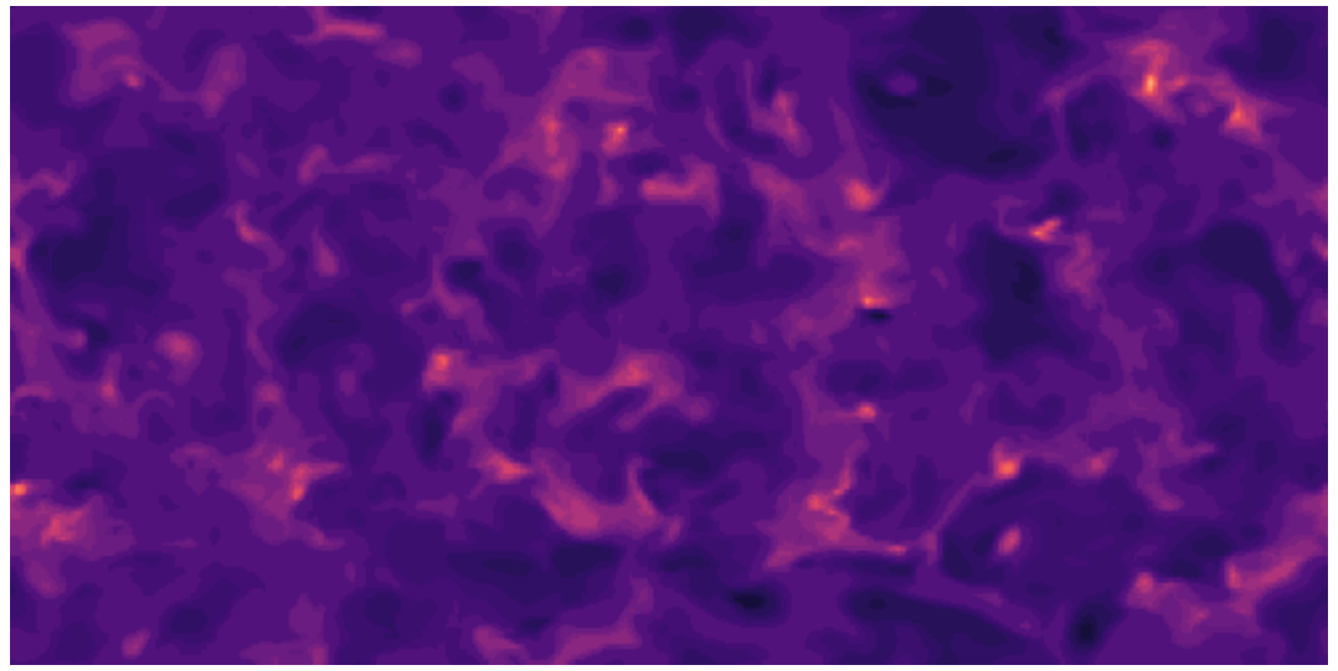
DISCO: learning to DISCover an evolution Operator for multi-physics-agnostic prediction

