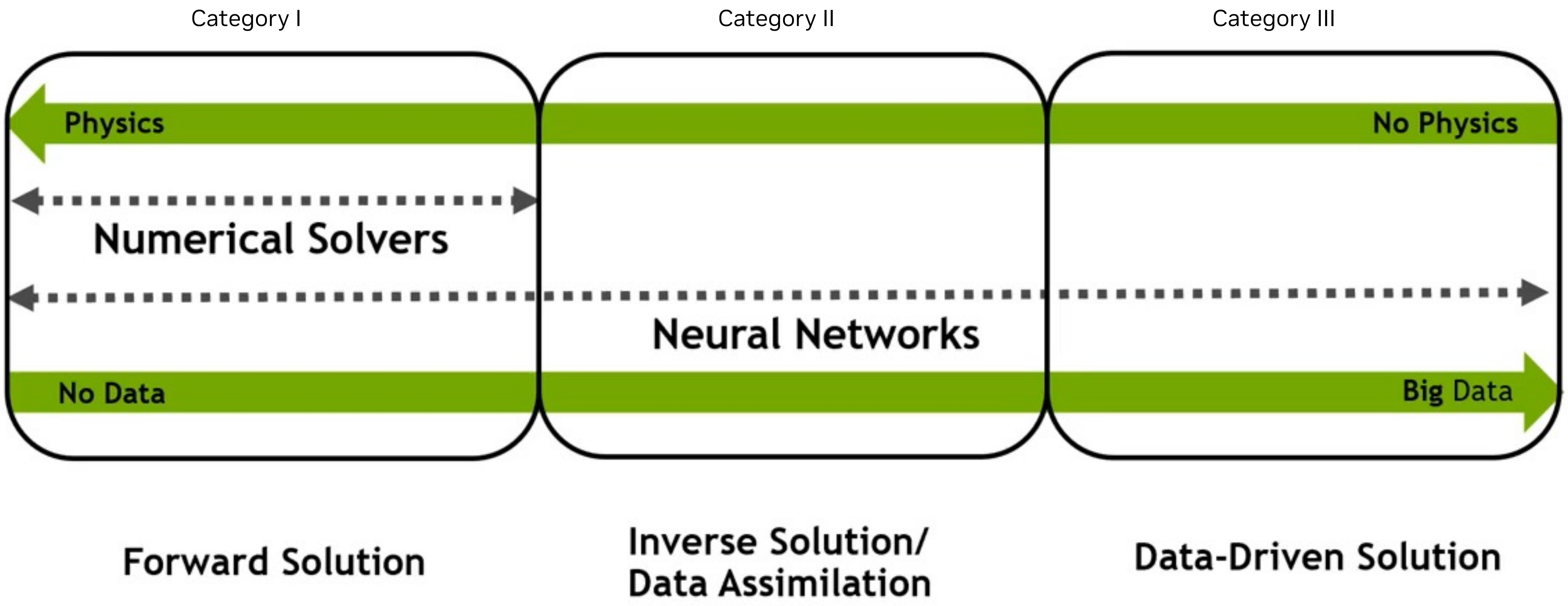
# Lab 2: Data-Driven

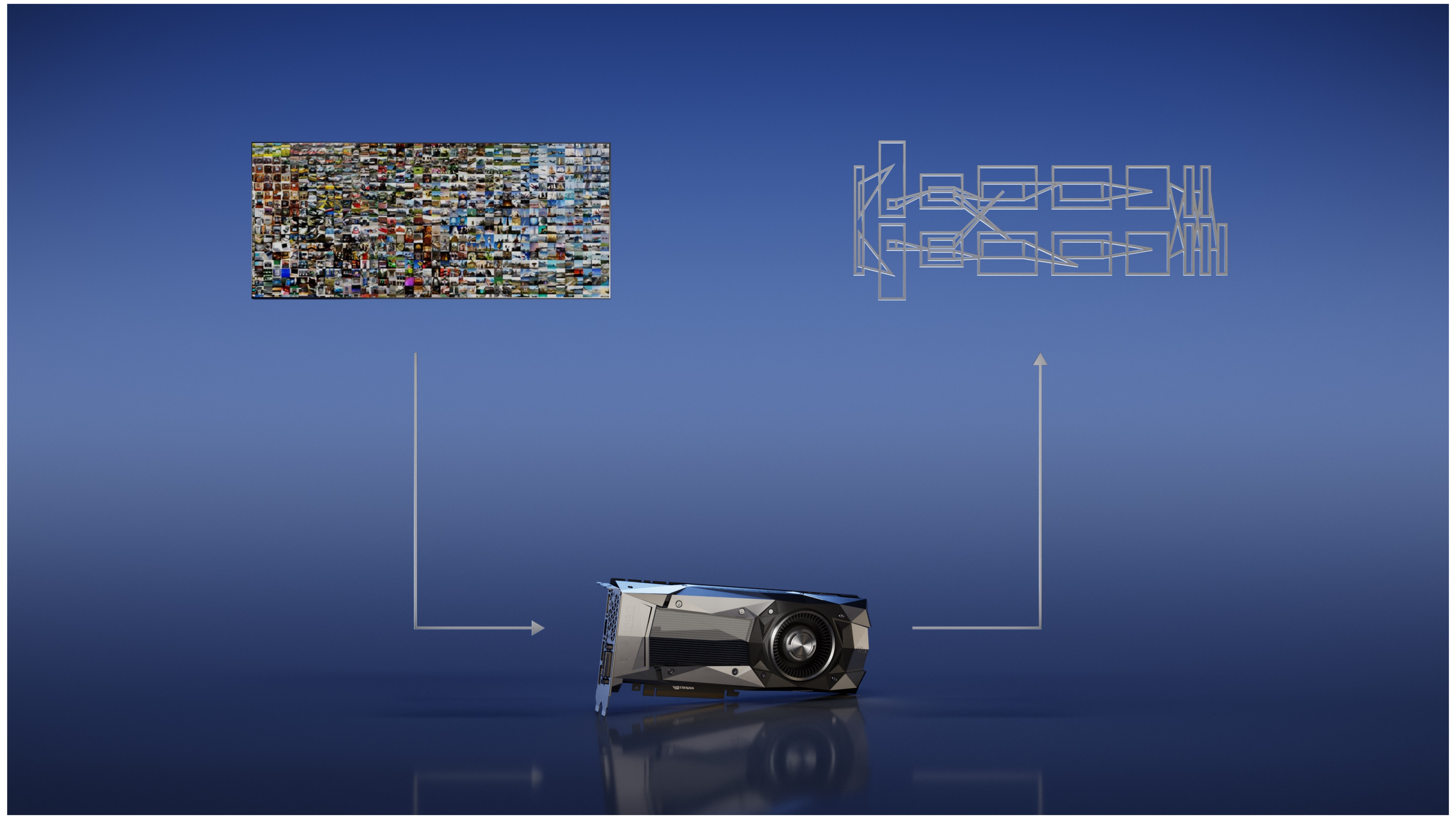
Jay Chen | Data Scientist, May 2023.

