

Masked autoencoder pretraining for stable PDE rollouts

Annabelle Miin

Mentors: Anima Anandkumar and Robert Joseph George

Neural operators can predict the next state of a physical system accurately, but small errors compound when the model is rolled forward over many steps. We ask whether masked autoencoder (MAE) pretraining, teaching the model to reconstruct corrupted inputs before any supervised training, can make these rollouts more stable. Prior work has applied MAE to PDE data using vision transformers and spatial patch masking, a corruption borrowed from vision. However, PDE information lives in frequency content, so we corrupt in the frequency domain instead, removing low, high, band, or random coefficient content in Fourier or wavelet space. We pretrain a MAE with a Fourier Neural Operator (FNO) backbone on two datasets from The Well benchmark with each strategy, finetune it for forecasting, and compare against a supervised baseline given the same total compute. Our results show that MAE pretraining in the frequency domain improves mean rollout by 40% on turbulent radiative layer and 12% on active matter compared to baseline.

Expanding past Miller-Schupp: Generating new presentation families for learning Andrews-Curtis trivialization paths using knot theory

Aaron Zhou

Mentors: Anima Anandkumar and Robert Joseph George

The Andrews-Curtis (AC) conjecture claims that all balanced presentations of the trivial group can be transformed into the trivial presentation via a sequence of moves known as AC moves. Finding trivialization paths in the AC conjecture is a difficult task due to the fact that trivialization paths can be thousands of moves long. One difficulty that arises when training machine learning models to predict the moves in an AC trivialization path is the lack of diversity in the families of presentations from which one can source data. Currently, many approaches use the Miller-Schupp family of presentations as their dataset to train their models, risking a lack of generalizability of models trained on this specific family to presentations of the trivial group at large. In order to expand the diversity of this data, we explore a connection between unknot diagrams and AC-trivial presentations by developing a pipeline to generate all unknot diagrams with at most 7 crossings and constructing their Wirtinger presentations. We then apply reduction moves to these presentations to generate a family of 4,046,634 2-generator AC-trivial presentations. We will also report on the ability of machine learning models trained on this dataset to learn the reverse mapping from presentations to knots.

Linearithmic scaling density functional theory via Fourier neural operators

Evan B. Xie

Mentors: Anima Anandkumar and Danish Khan

Kohn-Sham density functional theory (DFT) is limited by the repeated diagonalization of the Kohn-Sham Hamiltonian within the self-consistent field (SCF) loop, a step that scales cubically with system size and, for periodic systems, must be performed independently at every sampled k-point. We introduce a Fourier neural operator (FNO) surrogate that replaces this diagonalization step with a single near-linear operator evaluation, mapping the Kohn-Sham potential directly to the output electron density while leaving the outer SCF loop intact. For periodic systems, the surrogate predicts the already Brillouin-zone-summed density, eliminating the need for explicit k-point summations, achieving even greater speedups in compute time. Across molecular and periodic benchmarks, the resulting FNO-DFT method achieves accurate converged densities, total energies, and band structures at 86-fold reduced computational cost relative to conventional DFT, and outperforms graph neural network baselines trained to predict total energy directly from crystal structure. Our work enables accurate, first-principles electronic structure calculations at system sizes and timescales far beyond the reach of conventional DFT, opening the door to routine quantum-mechanical simulation of large, complex, and previously intractable materials systems.

Capturing membrane dynamics and spike timing of human neurons at scale using neural operators

Cindy Y. Zhao

Mentors: Anima Anandkumar and Valentin Duruisseaux

Characterizing the properties of neuronal cell types is central to understanding their functional roles in the brain. Yet, scaling biophysically realistic neuron models to large populations while preserving rich cellular dynamics and within-cell-type variability remains a major challenge. We present a cell-type-specific neural operator framework for scalable neuronal population modeling that learns a mapping between interpretable electrophysiological features and somatic voltage responses elicited by current stimuli. The model generates large ensembles of these responses while preserving variability across cells and model configurations, achieving a 3,640× speedup over numerical solvers. We evaluate the approach using metrics that describe membrane dynamics, action potential waveforms, spike timing accuracy, and input-output firing relationships. The predicted electrophysiological feature distributions show strong agreement with PDE-simulated outputs and human experimental recordings, including high explained variability in properties such as inter-spike intervals, firing frequency, action potential shape, and post-spike recovery. Through efficient ensemble simulation of cell-type-specific neuronal responses, the framework provides a scalable tool for studying cellular diversity, probing variability near functional thresholds, and generating synthetic yet biophysically grounded neuronal populations.