Jiequn Han

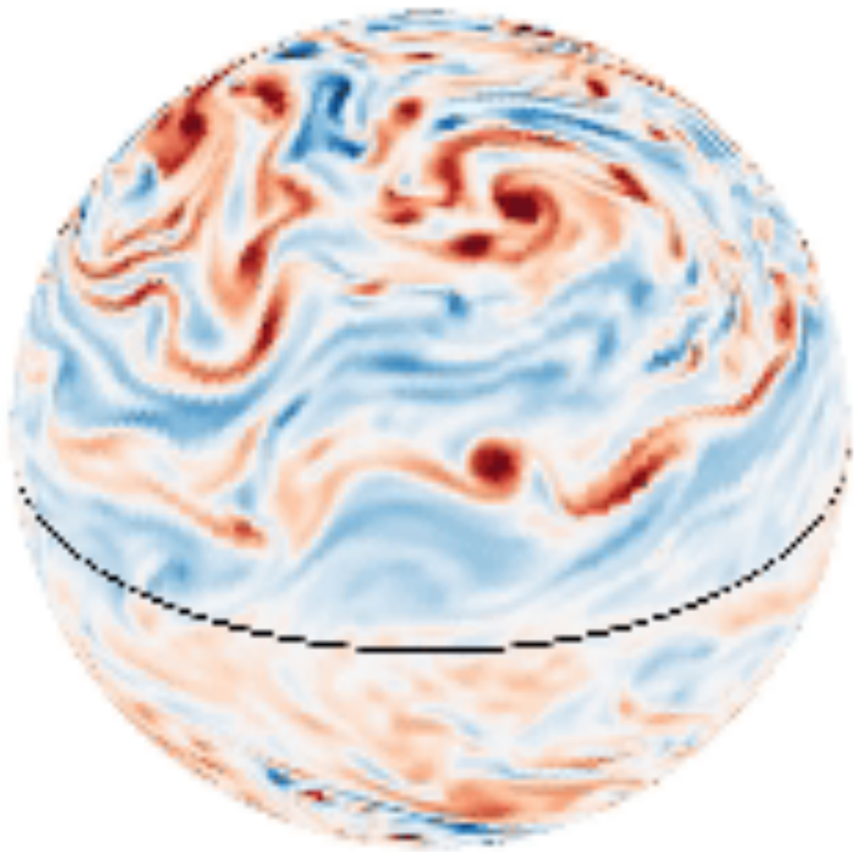
FLATIRON INSTITUTE

Machine Learning Models for Physics

Expensive simulations

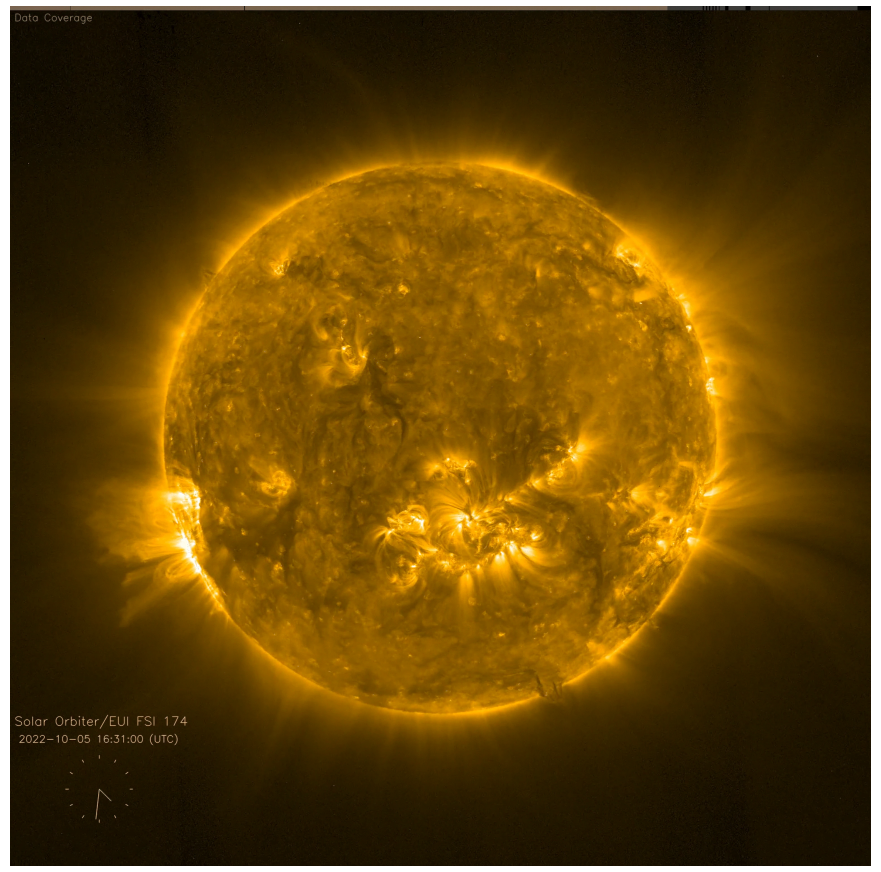


Weather forecast

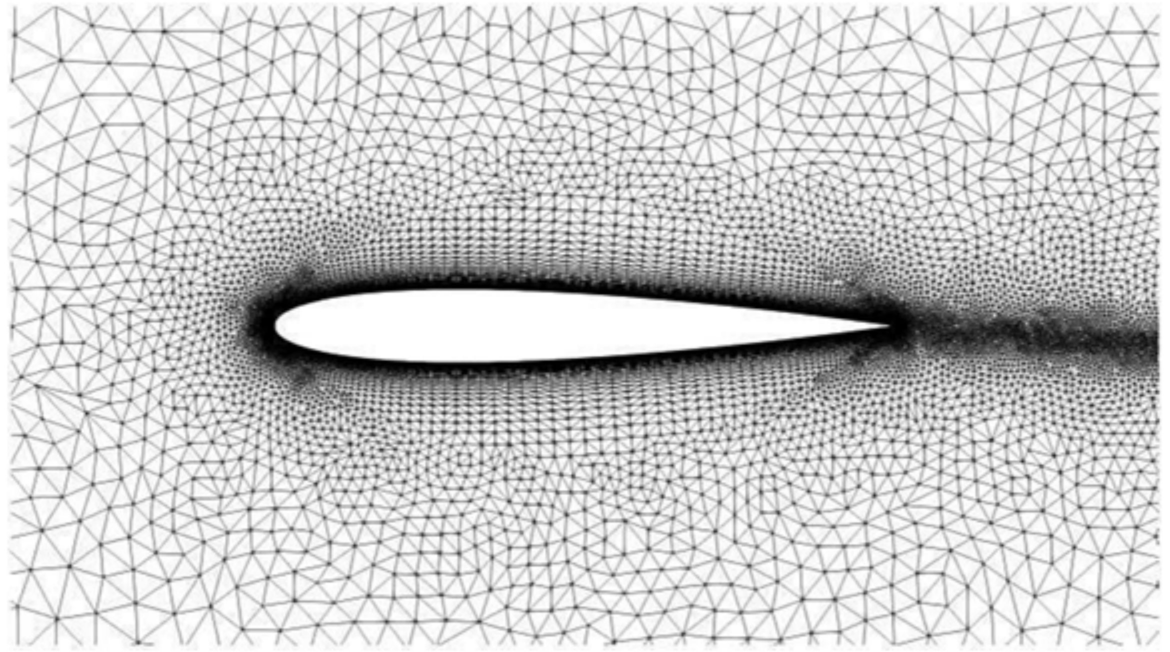


Various ML Models

Solar prediction



Non-uniform grids

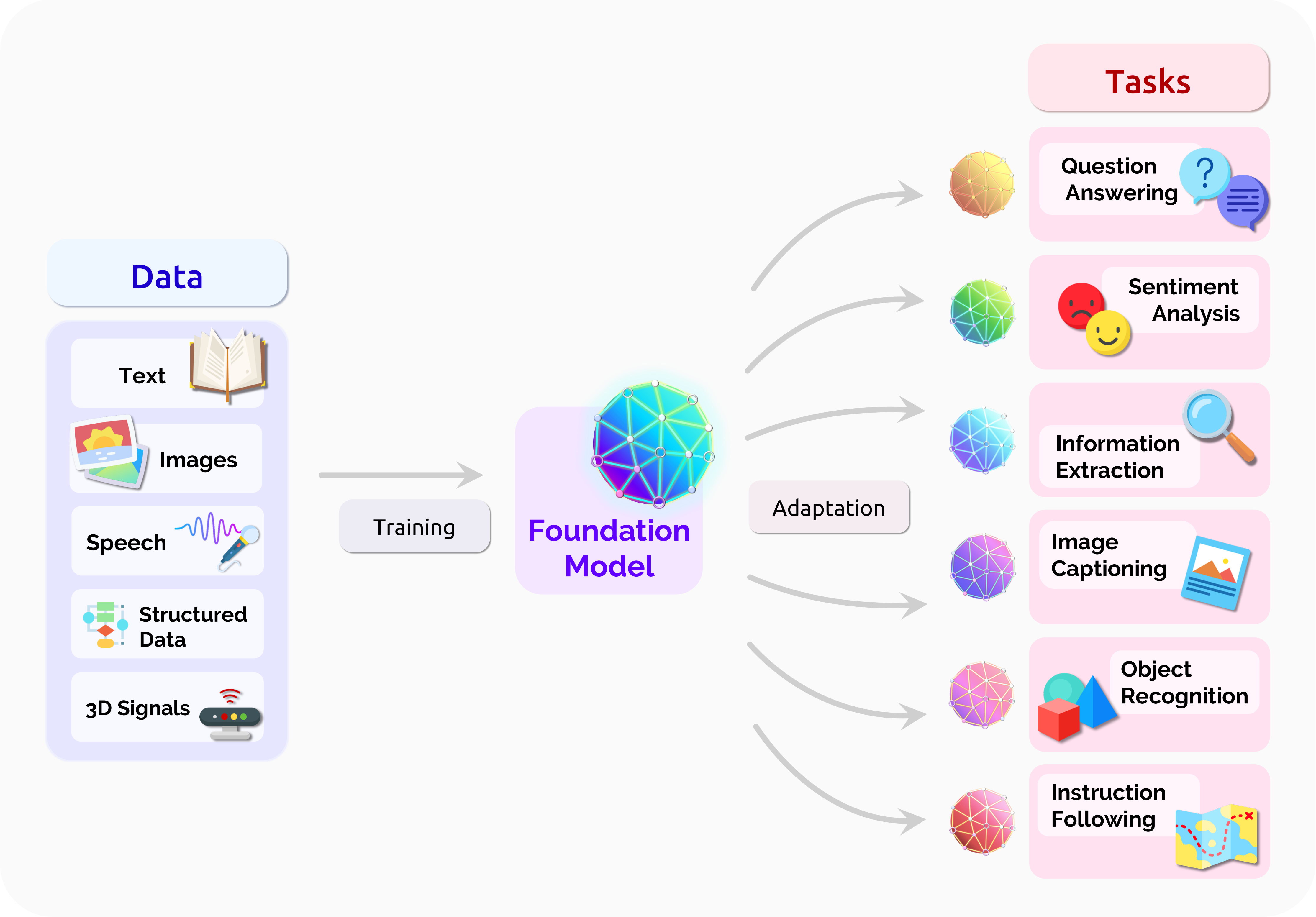


Why training from scratch every time?!

The Rise of the Foundation Model Paradigm

Foundation Model approach:

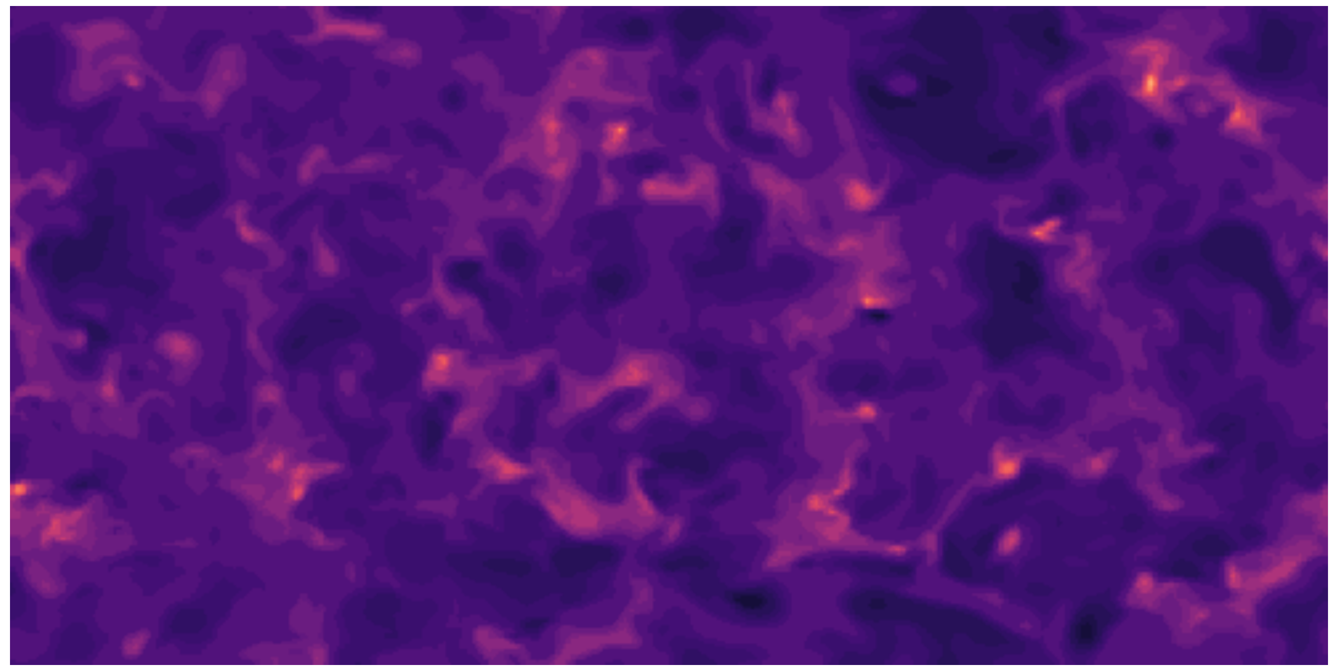
- **Pretrain** models on unlabeled massive datasets
- **Adapt** pretrained models to downstream tasks



