Slide 1 — Introduction

1. Summary of the Paper

This project is based on two main references: *The Well: A Large-Scale Collection of Diverse Physics Simulations for Machine Learning* (Ohana et al., 2024) and *Learning Fast, Accurate, and Stable Closures of a Kinetic Theory of an Active Fluid* (Agrawal et al., 2024). These works provide the foundation for revealing hidden order in active matter systems using autoencoders (AE) and variational autoencoders (VAE).
Active matter systems consist of self-driven particles whose collective motion gives rise to emergent phenomena such as alignment and phase transitions. Their evolution is described by kinetic equations, where the particle distribution $\Psi(x,p,t)$ changes in time according to transport and interaction processes.
Macroscopic quantities like the concentration **c(x,t)** and the nematic order tensor **Q(x,t)** are obtained from Ψ. However, the evolution equations for Q are not closed; they depend on higher-order moments. Neural network closures proposed in the cited work learn these missing terms while preserving rotational symmetry and physical consistency.

2. Description of the Data

The dataset comes from *The Well* benchmark, specifically the **active_matter** subset. It contains 2D simulations of active particles in a viscous fluid, with 360 trajectories of 81 frames each and a spatial resolution of 256×256. Each trajectory includes density, velocity, and orientation fields, capturing transitions such as isotropic–nematic ordering and vortex formation.
Each simulation frame represents a high-dimensional state **U(x,y,t)** = [ρ, v, Q], from which the model learns a compact latent encoding. These encodings are expected to organize frames according to their physical phase and temporal evolution.

3. Connection to the Project Goal

By combining data-driven models with kinetic theory, the project aims to learn latent manifolds that capture physical phases and transitions, showing smooth temporal trajectories that reflect the hidden order in active matter systems.

Slide 2 — Goal of the Project and Problem Statement

1. Goal of the Project

The goal is to uncover low-dimensional latent structures that capture the essential dynamics and transitions of active matter. The encoder $\mathbf{f}\theta$ maps each high-dimensional frame $\mathbf{U(x,y,t)}$ to a latent vector $\mathbf{z = f}\theta\mathbf{(U)}$, while the decoder $\mathbf{g}\phi$ reconstructs the original state minimizing the loss $\mathbf{L = ||U} - \mathbf{g}\phi\mathbf{(f}\theta\mathbf{(U))||^2}$.
The latent space should organize frames by physical regime (isotropic, nematic, turbulent) and reflect continuous temporal evolution, providing interpretable directions related to alignment strength, order, or activity.

2. Problem to Be Addressed

Active matter systems are nonlinear and high-dimensional, making their analysis computationally demanding. Traditional approaches struggle to capture the underlying low-dimensional dynamics or detect hidden transitions. The main challenges are:

- High dimensionality of the simulated fields.
- Hidden transitions between disordered (isotropic) and ordered (nematic) phases.
- Nonlinear coupling between microscopic alignment and macroscopic flow.
- Lack of physically interpretable latent variables.

Formally, given a dataset D = {U■(x,y,t■)}, we seek mappings ($f\theta$, $g\phi$) that compress U into z and reconstruct it with minimal error, while ensuring that distances in latent space correspond to physical similarities. This reveals hidden order and provides a reduced yet interpretable representation of active fluid dynamics.