How to revoke users in encryption schemes anonymously

Benjamin L. Ostrovsky

Mentors: Sanjam Garg and Leonard J. Schulman

Revocation is essential for keeping encryption secure as keys leak, devices are compromised or users leave. We revisit the problem of revocation for encryption schemes and give, to our knowledge, the first approach combining anonymous, collusion-resistant revocation with unchanged keys and access rules, succinct ciphertext overhead and a revocation authority that cannot decrypt. Revoking a user publishes one random-looking pseudonym, without telling encryptors whose key was revoked. A party decrypts future ciphertexts only if it is unrevoked and satisfy the original access rule.

Our core observation is that revocation is, at heart, witness encryption for the statement that one satisfies the access rule and that the pseudonym bound to one's key has not been revoked. Although general-purpose witness encryption remains far from practical, we observe that the statements arising from revocation are simple and algebraic, and we show that they fall within the special-purpose framework of Garg et al. (CRYPTO 2025), which compiles short arguments of knowledge into concretely efficient witness encryption using bilinear pairings. Our technical contribution is a pair of succinct non-membership gadgets for this framework, one for a single hidden pseudonym and one that certifies non-revocation jointly for all users selected by an access rule.

Both gadgets are generic and modular, as they compose with the gadgets expressing a wide range of access rules, and thus provide a single revocation mechanism that is designed once and reused across encryption schemes, without redesigning either the scheme or the revocation.

These gadgets yield the following applications. First, we give a black-box transformation adding revocation to arbitrary independently instantiated public-key encryption, with constant-size cryptographic ciphertext overhead. For unrevoked users, confidentiality continues to rely only on the security of the underlying scheme. Second, we construct revocable schemes for distributed broadcast, registered attribute-based, and registered threshold encryption, with succinct cryptographic ciphertext components and threshold decryption shares. We prove adaptive security of our schemes in the generic bilinear-group and random-oracle models.

Conversational video segmentation

Avni Garg

Mentors: Georgia Gkioxari and Aadarsh Sahoo

Language-grounded video segmentation is essential for spatio-temporal understanding. Referring video object segmentation can spatially localize entities described by simple language expressions and recent work has extended this to reasoning-based prompts. Both settings assume that the target should be segmented throughout the video and emit a mask track per expression with the full clip given offline for the model to see in its entirety before prediction. In this work, we address this gap by introducing Conversational Video Segmentation (CVS) and a benchmark for temporally grounded segmentation, where models must determine not only **what** to segment, but also **when** it should be segmented. We focus on two cases: state conditioned, where the model needs to segment the entity only while it satisfies the prompt (e.g. "the man while he is holding the cup"), and event based prompts, where the model needs to segment the entity before or after the provided event ("the dog before it jumps"). We show that current video segmentation models lack state-aware tracking and are inadequate for CVS. We propose an architecture that can handle this task by predicting a state score for every frame to determine if the entity should be segmented at that time. We also propose a fully automated data generation pipeline for scalable segmentation.

**Structural-functional integration for enhanced detection of subtle epileptogenic lesions:
A framework for a multimodal EEG-MRI clinically-informed approach**

Saanvi Skanda Subramanian

Mentors: Shreyas S. Vasanawala, Pietro Perona, Bruno Passebon Soares, and Mateus Aragao Esmeraldo