Bommasani et al. 2021

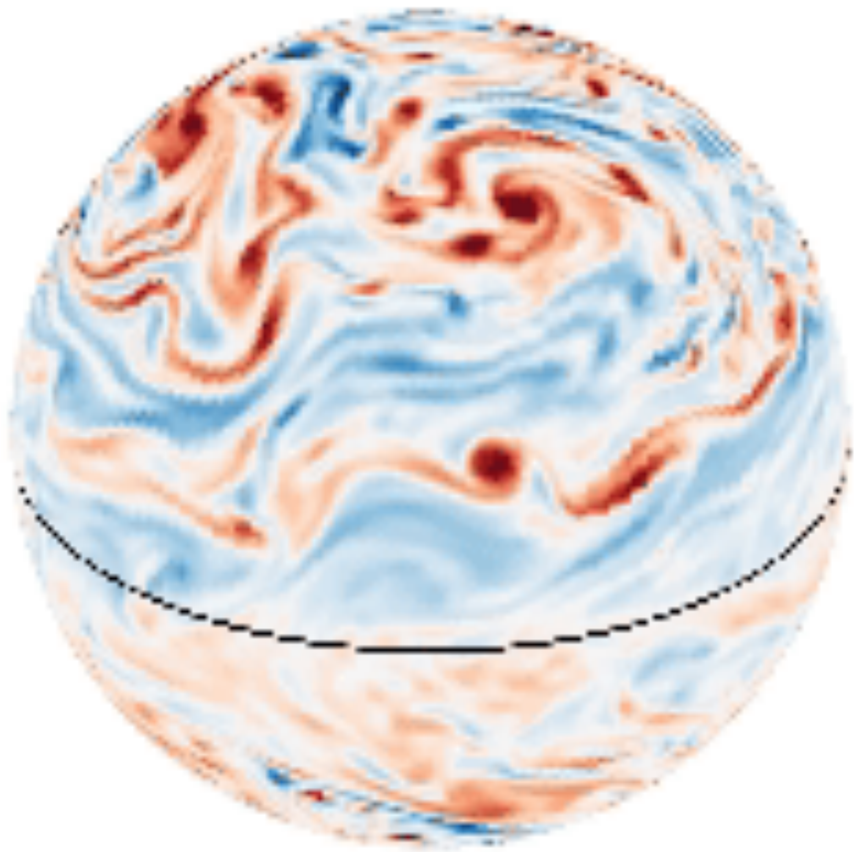
Towards Physical Foundation Models

Expensive simulations



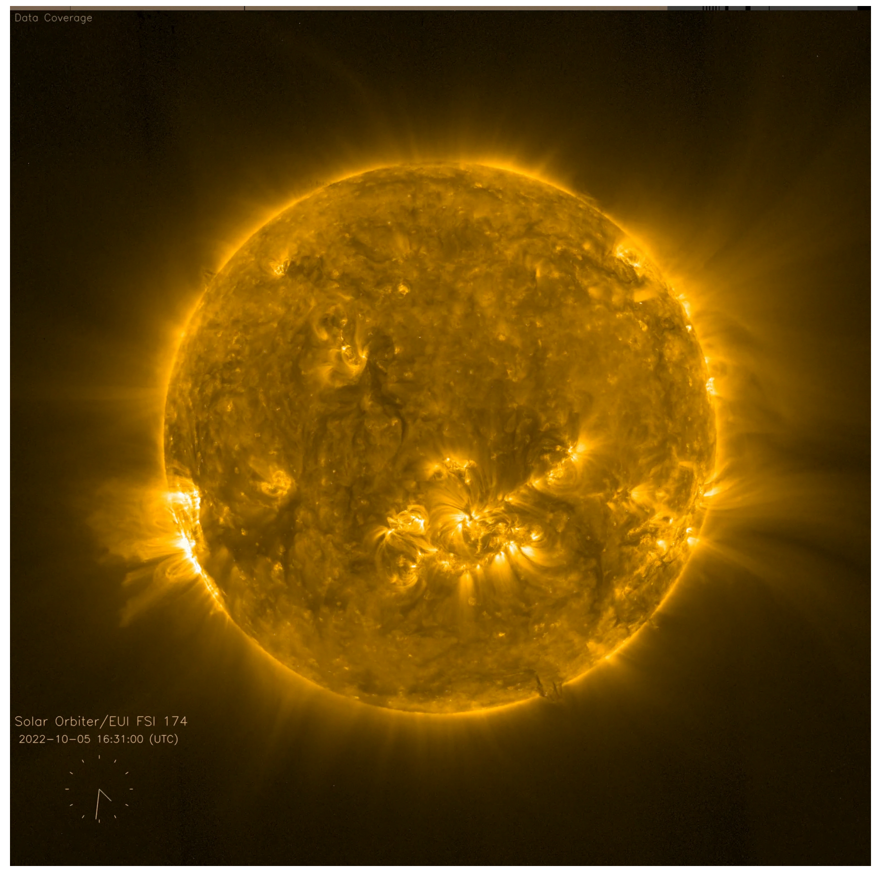
fine-tune

Weather forecast

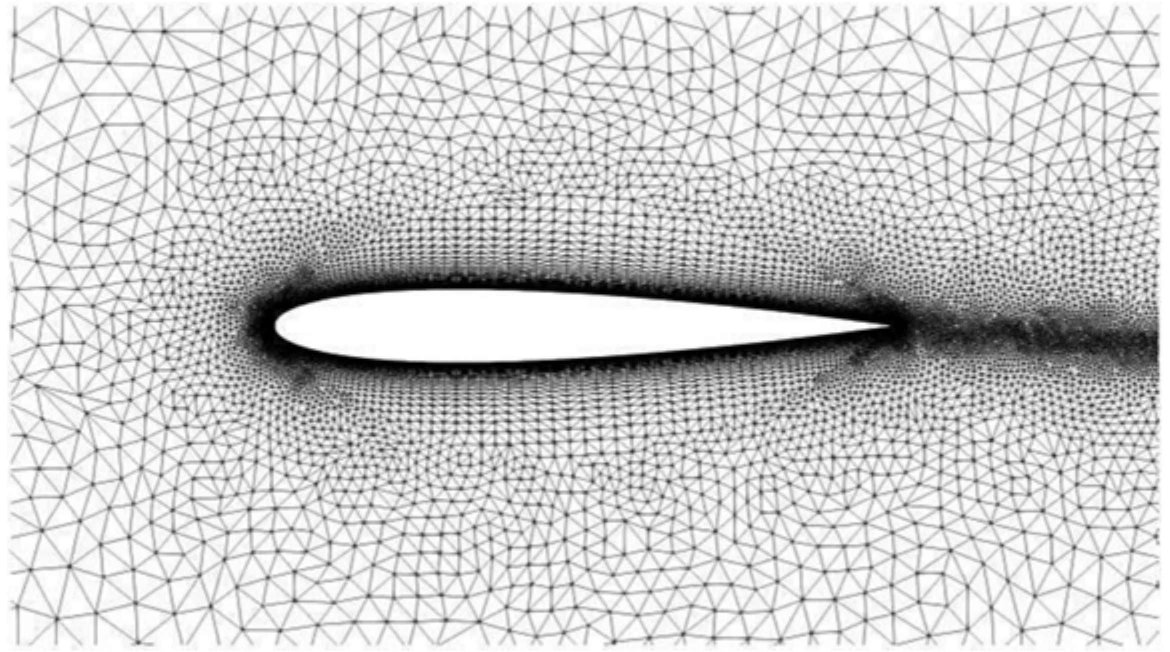


Foundation Model

Solar prediction

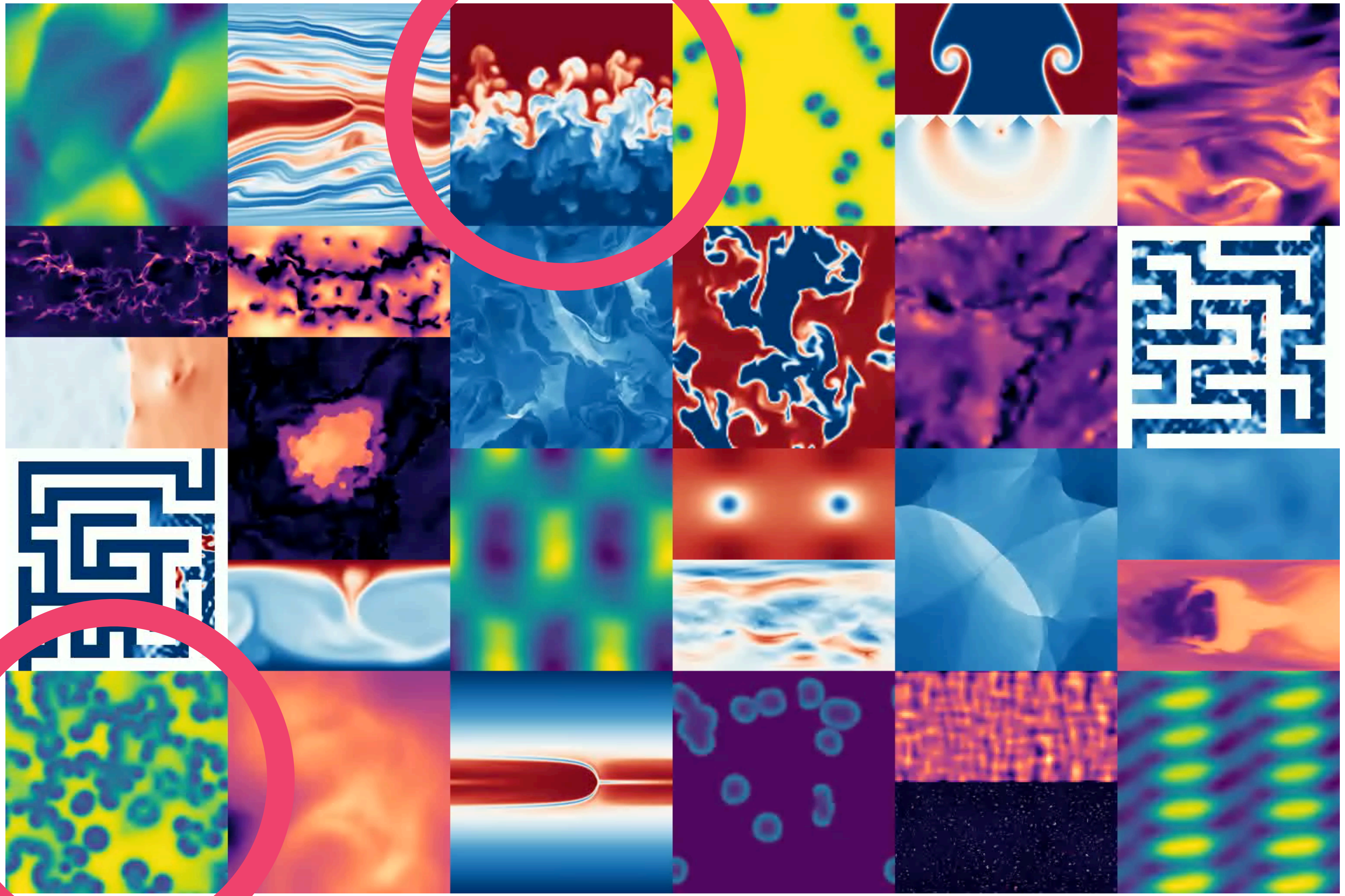


Non-uniform grids



training on multiple physics jointly!

Data variability



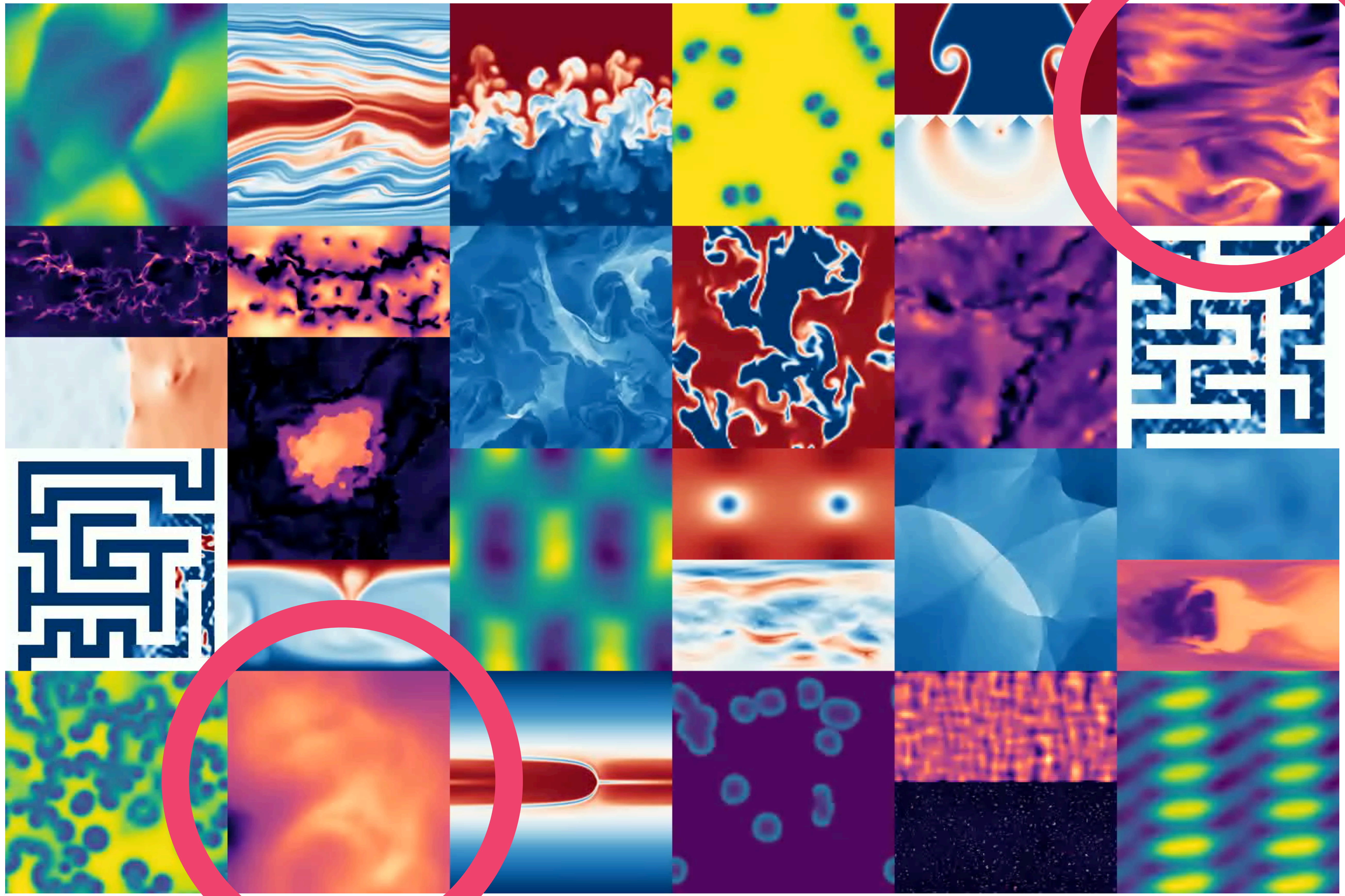
[*The Well*, R. Ohana*, et al. lead by Polymathic, 2024]

lvl1 PDE

lvl2 $\mathrm{PDE}_{\gamma,\eta}$ $\mathrm{PDE}_{\gamma',\eta'}$

lvl3 u_0 u_0'

Data variability



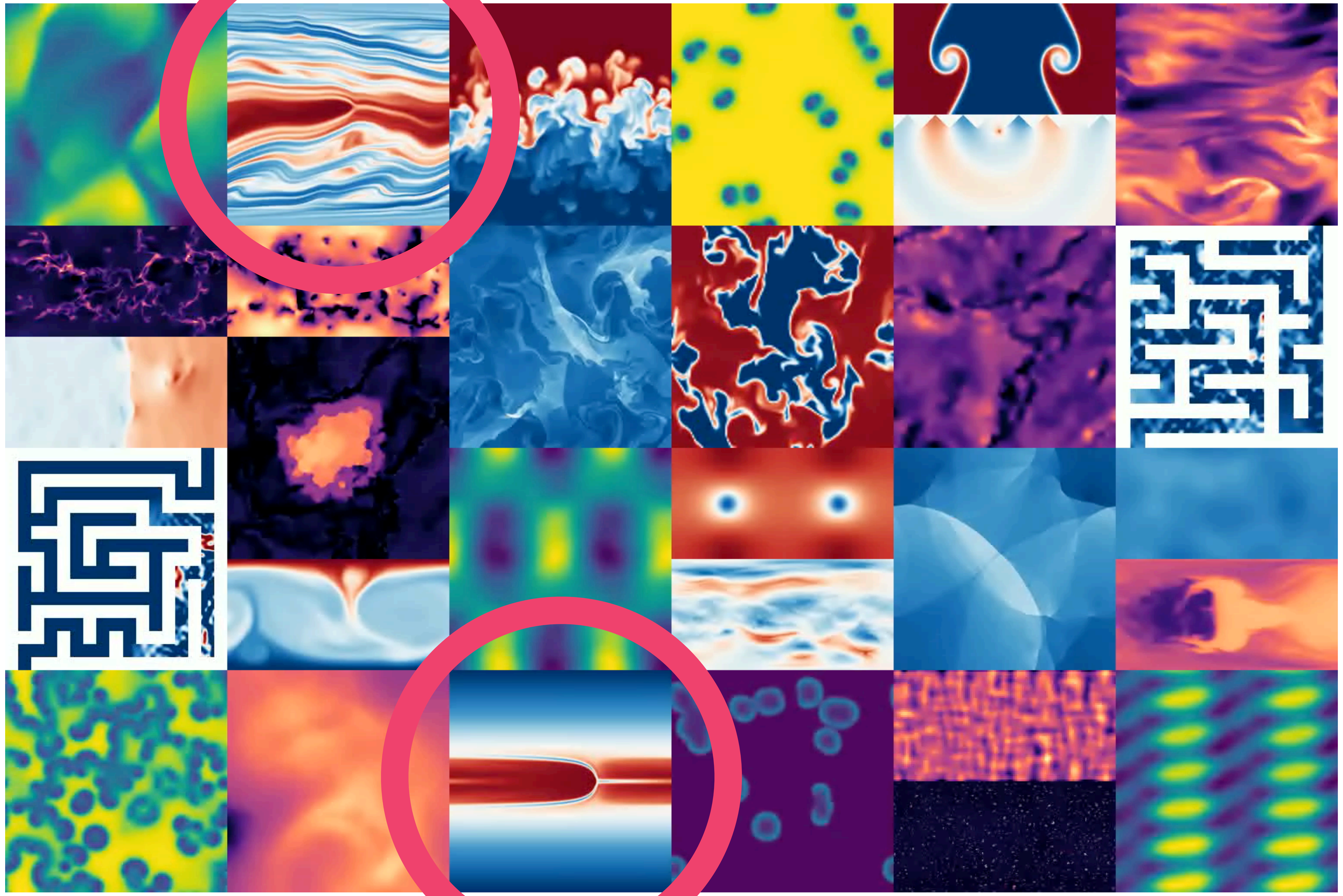
[*The Well*, R. Ohana*, et al. lead by Polymathic, 2024]