# Modulus

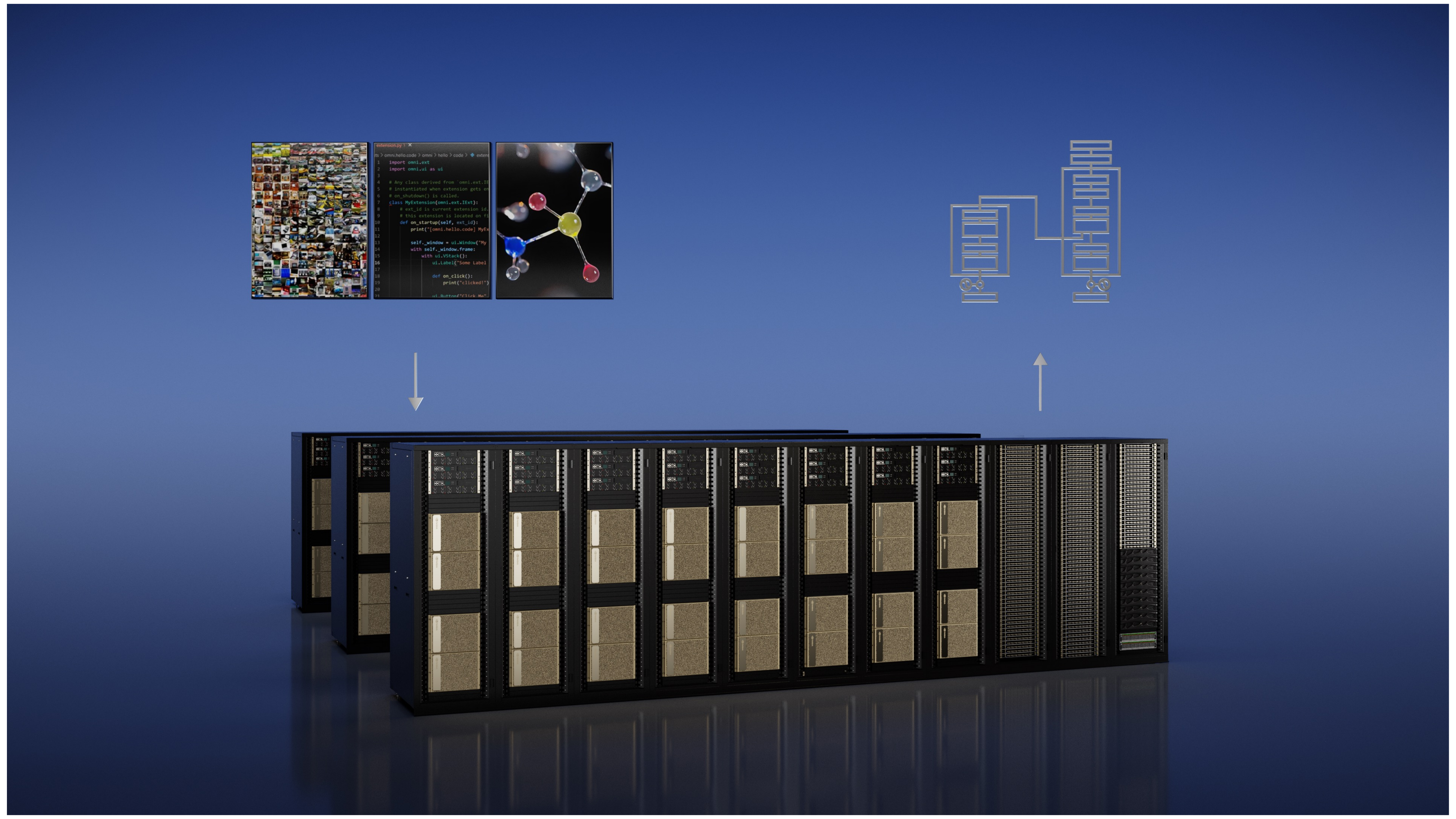



NVIDIA

## AlexNet: Big Bang Moment of AI



## ChatGPT: iPhone Moment of AI



**AlexNet, 2012**

Alex Krizhevsky

Ilya Sutskever

Geoffrey Hinton

10 years →

**GPT-3, ChatGPT, 2022**

OpenAI

Ilya Sutskever

And other brilliant researchers