Purpose: Focal cortical dysplasia (FCD) is a leading cause of treatment-resistant epilepsy and one of the most diagnostically challenging diseases in neuroimaging. Even the best deep learning models available to detect it rely on imaging data alone and achieve only suboptimal accuracy. A key insight motivates the approach taken here: experienced neuroradiologists and epilepsy specialists do not search the entire brain blindly. Instead, they integrate lectroencephalography (EEG) and clinical findings that pinpoint the region of abnormal activity where seizures originate, and use them to focus and prioritize the lesion search. This study tests whether an automated pipeline can mirror this search pattern by dynamically adjusting the detection threshold of a 3D nnU-Net according to the lesion location based on different localization scales (hemisphere, lobe, and sub-region). This design measures the upper-bound performance improvement that EEG or clinical priors could offer. **Methods:** A 3D nnU-Net was trained on 3D T2-FLAIR images from 175 FCD cases. Anatomical search zones at varying scales were built with SynthSeg, an atlas-based parcellation that runs in seconds per case, far below the roughly eight hours of model training per fold. Rather than applying the fixed 0.5 probability threshold everywhere, the pipeline lowered it inside the localization zone to surface faint sub-threshold peaks and raised it to 0.90 outside the zone to suppress false positives, scoring a case as recovered when any of the top five candidate peaks overlapped the lesion. **Results:** The model detected 86.3% of lesions (mean Dice 0.57), missing 24 cases (13.7%). Restricting a multi-peak search to the lesion lobe recovered 14 of the 24 misses (58%), which a chance baseline confirms is well above the 1.4 expected from a spatially uninformative search ($P < 0.001$). Most of the benefit came from restriction to a moderate scale, the correct hemisphere or lobe; finer subregional restriction added no significant improvement over the lobe. **Conclusion:** This study shows that sub-threshold, spatially coherent signal often persists in cases where a standard nnU-Net reports the case as negative. Future work should focus on integrating real EEG and clinical data to improve deep learning detection pipelines.

Joint optimization of k-space sampling and water-fat separation for accelerated dual-echo MRI

Xiaoli Zhao

Mentors: Shreyas S. Vasanawala, Houman Owhadi, Aiqi Sun, and Yan Wu

3D dual-echo spoiled gradient echo MRI provides strong T1 contrast and enables water-fat separation, but 3D acquisitions are time-consuming, and conventional Dixon-based methods are susceptible to water-fat swaps. No prior work has built self- and cross-attention between echoes directly into the reconstruction network, letting the two echoes share anatomical information during reconstruction itself. We formulate dual-echo reconstruction as a data-driven inverse problem, benchmarking U-Net and T-Net backbones augmented with self-attention, cross-attention, and a physics-informed fusion gate conditioning attention weights on inter-echo phase difference, trained with data-consistency at $R=10$. T-Net with cross-attention performs best (Echo 0: 29.08 dB PSNR, 0.9512 SSIM, 0.0577 NMSE; Echo 1: 32.63 dB, 0.9856, 0.0168), outperforming the U-Net baseline (20.50 dB, 0.7439, 0.4325); cross-attention alone matches the fusion-gated variant, suggesting inter-echo information sharing is the primary driver of improvement. We further extend the fusion-gate backbone into a timestep-conditioned diffusion denoiser with flow-matching (ODE-based) sampling and measurement-consistency guidance; while this trails the deterministic baseline in raw reconstruction quality, its multi-sample uncertainty estimates correlate strongly with reconstruction error, showing predictive value for uncertainty-aware reconstruction.

Learning anisotropic prediction targets for flow models

Rayhan A. Zirvi

Mentors: Leonidas Guibas and Georgia Gkioxari

Flow-based generative models are trained on a hand-chosen prediction target, such as the velocity, the noise, or the clean data. This choice strongly affects training and sample quality. We introduce Anisotropic Flow Models (AnisoFlow), which replaces a fixed prediction target with a fully learnable one. AnisoFlow learns to scale the noise component of the target anisotropically, assigning a separate nonnegative weight to each direction in a jointly learned basis. A closed-form map recovers the velocity from the network's prediction, leaving the architecture, flow-matching loss, and sampler unchanged. Since recovery is exact, training optimizes the target purely to make it easier for the network to predict. We theoretically analyze the resulting prediction error and predict that the preferred noise weight in each direction scales near-linearly with data variance, in agreement with the weights learned empirically in pixel space. AnisoFlow improves FID from 1.90 to 1.55 over the JiT baseline on ImageNet 256x256, by learning only the prediction target, without representation alignment or architectural changes. In pixel-space text-to-image generation, AnisoFlow improves image quality and, through better-conditioned velocity recovery, largely removes patch-seam artifacts under reduced-precision inference. Code and models will be released.