lvl1 $\quad$ PDE

lvl2 $\quad$ $\mathrm{PDE}_{\gamma,\eta}$ $\quad$ $\mathrm{PDE}_{\gamma',\eta'}$

lvl3 $\quad$ u_0 $\quad$ u_0'

Data variability



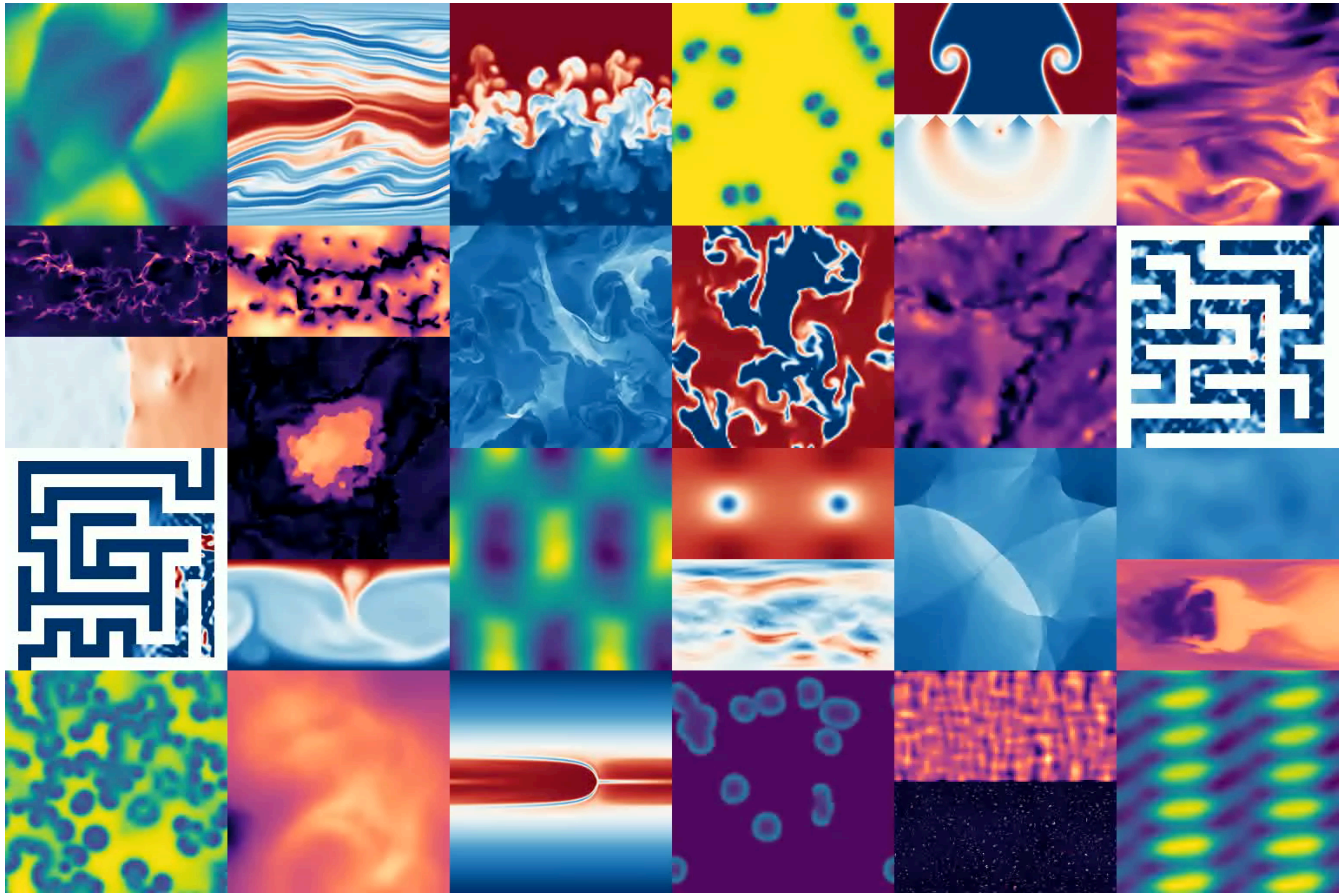
[*The Well*, R. Ohana*, et al. lead by Polymathic, 2024]

lvl1 PDE

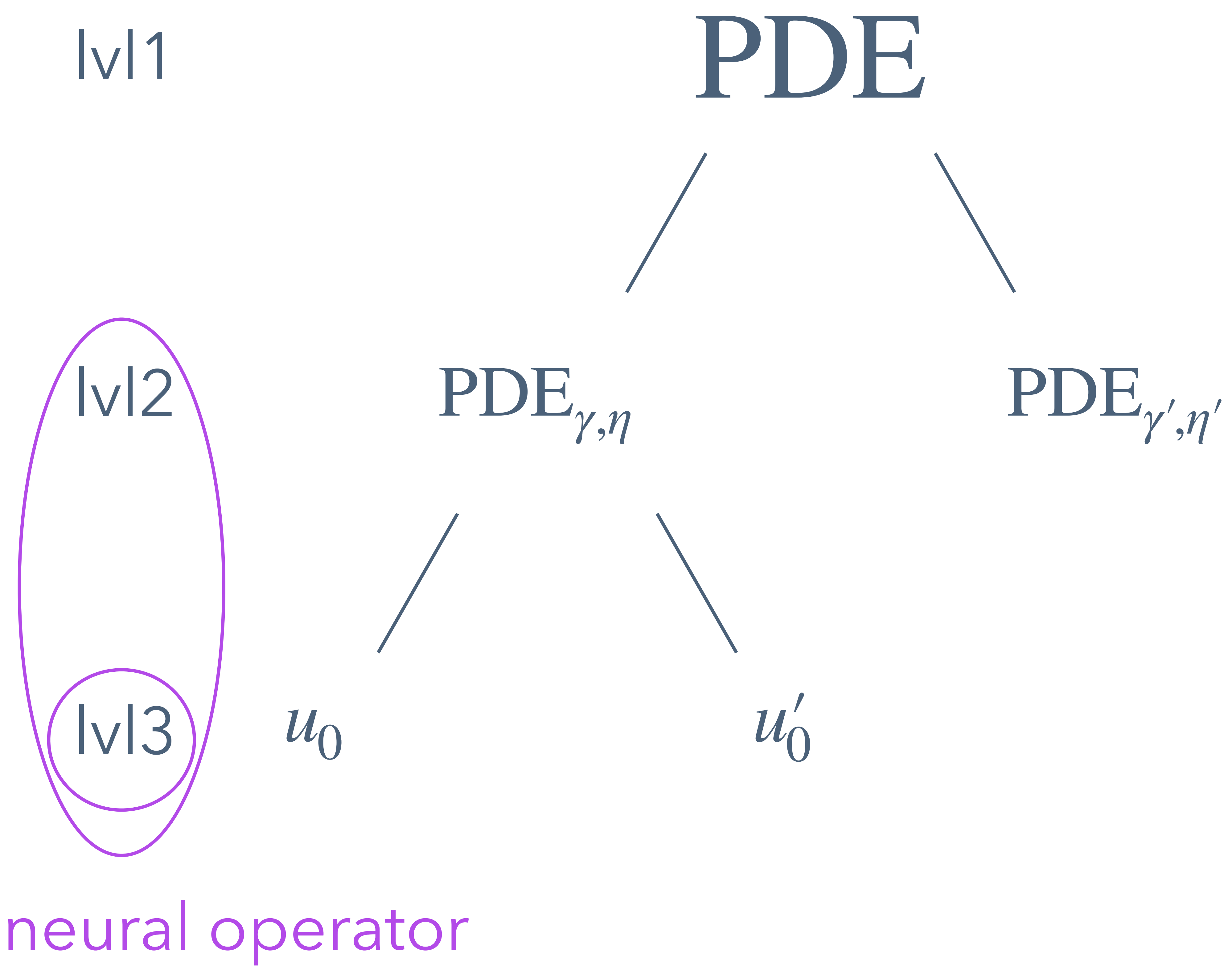
lvl2 $\mathrm{PDE}_{\gamma,\eta}$ $\mathrm{PDE}_{\gamma',\eta'}$

lvl3 u_0 u_0'