**61M** Parameters

**262** PetaFLOPS

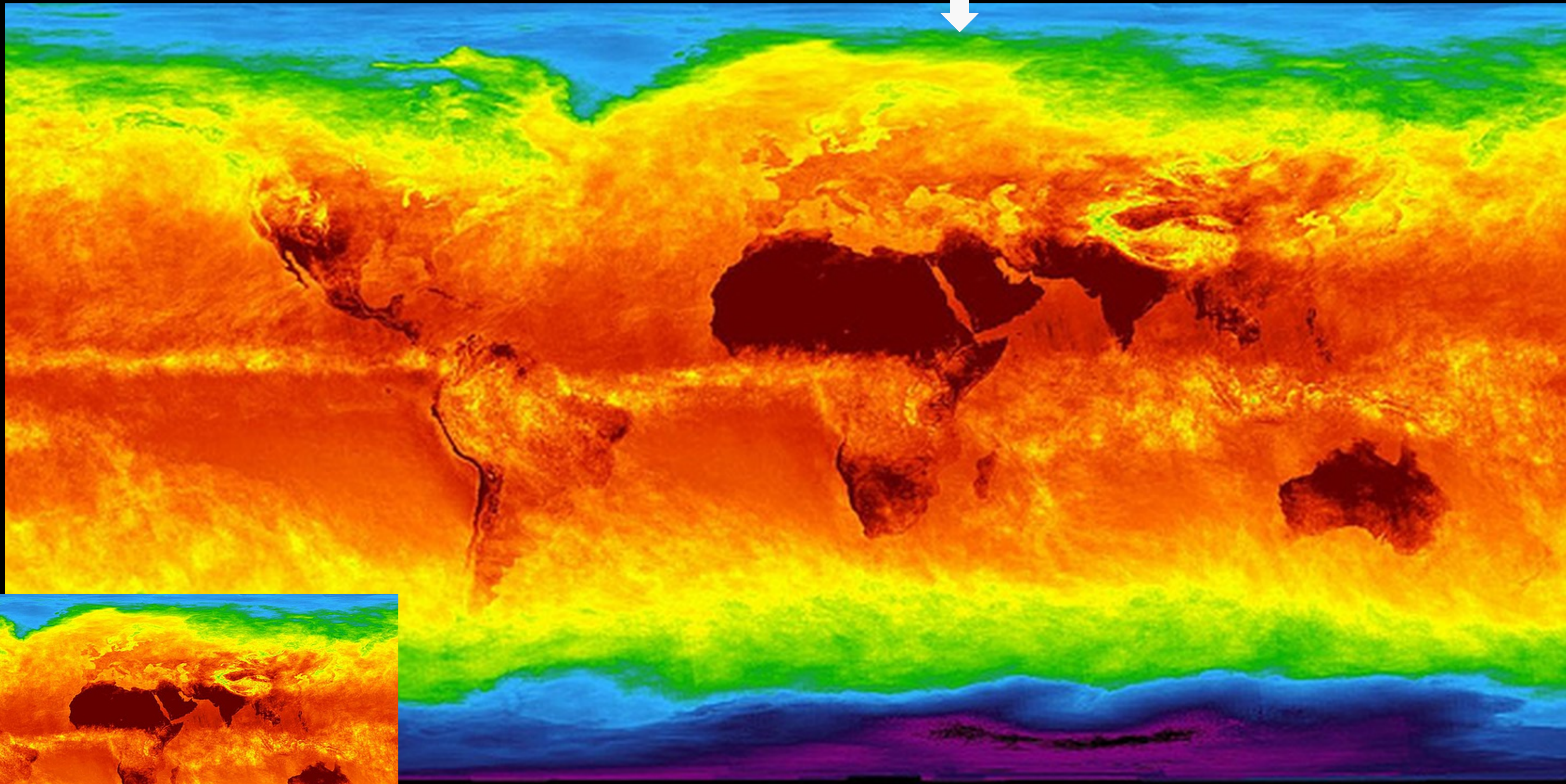
1000 X →

**175B** Parameters

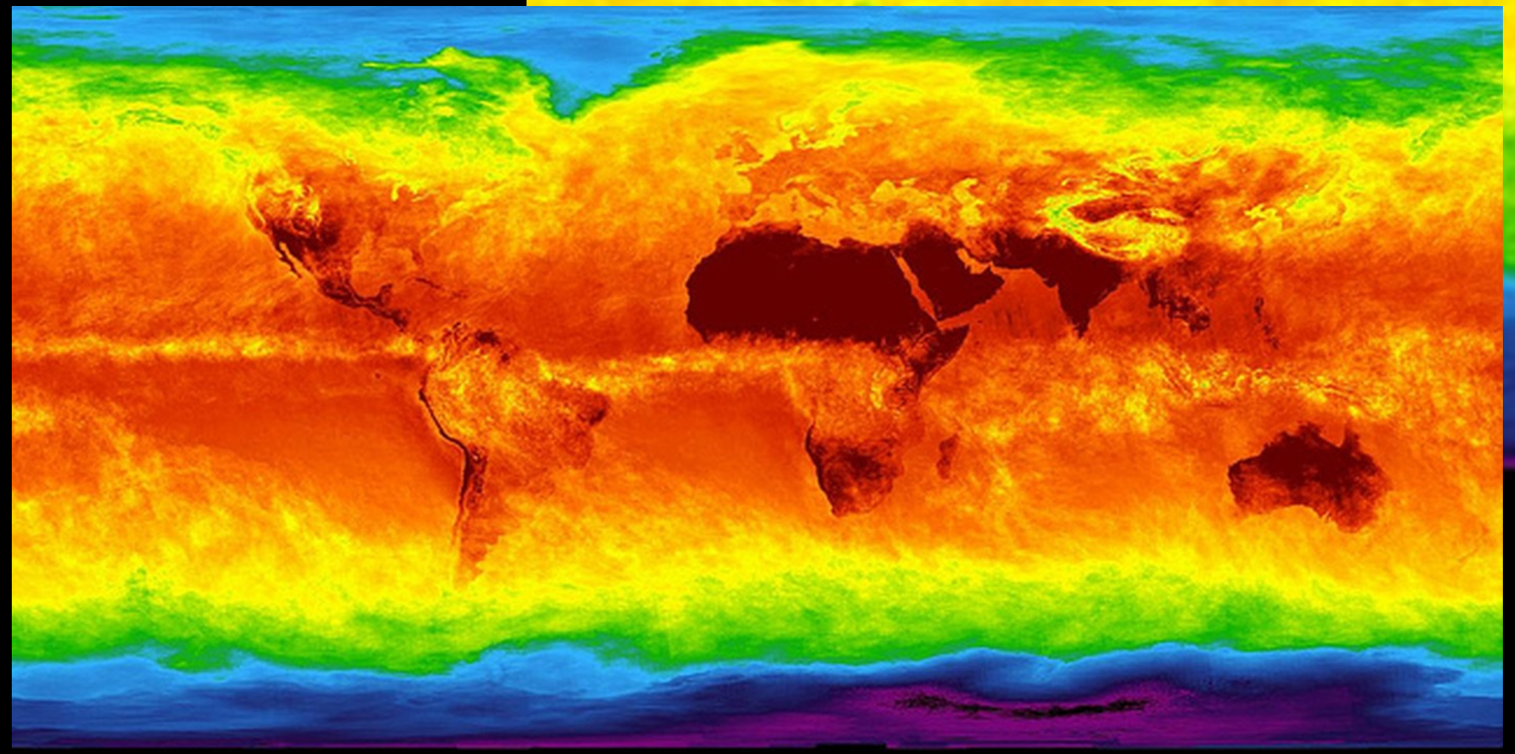
**323** ZettaFLOPS

<u>OpenAI's Believe "Bigger is Better"</u>
by Ilya Sutskever

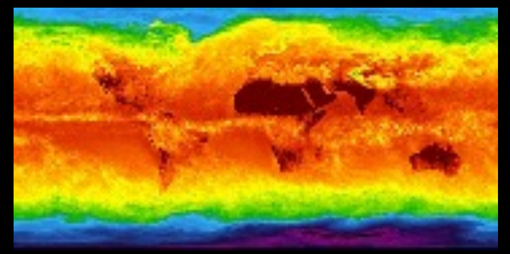
# FourCastNet: NV's DDWP, first to be trained at ambitious 0.25-deg global resolution



