FLATIRON INSTITUTE

FWAM 25' AGENDA

OCTOBER 15TH
WEDNESDAY

162 2ND FLOOR IDA
9:30AM-4:30PM

WELCOME ***9:30AM-9:45AM***

GETTING STARTED WITH DEEP LEARNING FOR SCIENCE AT FLATIRON: MODEL CHOICES

9:45AM-10:45AM

RUDY MOREL, FRF, CCM

Deep learning has become a powerful tool for advancing scientific research, but its success often relies on many model choices, some made unconsciously. This talk will discuss how the problem being addressed, for example, predicting a fully or partially observable system, influences modeling strategies. Several architectures, including transformers, neural operators, and diffusion models, will be considered across tasks such as prediction, denoising, probabilistic modeling, and inverse problems. As an illustration, I will use a project on emulating solar dynamics to demonstrate how these modeling choices influence performance and outcomes.

IMPLEMENTING YOUR ML TOOLBOX

10:45AM-11:45AM

TANYA MARWAH, FRF, FOUNDATION MODELS

In this talk, we will explore how to effectively make use of the modern machine learning toolbox to design your experiment. We will look at the tools available to us through the lens of data scale, model complexity, and—most importantly—the task at hand. We will discuss practical strategies for training and tuning models, managing experiments, and interpreting results, with an emphasis on good habits like logging, tracking, and systematic evaluation. The goal is to build an intuitive sense of what to use and why when developing ML systems in practice.

COFFEE BREAK ***11:45AM-12:00PM***

FROM SEQUENCES TO GEOMETRIC STRUCTURES AND BACK

12:00PM-1:00PM

TERESA HUANG, FRF, CCM

Machine learning (ML) has expanded to a wide range of data modalities including sequences, point clouds, graphs, and manifolds. In this talk, we begin by reviewing the evolution of ML designed specifically for each modality, culminating in the rise of Transformers as general-purpose sequence models. The success of Transformers has sparked a fascinating paradigm shift: flattening complex geometric structures into sequences to be processed by a single architecture. This approach offers a path towards foundation models that can leverage multimodal information, allowing new structural data to be easily integrated with pre-trained sequence models. We will highlight the success of this approach but also its pitfalls, such as the loss of structural inductive biases and scalability issues. We will conclude with open research directions, including enhancing sequence models through hybrid architectures and structural positional encodings, as well as designing diverse geometric benchmarks to drive progress.

FWAM 25' AGENDA

OCTOBER 15TH
WEDNESDAY

162 2ND FLOOR IDA
9:30AM-4:30PM

OUR LEARNING UNIVERSE: AI MEETS THE COSMOS — *1:00PM-1:30PM*

ADRIAN BAYER, FRF, CMB ANALYSIS AND SIMULATION

Artificial intelligence is reshaping how we explore our Universe. In cosmology, machine learning is now used to accelerate simulations, uncover hidden information in the cosmic web, and even to reconstruct the initial conditions of the Universe. As AI becomes central to discovery across ever-larger datasets and higher-resolution telescopes, it challenges us to build models that are not only powerful but also robust, interpretable, and worthy of trust. In this talk, I'll present what happens when AI meets the cosmos—how machines are helping us learn about our Universe, and how, in turn, the Universe is teaching us about learning itself.

LUNCH — *1:30PM-2:30PM*

DECODING THE HUMAN GENOME WITH AI — *2:30PM-3:00PM*

YUN HAO, FRF, CCB

Interpreting the human genome remains a fundamental challenge at the intersection of computational biology and medicine. Recent advances in machine learning—particularly in deep representation learning—are redefining how we model genomic function and disease. In this talk, I will present AI frameworks that learn directly from DNA sequences using architectures such as convolutional and transformer-based models to predict regulatory activity, functional consequences of noncoding variants. I will discuss how these models capture higher-order dependencies, encode sequence–function relationships, and generalize across cell types. Finally, I will illustrate how applying these frameworks to large-scale genomic and clinical datasets has revealed novel regulatory mechanisms and causal drivers of complex diseases, paving the way for interpretable and scalable precision genomics.

AI AGENTS FOR DATA SCIENCE: HOW TO AND SHOULD WE — *3:00PM-3:30PM*

JEREMY MAGLAND, SENIOR DATA SCIENTIST, CCM

Recent advances in large language models (LLMs) and agent-based systems promise new tools for automating aspects of data science: exploring datasets, generating code, running analyses, and even interpreting results. In this talk, I'll walk through how to build LLM-based agents that can access data, execute code, and interact with plots and outputs, demonstrating how these systems can assist in exploratory data analysis. But alongside these capabilities come risks. I'll show concrete examples of how agents can go astray, falling into common statistical traps such as neglecting multiple comparisons, misinterpreting noisy data, or overlooking temporal dependencies. We'll explore not just how to make these tools work, but whether we should trust them, when they fail, and what this could mean for the future of data science.

WHAT IS MACHINE LEARNING GOOD FOR (AND BAD FOR) IN THE PHYSICAL SCIENCES? — *3:30PM-4:30PM*

DAVID HOGG, GROUP LEADER, CCA

Machine learning has become indispensable in multiple scientific domains. I focus on three uses: supervised regression to label unlabeled data; emulation of (or surrogate for) expensive simulations; and unsupervised modeling or prediction of nuisances. I will argue that the first two may bring significant harm to any project, and the last one is of great scientific value.