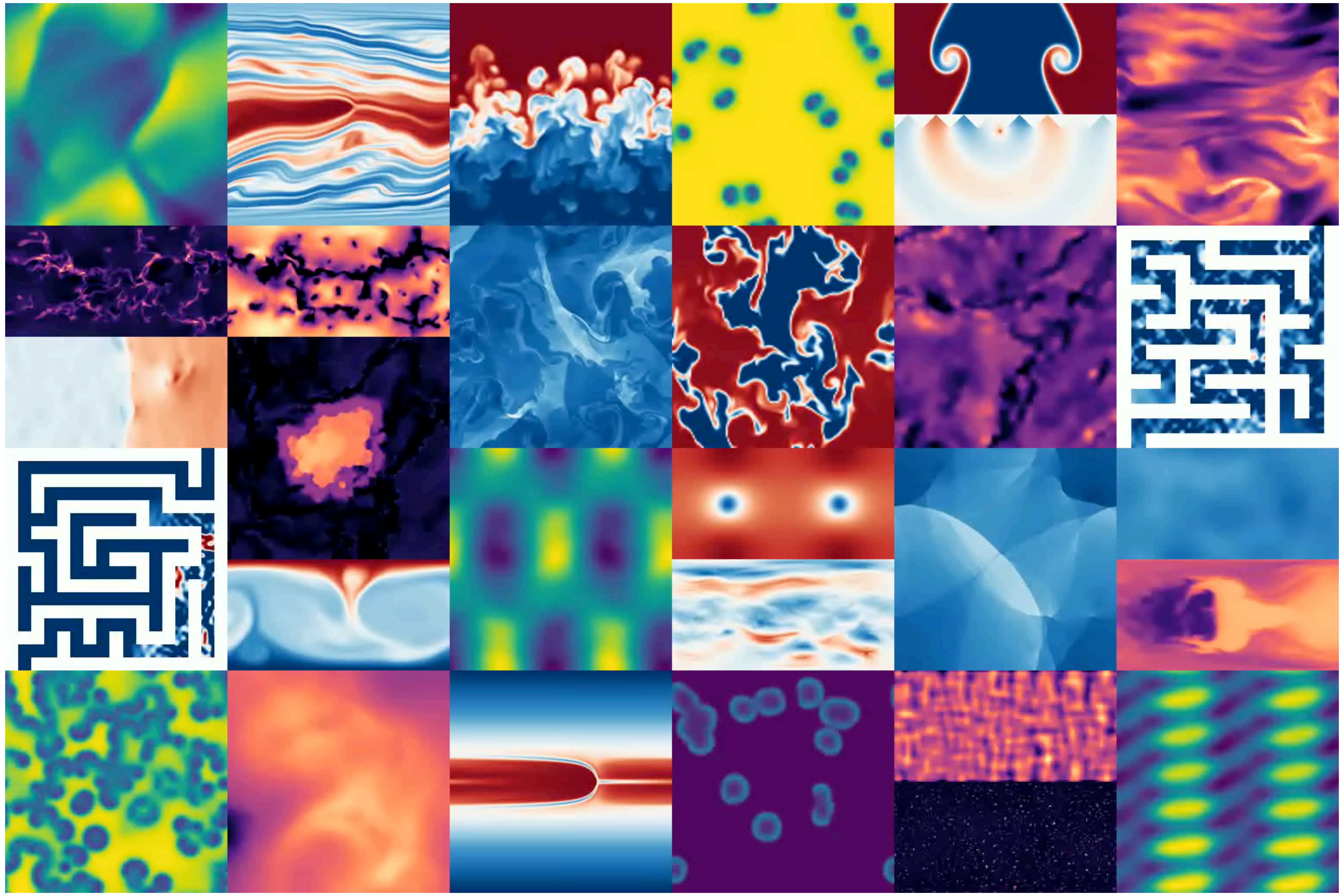
Data variability



[*The Well*, R. Ohana*, et al. lead by Polymathic, 2024]



Data variability



[*The Well*, R. Ohana*, et al. lead by Polymathic, 2024]

lvl1 PDE

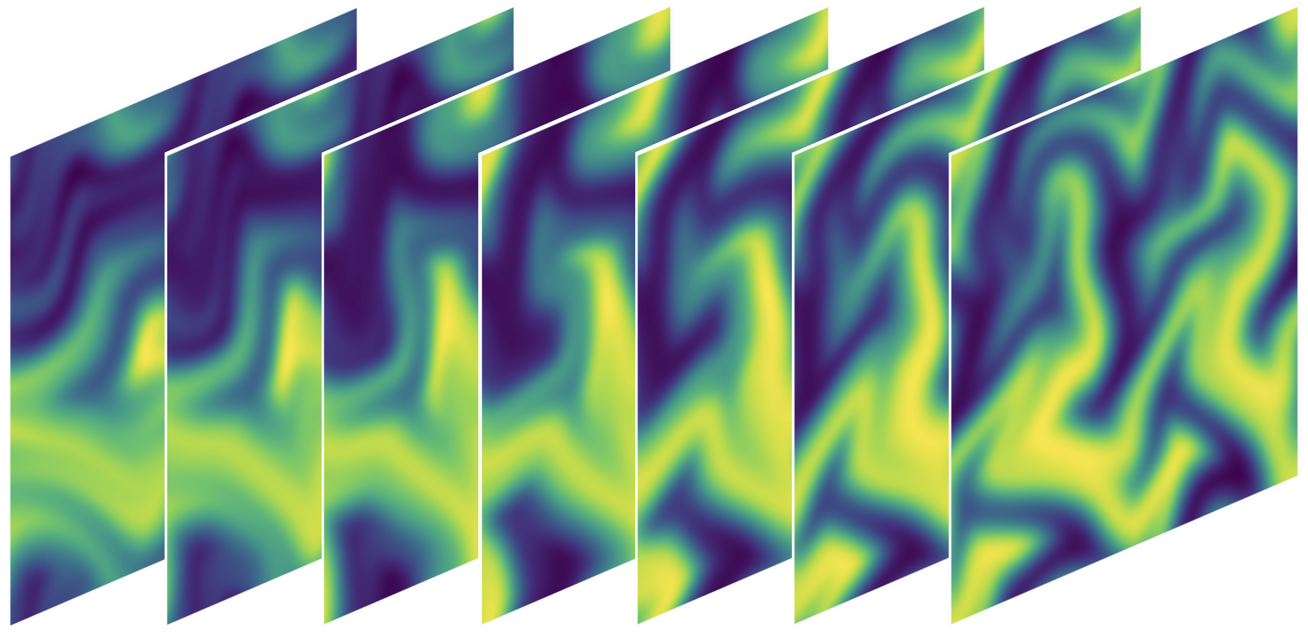
lvl2 $\mathrm{PDE}_{\gamma,\eta}$ $\mathrm{PDE}_{\gamma',\eta'}$

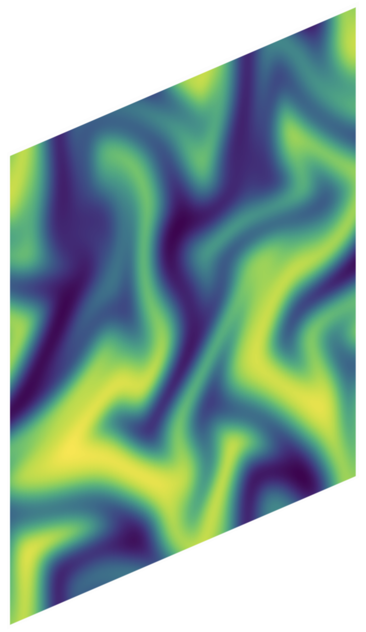
lvl3 u_0 u_0'

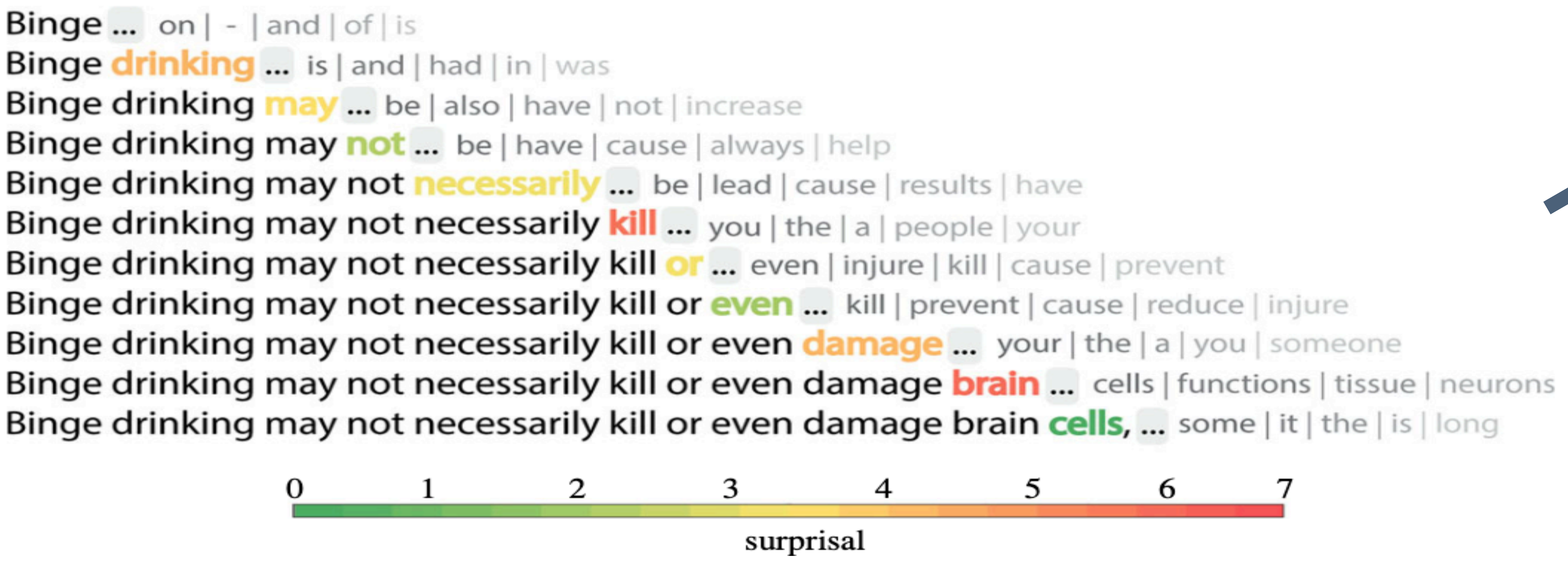
multi-physics-agnostic prediction

Multi-physics-agnostic prediction

Task: From a context of T states $(u_{t-T+1}, \ldots, u_t)$ predict the next state u_{t+1}







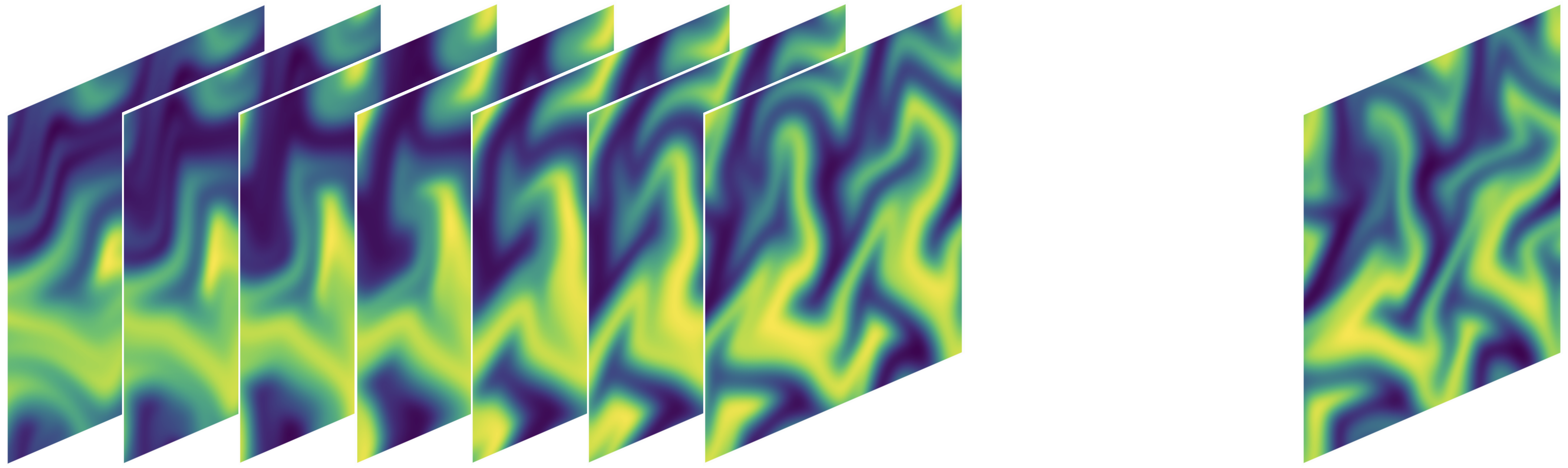
Task introduced in [Multiple Physics Pretraining, McCabe et al., Polymathic, 2024]