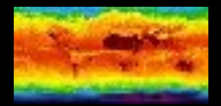
FourCastNet, Pathak et al. (2022), 0.25°, ~1,000,000 Pixels, ViT+AFNO



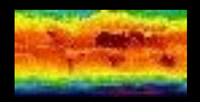
GNN, Keisler et al. (2022), 1°, 64,000 Pixels, Graph Neural Networks



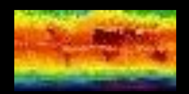
DLWP, Weyn et al. (2020). 2°, 16K pixels, Deep CNN on Cubesphere/(2021) ResNet



Weyn et al. (2019), 2.5° N.H only, 72x36, 2.6k pixels, ConvLSTM



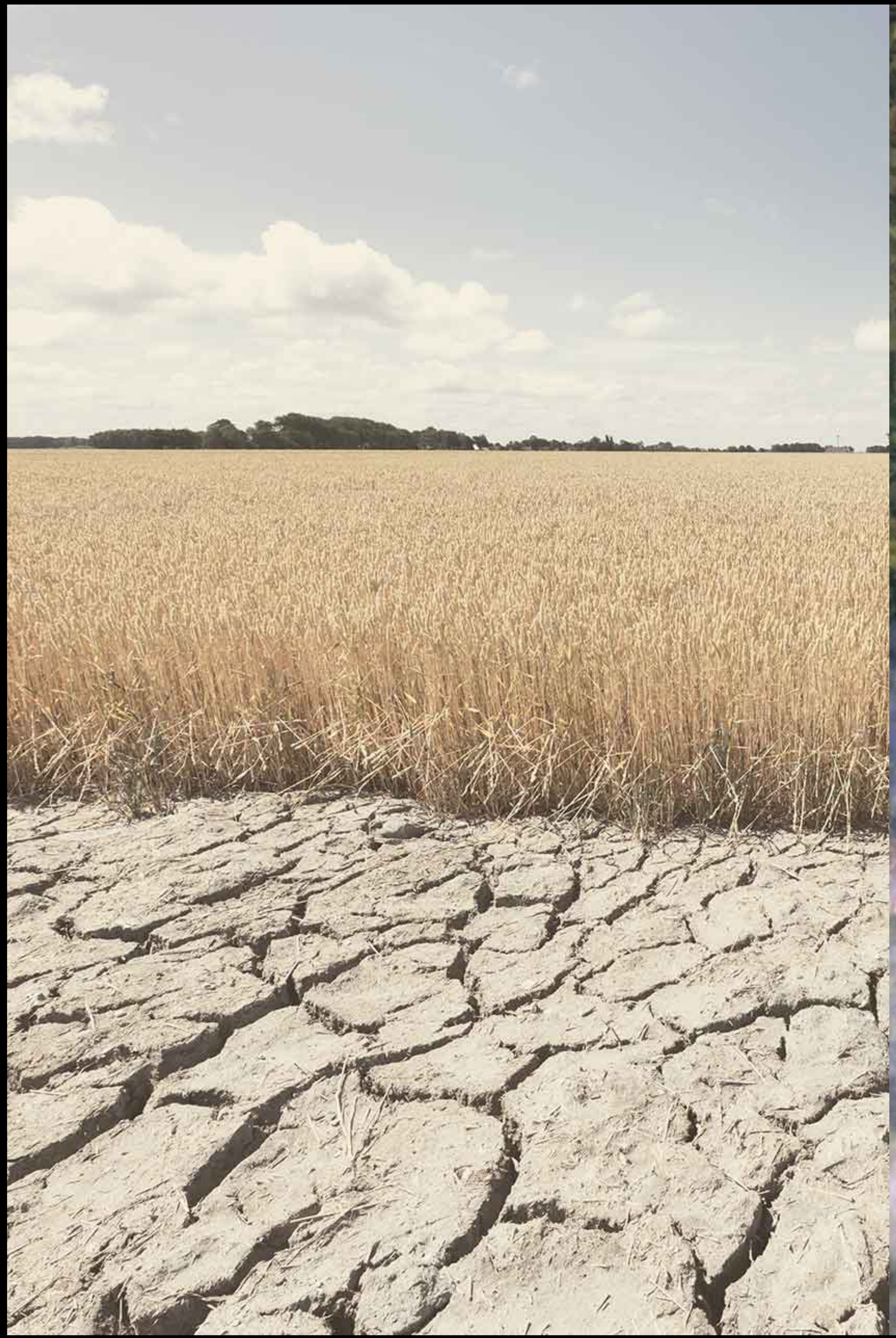
WeatherBench, Rasp et al. (2020). 5.625°, 64x32, 2K pixels, CNN