FLATIRON INSTITUTE

FWAM 25' AGENDA

OCTOBER 16TH
THURSDAY

162 2ND FLOOR IDA
9:45AM-6:30PM

WHAT EVERYONE SHOULD KNOW ABOUT FUNCTION APPROXIMATION IN 1 AND MORE DIMENSIONS
9:45AM-10:45AM

ALEX BARNETT, GROUP LEADER, CCM

I will overview several practical tricks for handling (interpolating) functions efficiently, and discuss how they can help your numerical simulations. This may include speeding up slow functions, numerical differentiation, locating singularities, and smoothly extending or "rolling off" functions. While many are classical, some are active research areas. A lot will be in 1D, some in low dimensions, and a few comments about high dimensions.

QUANTUM SIMULATION WITH TENSOR NETWORKS IN TWO- AND THREE- DIMENSIONS
10:45AM-11:45AM

JOEY TINDALL, ASSOCIATE RESEARCH SCIENTIST, CCQ

Simulating the dynamics of many-body quantum systems is one of the foremost challenges in modern-day physics. In this talk I will demonstrate how structured tensor networks are being used to meet this challenge here at the Flatiron Institute. I will begin by introducing a graphical approach to working with tensor networks and show how it allows the seamless implementation of efficient, GPU-accelerated message-passing-based contraction schemes for their contraction and subsequent optimization.

I will then focus on several, prominent non-equilibrium experiments, and show how a tensor-network ansatz which reflects the underlying lattice geometry can be used to efficiently simulate them. The most recent of these experiments involves both two and three-dimensional lattices of spins driven through a continuous phase transition. For finite-time quenches, we can perform accurate simulations which scale only linearly in the number of qubits and capture the universal physics present via extraction of the corresponding Kibble-Zurek exponent. I will conclude by discussing the future prospects of a structured tensor-network based approach to simulating quantum dynamics, with two and three-dimensional out-of-equilibrium fermionic and spin systems now within reach.

COFFEE BREAK
11:45AM-12:00PM

JAX TRAINING
12:00PM-1:00PM

DAN FOREMAN-MACKEY, CCA ALUM, GOOGLE DEEPMIND

In this interactive session, I will introduce JAX (https://github.com/jax-ml/jax), an open source project that enables differentiable, parallelizable, and hardware accelerated numerical computing. JAX is widely used for machine learning research, but it is also a useful tool for other computing tasks, including the area where I have the most experience: probabilistic programming. After demonstrating the core JAX programming model, I will share some case studies to demonstrate how the community can build on top of JAX to produce performant and user-friendly Python libraries. This session will include some live coding, and plenty of time for discussion.

FWAM 25' AGENDA

OCTOBER 16TH
THURSDAY

162 2ND FLOOR IDA
9:45AM-6:30PM

SOLVING LARGE SYSTEMS OF NONLINEAR PDES IN JULIA — *1:00PM-1:30PM*

DAVID STEIN, RESEARCH SCIENTIST, CCB

Numerically evolving systems that are governed by nonlinear partial differential equations (PDEs) typically requires several routines --- not just the objective function that you seek to solve, but routines to generate Jacobians, Jacobian-vector products, and, when solving at larger scales, preconditioners. This can result in fractured code that is challenging to implement, maintain, and update. The use of automatic differentiation (AD) can eliminate the need for explicitly coding all but the objective function, yielding concise code that is easy to add new features to. As an experiment, we decided to attempt to rewrite SkellySim, our group's large-scale code for simulating the dynamics of many hydrodynamically interacting flexible filaments, within an AD framework in Julia. That process was surprisingly smooth, though not without some difficulties, which we discuss in this talk.

LUNCH — *1:30PM-2:30PM*

BEYOND AUTOCOMPLETE: AI-ASSISTED DEVELOPMENT IN VS CODE — *2:30PM-3:00PM*

THOMAS HAHN, FI SOFTWARE RESEARCH FELLOW, CCQ

Tired of waiting for code reviews? Ready to code with an AI pair programmer?

Discover how GitHub Copilot can transform your development workflow from simple autocomplete to autonomous coding agents.

This talk is aimed at developers who are new to AI-assisted development and want to understand what these tools can do and how to get started using them. I will show how to set up GitHub Copilot in VS Code and demonstrate its core features through live coding: code completion as you type, chat mode for questions and explanations, inline editing for quick fixes, and agent mode for autonomous multi-file tasks.

By the end, you should have a better understanding of what AI can do for your coding workflow and be ready to try it in your own projects.

USING AI EFFECTIVELY: CONTEXT OVER VIBES — *3:00PM-3:30PM*

KYLE ESKRIDGE, FI SOFTWARE RESEARCH FELLOW, CCQ

AI tools are becoming part of everyday research, but getting good results isn't about "vibing" with the model - it's about structure and context. This talk lays out some best practices for using AI productively. The focus is on providing the information and framing that let AI systems perform well, while making sure their work gets double-checked through automated tests and hands-on review. The examples in this talk are primarily based on AI tools for software development, but the best practices apply broadly to AI tools for general tasks. Whether you're experimenting with AI in your workflow or already using it daily, you'll walk away with tips for keeping it focused on the result you want to achieve.

FAST NUMERICS, SMALL CHANGES: THE MINIMAL PLAYBOOK — *3:30PM-4:30PM*

MARCO BARBONE, SOFTWARE ENGINEER, CCM

This is a practical session that requires a laptop!
This tutorial delivers hardware-aware speedups for scientific code with small, targeted changes. We will start from a working baseline and, step by step, move towards more optimised implementations by means of vectorization, jitting and existing plug-and-play tools that are widely available to the scientific community.

RECEPTION & END OF FWAM CELEBRATION! — *4:30PM-6:30PM*