Transformer Neural Network

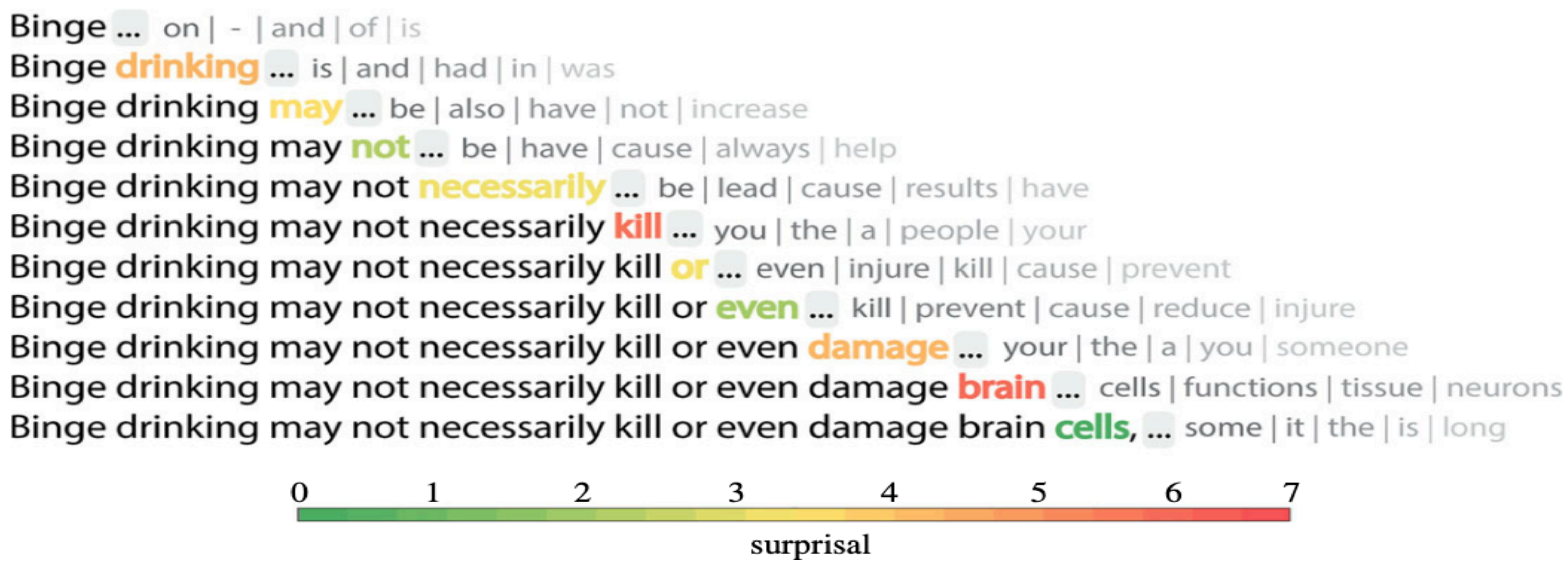
just like next token prediction

Difference between Language and Physics

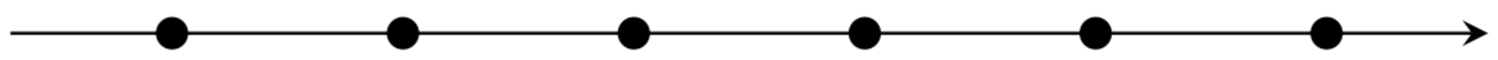
Task: From a context of T states $(u_{t-T+1}, \ldots, u_t)$ predict the next state u_{t+1}



inherently **continuous** in time



inherently **discrete** in time



How to solve the continuous-time Physics?

Learn the evolution operator, i.e. the "update rule"

$$\partial_t u_t(x) = f\left(u_t(x), \nabla u_t(x), \nabla^2 u_t(x)\right)$$

Numerical solver

time discretization

$$u_{t+\Delta t} \approx u_t + f\left(u_t, \nabla u_t, \nabla^2 u_t\right) \Delta t$$

space discretization

$$u_{t+\Delta t} \approx u_t + f_\theta\left(u_t\right) \Delta t \quad \left(e.g.\ u'(x) \approx \frac{u(x+\Delta x) - u(x-\Delta x)}{2\Delta x}\right)$$

Neural solver -> learn θ from data

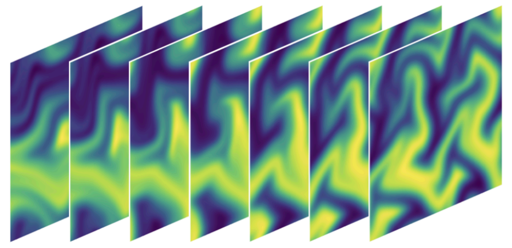
[Neural ODE, Chen et al., 2018]
[Bar-Sinai, Hoyer et al. 2019]
[Brandstetter, Worrall, Welling, 2022]

Task: From a context of T states $(u_{t-T+1}, \ldots, u_t)$ predict the next state u_{t+1}

→ How to obtain an operator f_θ for each context $(u_{t-T+1}, \ldots, u_t)$?

How to obtain an operator f_θ for each context $(u_{t-T+1}, \ldots, u_t)$?

1. Gradient adaptation. Learn an operator network f_θ every new context

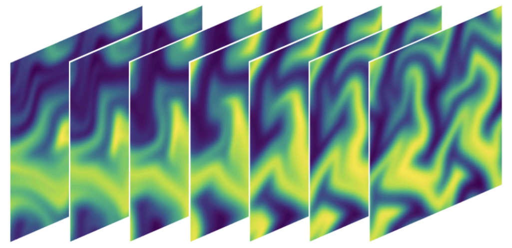


gradient descent on θ $\longrightarrow$ $u_{t+\Delta t} \approx u_t + f_\theta\left(u_t\right) \Delta t$

Very costly

How to obtain an operator f_θ for each context $(u_{t-T+1}, \ldots, u_t)$?

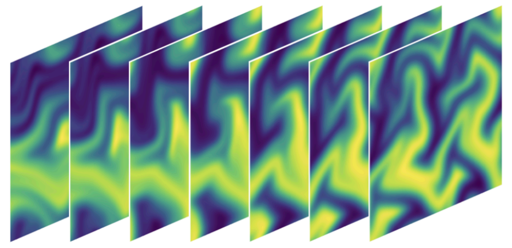
1. Gradient adaptation. Learn an operator network f_θ every new context



gradient descent on ξ $\longrightarrow$ $u_{t+\Delta t} \approx u_t + f_\theta\left(u_t\right)\Delta t \quad \theta = \theta_c + W\xi$ *[GEPS, Koupaï et al. 2024]*

Very costly

2. DISCO: output the operator itself

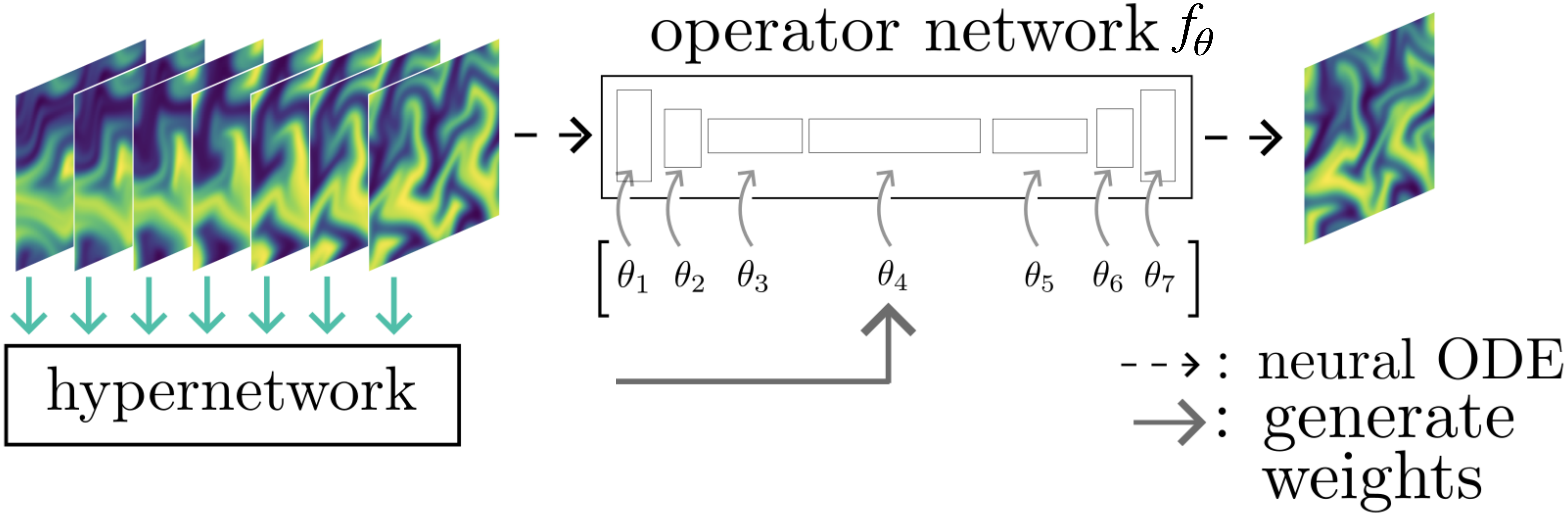


single forward $\longrightarrow$ θ

Rudy Morel

Edouard Oyallon

DISCO: learning to DISCover an evolution Operator from data



DISCO (ICML, 2025):

- decouple parameter estimation from state evolution
- enforce an "information bottleneck" in the operator: intrinsic $\dim \theta = 384$

DataSets

PDEBench + The Well: 1D, 2D, 3D, different resolution/quantities/boundary conditions

lvl1 lvl2 lvl3

Table 1. The datasets from PDEBench (Takamoto et al., 2022) and The Well (Ohana et al., 2024) used in this paper.

Dataset Name	Physical Dimension	# of Fields	Resolution (Time)	Resolution (Space)	Boundary Conditions
Burgers	1D	1	200	1024	periodic
Shallow water equation	2D	1	100	128×128	open
Diffusion-reaction	2D	2	100	128×128	Neumann
Incomp. Navier-Stokes (INS)	2D	3	1000	512×512	Dirichlet
Comp. Navier-Stokes (CNS)	2D	4	21	512×512	periodic
active matter	2D	11	81	256×256	periodic
Euler multi-quadrants	2D	5	100	512×512	periodic / open
Gray-Scott reaction-diffusion	2D	2	1001	128×128	periodic
Rayleigh-Bénard	2D	4	200	512×128	periodic $\times$ Dirichlet
shear flow	2D	4	200	256×512	periodic
turbulence gravity cooling	3D	6	50	$64 \times 64 \times 64$	open
MHD	3D	7	100	$64 \times 64 \times 64$	periodic
Rayleigh-Taylor instability	3D	4	120	$64 \times 64 \times 64$	periodic $\times$ periodic $\times$ slip
supernova explosion	3D	6	59	$64 \times 64 \times 64$	open