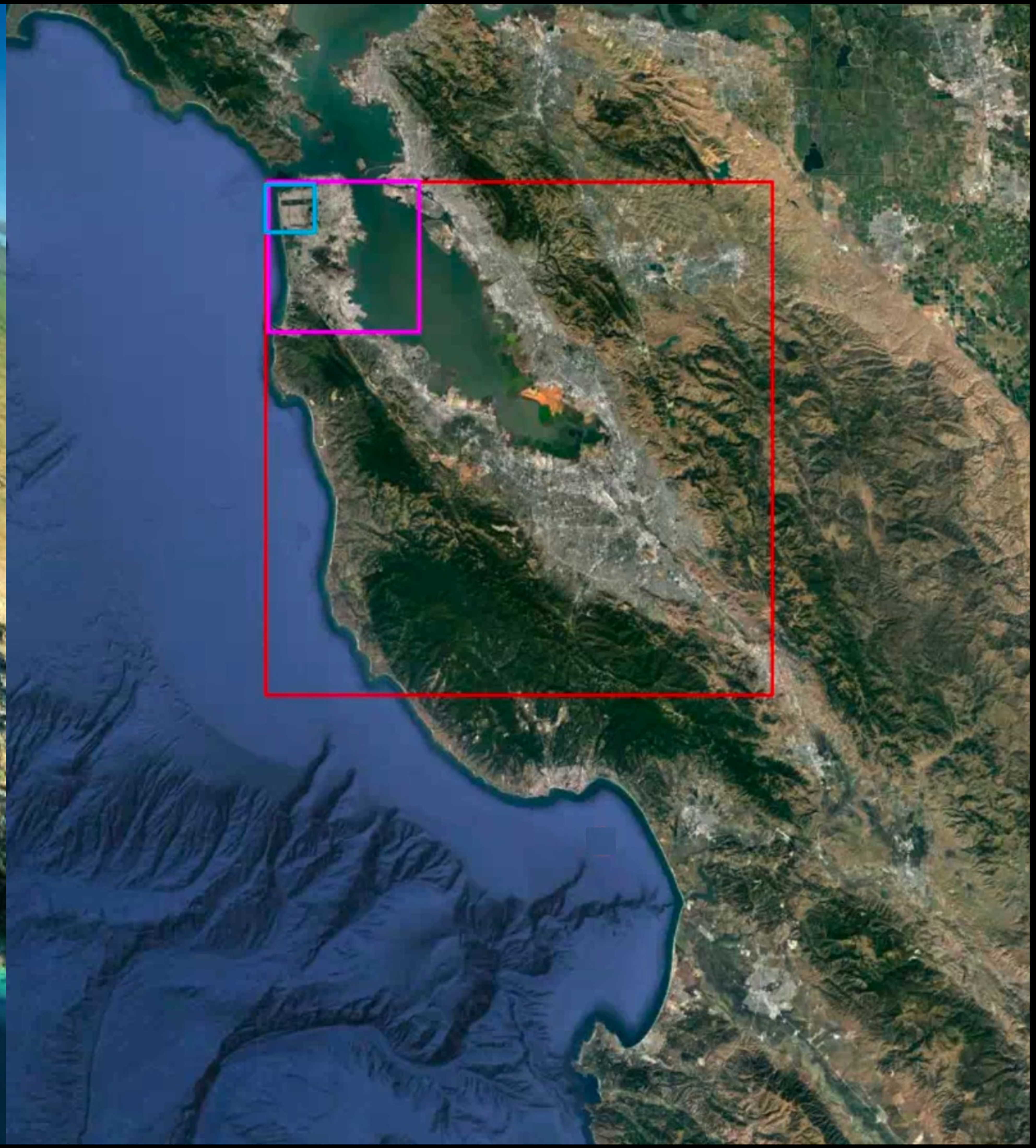
Deuben & Bauer (2018), 6° , 60x30, 1.8K pixels, MLP

# … Ask Questions about Climate Change's Impacts…

On Food, Health, Infrastructure, Energy systems, and more…



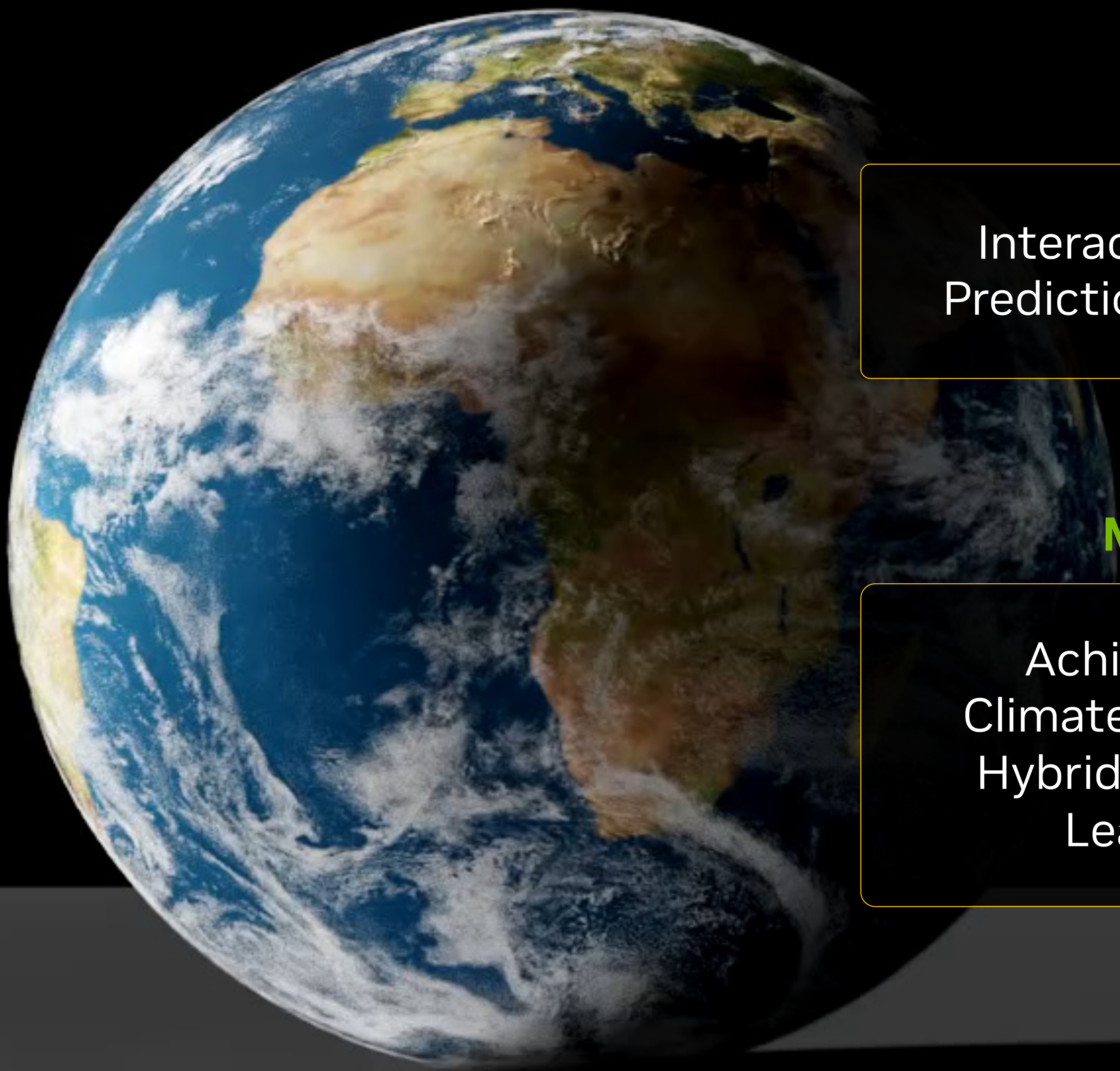

# Imagine you could Select a Region of the Planet...



# Imagine the System Evolving in Scientific Fidelity and Computational Ambition

Eventually fed by a new library of climate predictions so high-resolution they seem impossible today.

