching instead learns a single time-independent scalar field that connects transport-aligned generation with equilibrium modeling.

Far from the data manifold, samples follow irrotational, optimal-transport-like paths from noise to data; near the data manifold, an entropic energy term guides the system toward a Boltzmann equilibrium distribution. The method advanced the state of the art in energy-based modeling across image generation and controlled protein optimization, while retaining an explicit energy prior that can be composed with observations, constraints, and property objectives. I designed the objective, implemented the training and sampling pipelines, ran the empirical evaluation, and released the code.



Trajectories compare transport from black noise samples to blue two-moons data under Action Matching, optimal-transport Flow Matching, a contrastive-divergence EBM, and Energy Matching. Energy Matching keeps a time-independent scalar landscape, moving efficiently toward data while preserving exploration along the manifold.

The key contribution is that the learned scalar landscape encodes the likelihood structure of the data near the data manifold while remaining well behaved off-support, so that sampling remains efficient rather than becoming trapped in spurious local minima. This issue becomes even sharper for discrete scientific objects such as molecular graphs, where probabilistic inference is valuable because constraints on structure, validity, and function often need to be composed at inference time. In submitted work on *Graph Energy Matching* (GEM), I extended this perspective to discrete graph generation by learning a permutation-invariant potential energy inspired by transport-map optimization. GEM uses this potential both to guide rapid transport from noise toward high-likelihood graph regions and to refine samples once they reach plausible regions of graph space. This bridges the compositional flexibility of EBMs with the sampling efficiency needed for high-quality discrete generation. Beyond unconditional generation, GEM’s relative-likelihood structure enables targeted exploration, property-constrained sampling, and graph interpolation, moving toward scientific foundation models that adapt outputs through explicit inference-time objectives.