State-of-the-art prediction on PDEBench

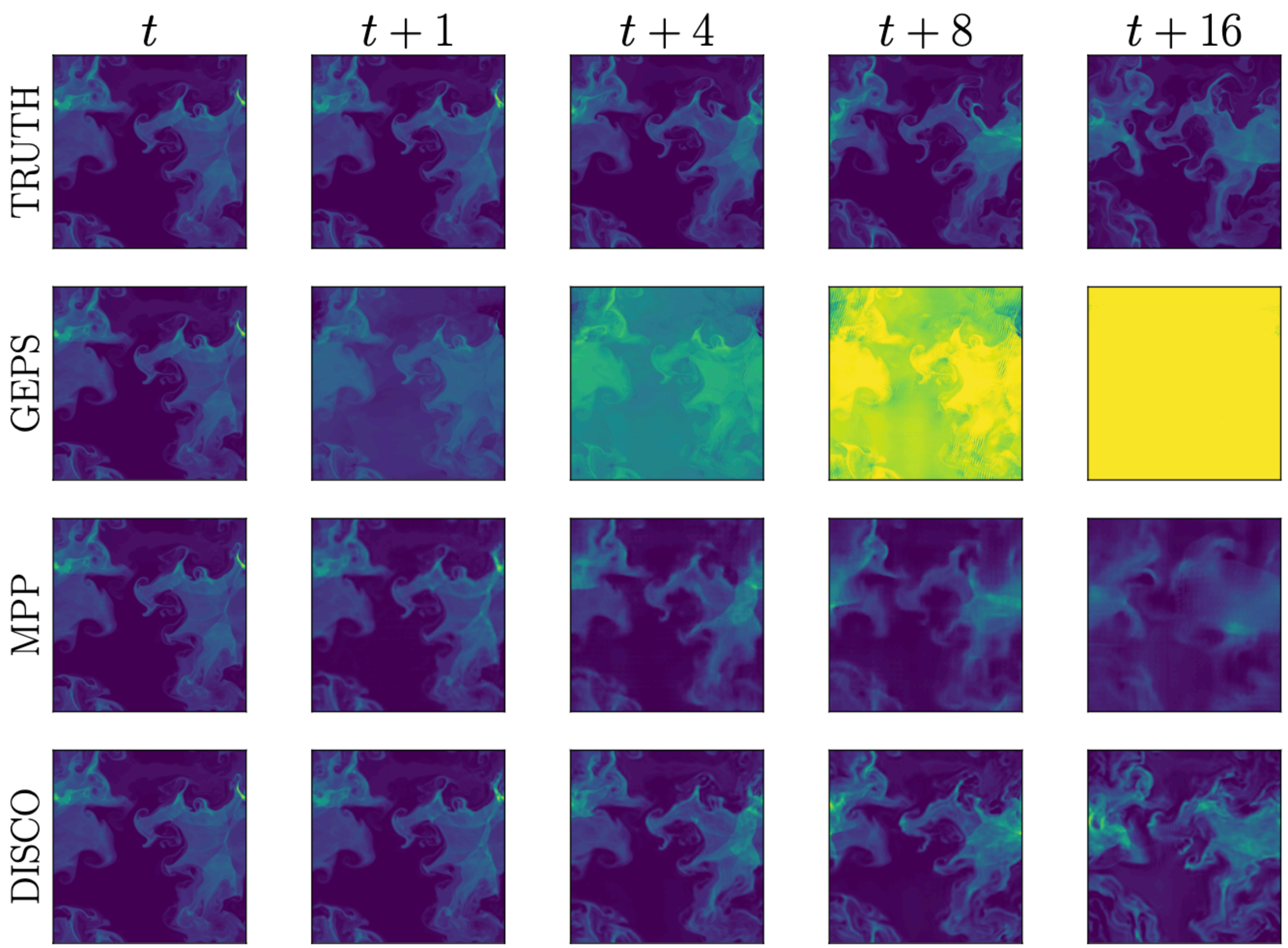
(a) Burgers (b) Shallow water (c) Diff.-reaction (d) Incomp. NS (e) Comp. NS (f) Comp. NS (alone)

Table 5. **Number of epochs to reach SOTA performance on PDEBench datasets.**

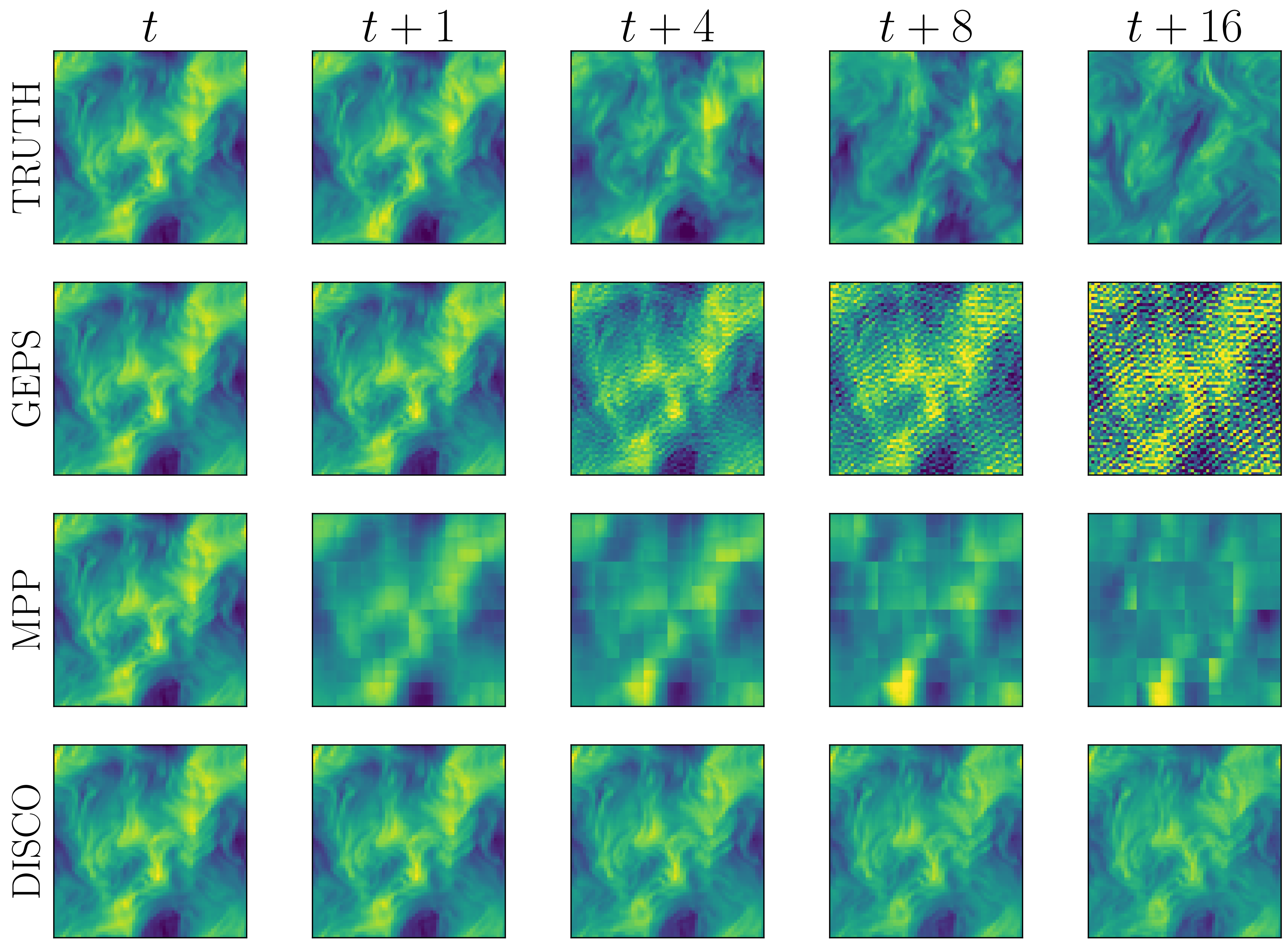
Model	# parameters	Burgers	SWE	DiffRe2D	INS	CNS	CNS (alone)
MPP (retrained)	160m	500	500	500	500	**500**	50
DISCO (ours)	119m	**277**	**70**	**55**	**35**	-	**7**