Earth



Twin Earth

**Mission #1**

Interacting with Climate Predictions at Low Latency.

**Mission #2**

Achieving Next-Gen Climate Predictions using Hybrid Physics, Machine Learning & HPC.

# EVE: Earth Visualization Engine

The Berlin Summit, 3~7th July.
https://eve4climate.org/

- Conference Keynotes
  - **Jensen Huang, Founder and CEO NVIDIA, USA**

  - Frans Timmermans, Executive Vice President, European Commission, EU
  - Deborah R. Coen, Professor, Chair of the History of Science & Medicine Program, Yale University, US
  - Fabiola Gianotti (tbc), Director-General of CERN, EU
  - Aromar Revi, Director Indian Institutes for Human Settlements, India
  - Petteri Taalas, Secretary-General of the World Meteorological Organization, INT
  - Hao Xu, Tencent 腾讯, Vice President of Sustainable Social Value, China

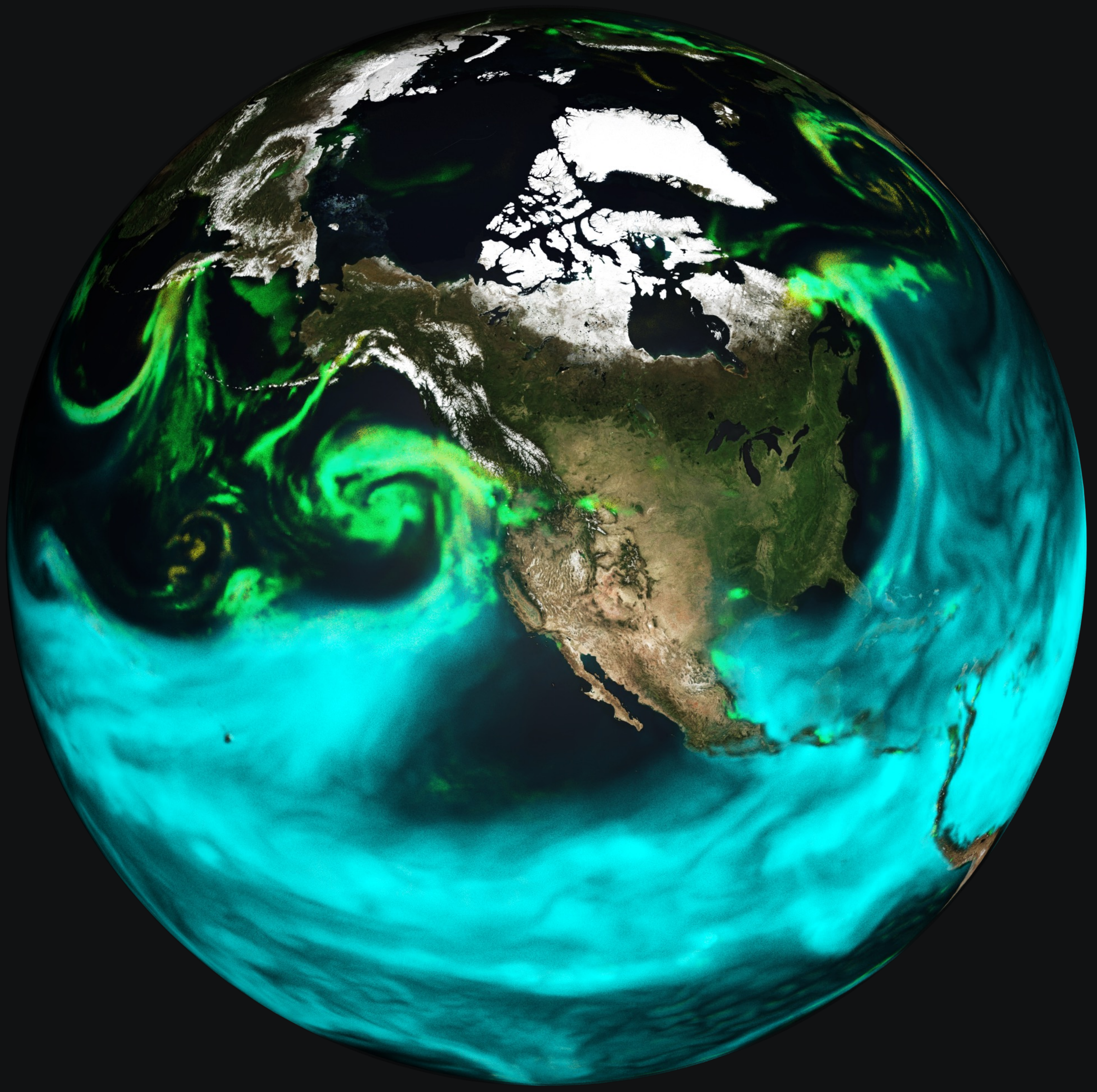
# Testing ML methods with potential to sidestep Moore's Law

Today's climate models are too low in resolution. Brute force numerical solvers are decades away from what is needed.