What heads are you really finding? Single-token vs. free-form localization

Ayushi Mehrotra

Mentors: Dylan Hadfield-Menell, Adam C. Wierman, and Aruna Sankaranarayanan

Task settings that elicit a single-token response from a Language Model are widely used in interpretability research to causally localize the circuits underlying the task. Yet, this formulation is poorly suited to complex behaviors such as factual recall, reasoning, summarization, and style transfer. Prior work therefore often studies these behaviors through multiple-choice question answering (MCQA) formulations, or leaves them unexplored altogether. In these constrained settings, a model is said to exhibit the target behavior when it predicts the aligned answer choice. We propose, and defend, an alternative: localizing behavioral circuits directly from free-form generative responses. Across five model families and four tasks with matched MCQA and generative variants, we find that circuits localized from free-form responses are more precise than their MCQA-derived counterparts. We employ mechanistic analyses to explain this difference: MCQA introduces intermediate computations that are absent during direct generation, and contains high-rank representations. Taken together, our results suggest that localizing circuits from free-form responses provides a more direct and effective basis for model control.

HomeBody: A humanoid that explores, remembers, and acts on its own

Gio Huh

Mentors: Karen Liu, Georgia Gkioxari, Guy Tevet, and Takara Truong

A hierarchical approach to humanoid loco-manipulation connects task reasoning (System 2) to motor execution (System 0) through a learned action policy (System 1). This intermediate policy converts visual context and instructions into motion targets. As frontier vision-language models become more capable, we revisit whether System 2 can directly select reusable skills and spatial targets for System 0. We build HomeBody to demonstrate the feasibility of this interface for long-horizon humanoid loco-manipulation. To identify where to act, HomeBody builds persistent spatial memory from the humanoid's own exploration, linking camera observations to measured room geometry. System 2 uses the spatial memory to select a skill and target, while the skill uses local perception and control to execute the motion. On a Unitree G1, this loop supports room-wide tidying and retrieval of medicine from a remembered drawer in a previously unseen kitchen. These qualitative demonstrations use pretrained components without additional environment-specific policy training or task demonstrations.

Spectral methods for causal effect estimation in linear time-series models

Rafael I. Castro

Mentors: Bijan H. Mazaheri and Leonard J. Schulman

We study causal inference in linear time-series systems and perform soft interventions on the causal variables, driving them at chosen frequencies and taking the Fourier projection of observed signals to separate causal structure from confounding. When estimating the effect α of one variable X on another Y and using a soft driving intervention $D(t) = e^{i\omega^*t}$ with effect strength δ , we find the signal-to-noise ratio is proportional to the quantity $\delta\omega^*\sqrt{T}$, increasing with stronger, higher frequency, and longer drives. We then extend the mechanism to DAG recovery on n nodes by driving each node simultaneously at a distinct frequency and de-mixing the response from a single time record. This scheme improved the error of each recovered edge by a factor of $\sqrt{n}$ for equal total observation time. Our results were verified on simulated systems for random Erdős-Rényi DAGs.

What to reward and where to start: Guiding robots that learn by trial and error

Jadon T. Lam

Mentors: Pietro Perona and Ilona Demler

Robots can learn skills in simulation by trial and error, guided by a numerical score, the reward, that tells them how well they did. When the reward is given only for success and success is rare, the learner receives almost no guidance. Two common remedies are to reshape the reward so that it also pays for progress, or to start practice attempts partway through the task, at states taken from a demonstration. This project examined both. In a simulated block-stacking task, a standard shaped reward trained a seven-joint robot arm to bring a cube to its target position in about 90% of attempts. In this first training run, however, a stricter measure requiring the cube to be released and at rest recorded no completed stacks. The reward multiplies progress by the gripper's closeness to the cube, so letting go forfeits most of it. In a second, controlled study, a simulated rocket had to cross a maze with four narrow gaps and was rewarded only for reaching the goal in time. Starting practice from 128 points along a single demonstration made the task learnable: 16 of 16 training runs succeeded, compared with 0 of 5 without these starting points. Unlocking the points backward from the goal made learning about twice as fast. In pre-registered comparisons with up to 32 independent runs per condition, no smarter placement or learned weighting of the points did significantly better than spacing them evenly, and advantages seen in smaller experiments disappeared as the samples grew. A reward for progress along the same demonstration produced no successes in 34 runs. The results suggest that how a demonstration is used matters more than which of its states are chosen, and that shaped rewards can fail in ways a headline success rate does not reveal.