→ requires far fewer epochs on most datasets

Rollout trajectories on the Well dataset

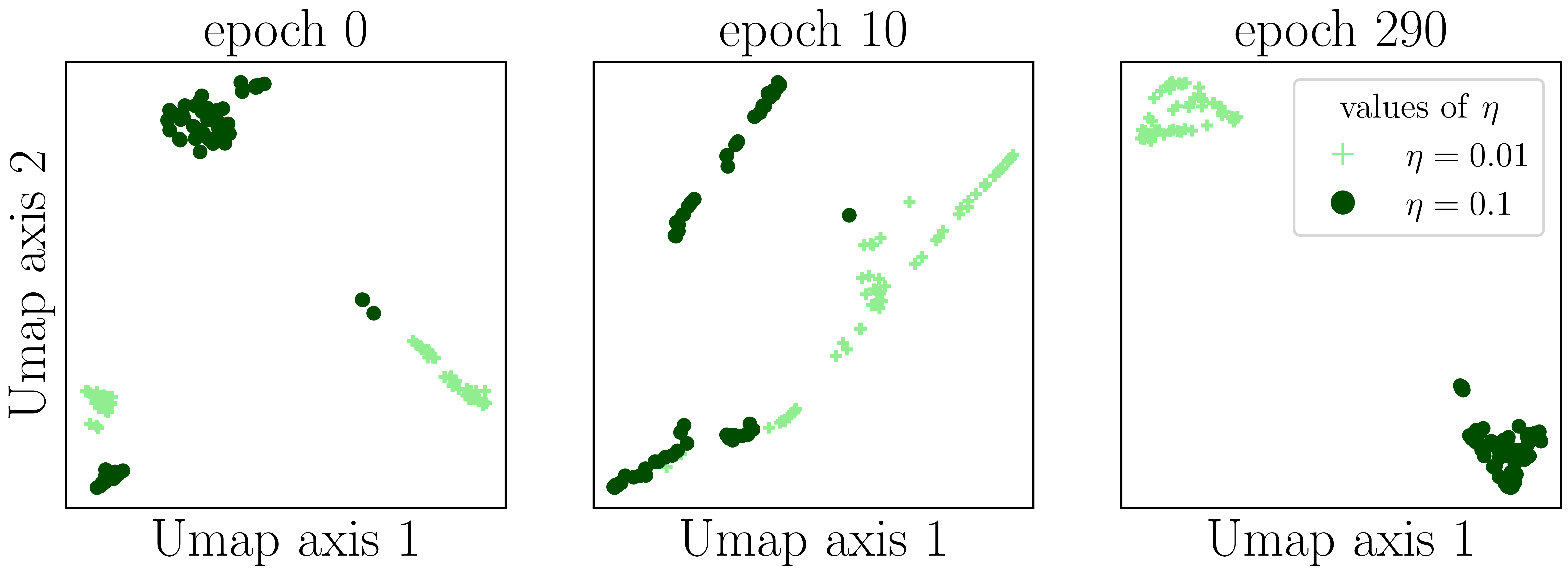


Euler



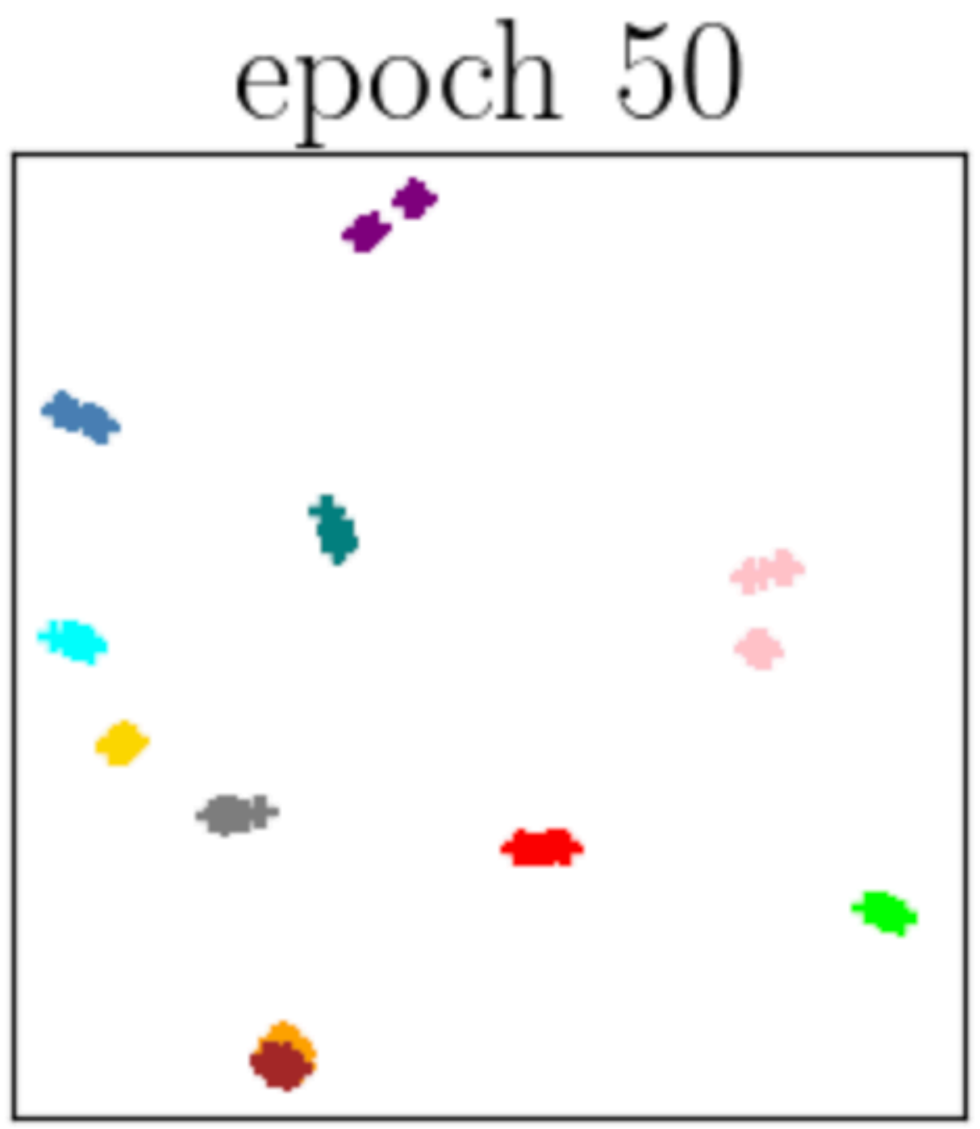
2d slices of 3d MHD

A shared latent space for Physics

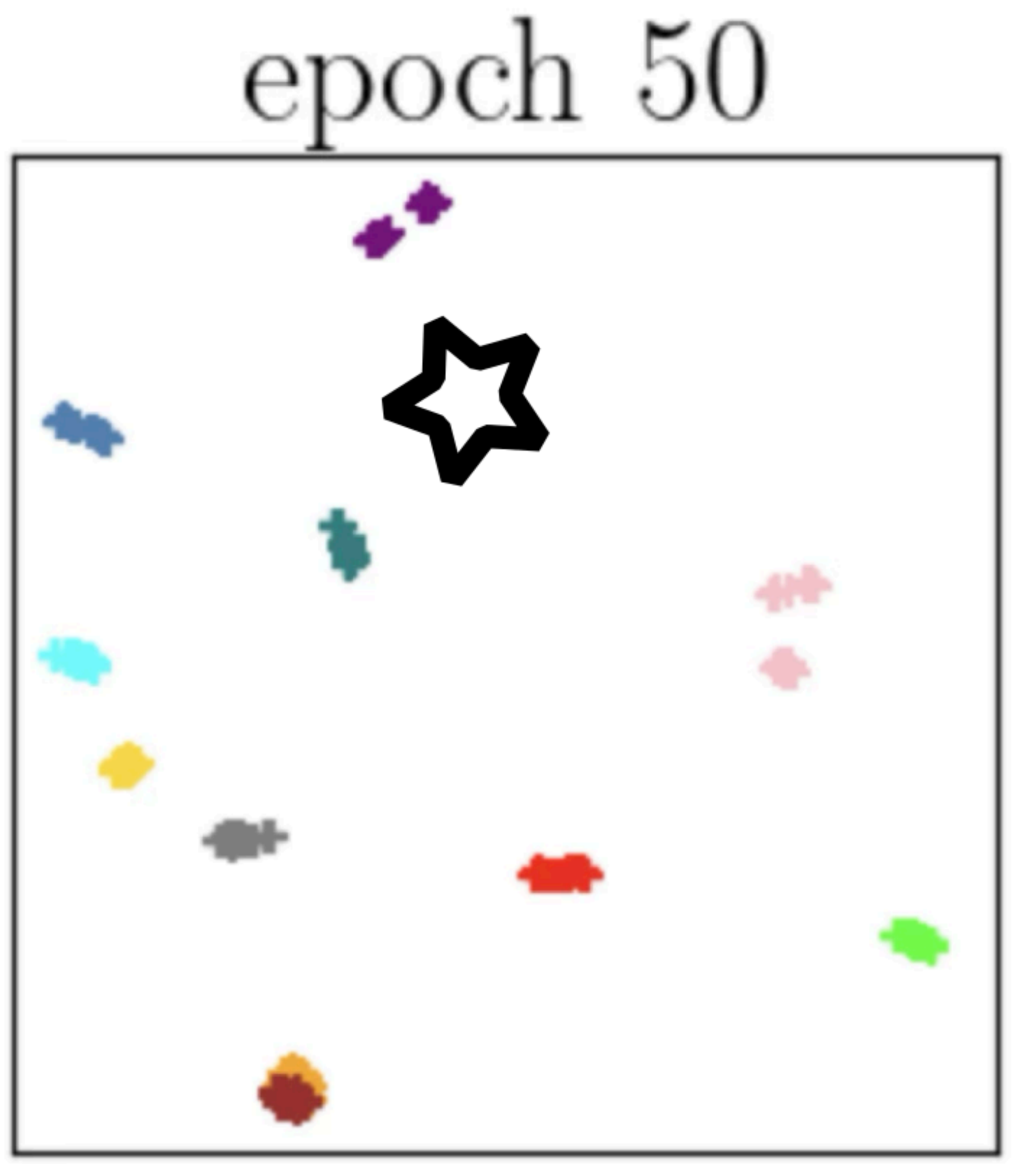


Space of the evolution operators

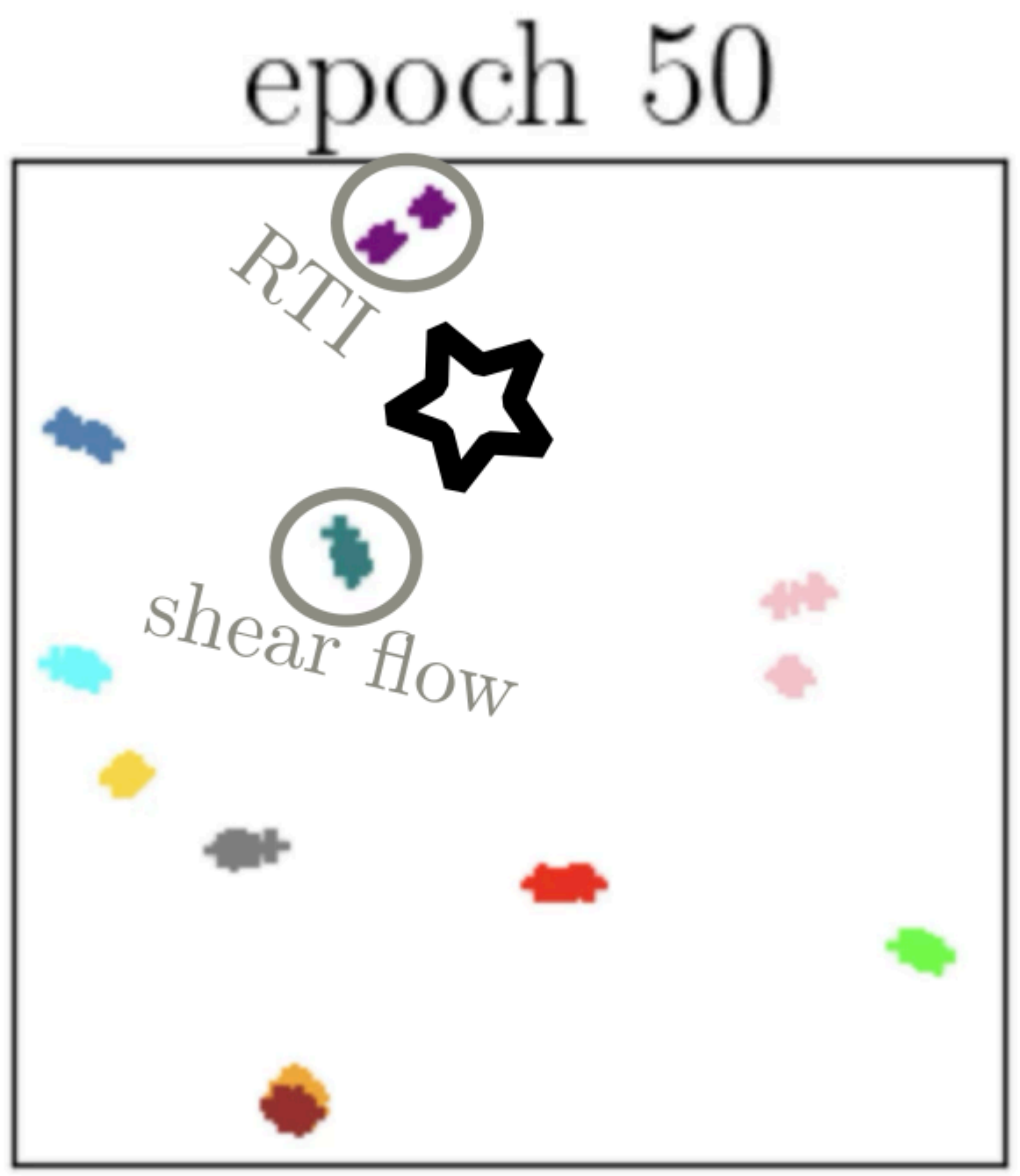
A shared latent space for Physics



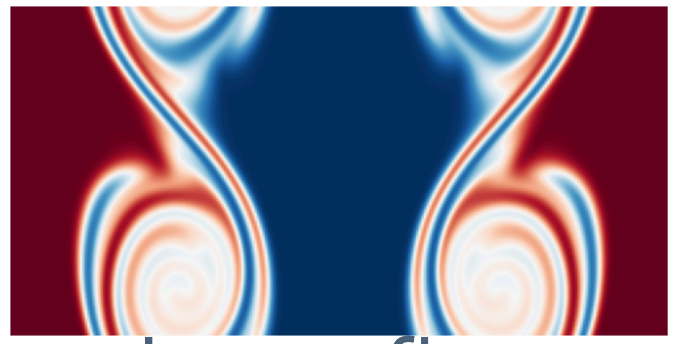
pretrain



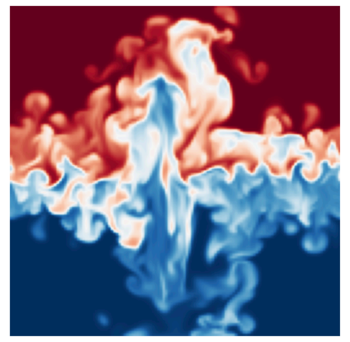
project your unseen physics



finetune on



shear flow



RTI

Takeaway

1. Multi-physics-agnostic prediction with multi-level variability is essential for generalization

2. Physical evolution is inherently continuous in time

3. DISCO exploits this inductive bias while harnessing the power of transformers

4. The latent space in DISCO offers interpretability

paper

code