**Figure adapted from:** Schneider, T., Teixeira, J., Bretherton, C. *et al.* *"Climate goals and computing the future of clouds". Nature Climate Change* **7,** 3–5 (2017)

# CHALLENGES: NWP MODELS ARE EXTREME-SCALE SUPERCOMPUTING APPLICATIONS

- Complexity: solution of hundreds of coupled nonlinear partial differential equations (PDE)
- Resolution: achieving 1 km global resolution will not be possible before the year 2060
- Dimensionality: at 1 km resolution → $10^{12}$ degrees of freedom
- Ensemble size: Many sources of uncertainties → Large ensembles are required
- Throughput + Scalability & performance



# Weather Simulation with FourCastNet

## Fully data-driven weather prediction.

| | |
|---|---|
| • Scope | Global, Medium Range |
| • Model Type | Full-Model AI Surrogate |
| • Architecture | AFNO (Adaptive Fourier Neural Op.) |
| • Resolution: | 25km |
| • Training Data: | ERA5 Reanalysis |
| • Initial Condition | GFS / UFS |
| • Inference Time | 0.25 sec (2-week forecast) |
| • Speedup vs NWP | $O(10^4\text{-}10^5)$ |
| • Power Savings | $O(10^4)$ |

# EARTH DIGITAL TWIN


ERA5 ECMWF
Atmospheric Winds & Geopotential
10 TB | 25km | 5 Atmos Layers

100,000X Speed-Up
< 1 Second for 7-Day Forecast
Training: 16 Hours on 64 A100 GPUs

Digital Twin
Interactivity & Verification

RAPIDS

Modulus

Omniverse

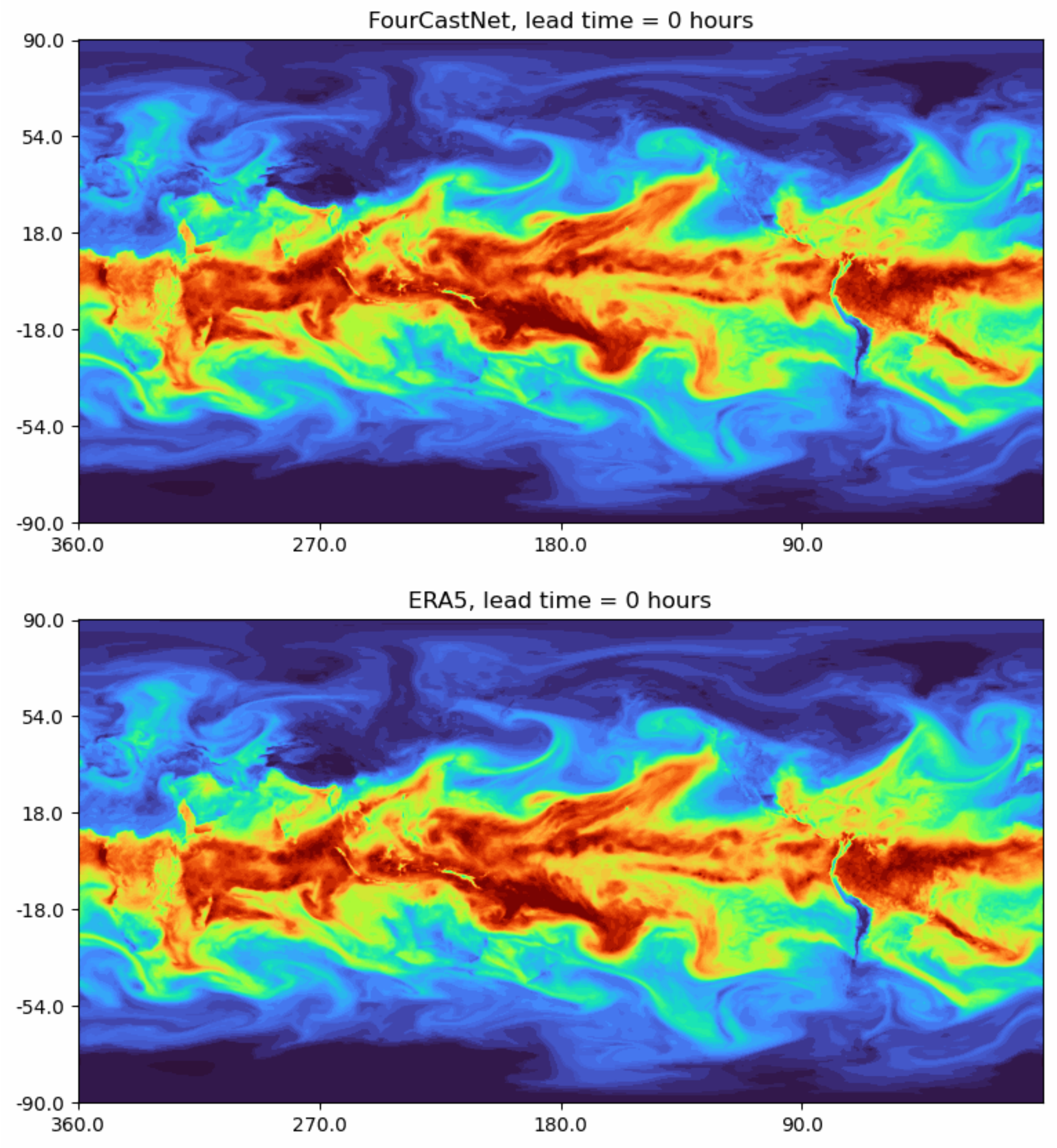
**DATA PROCESSING**

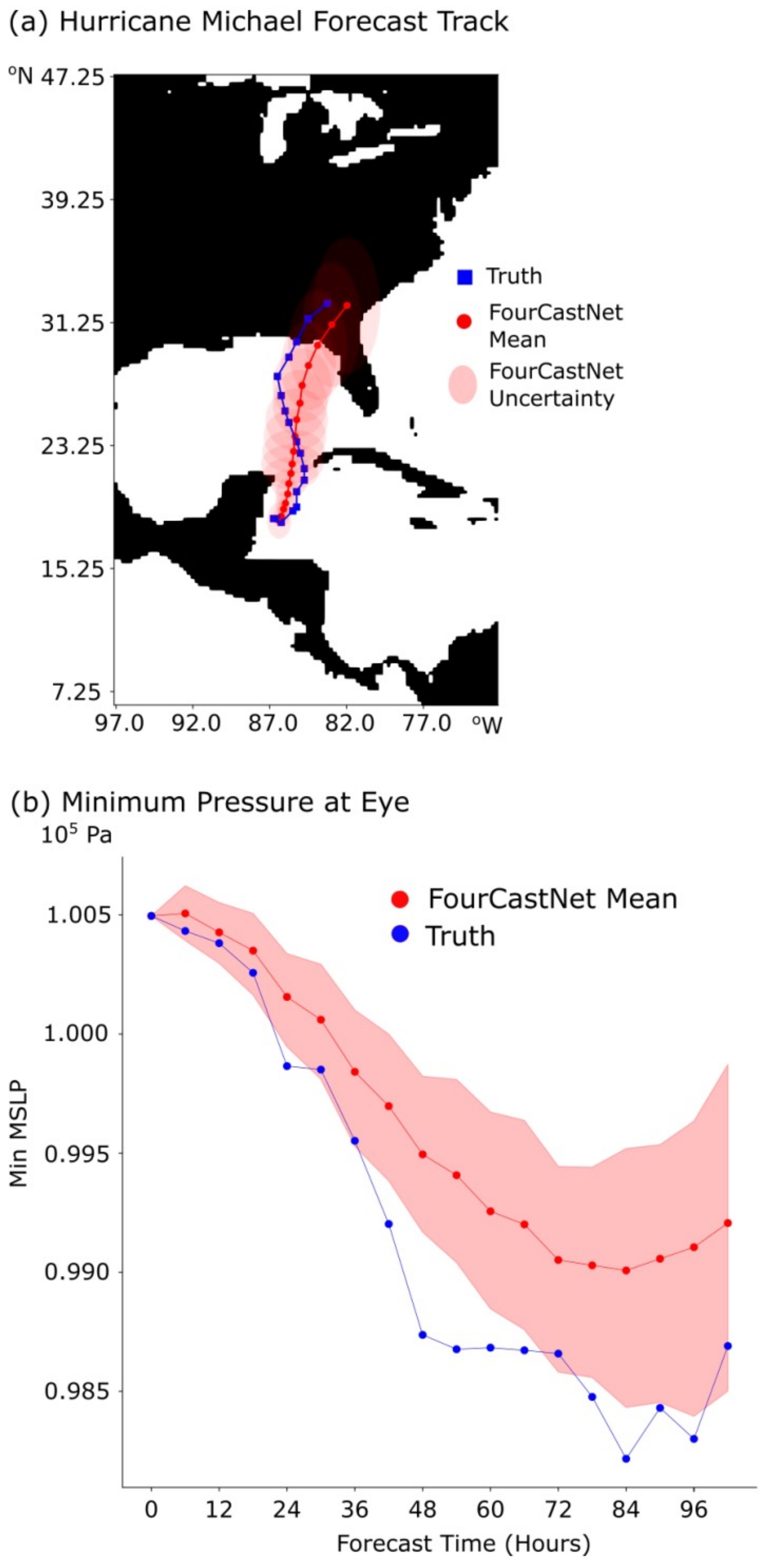
**AI-SIMULATION**

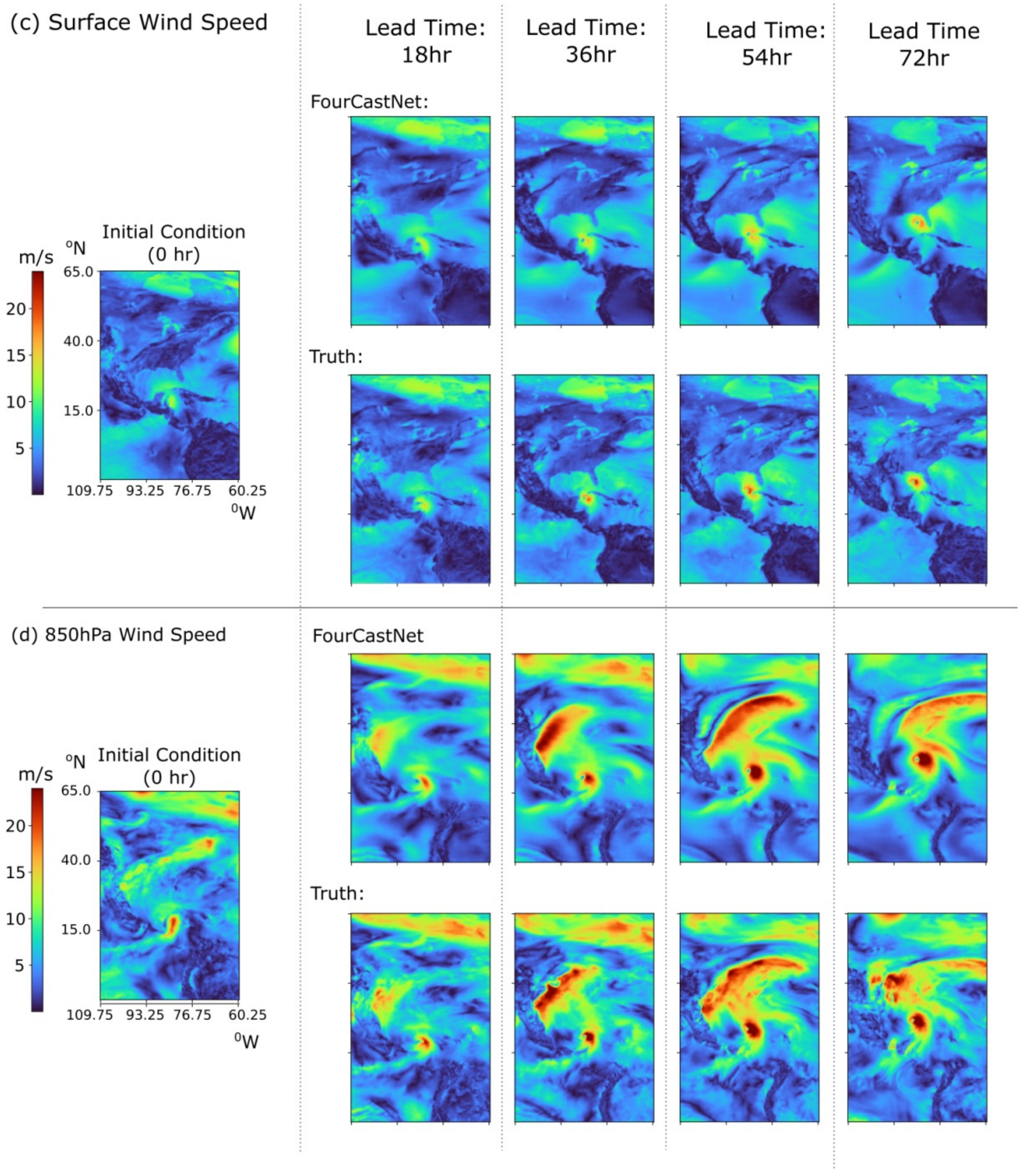
**VISUALIZATION**

# FourCastNet

[Modulus FourCastNet Document]



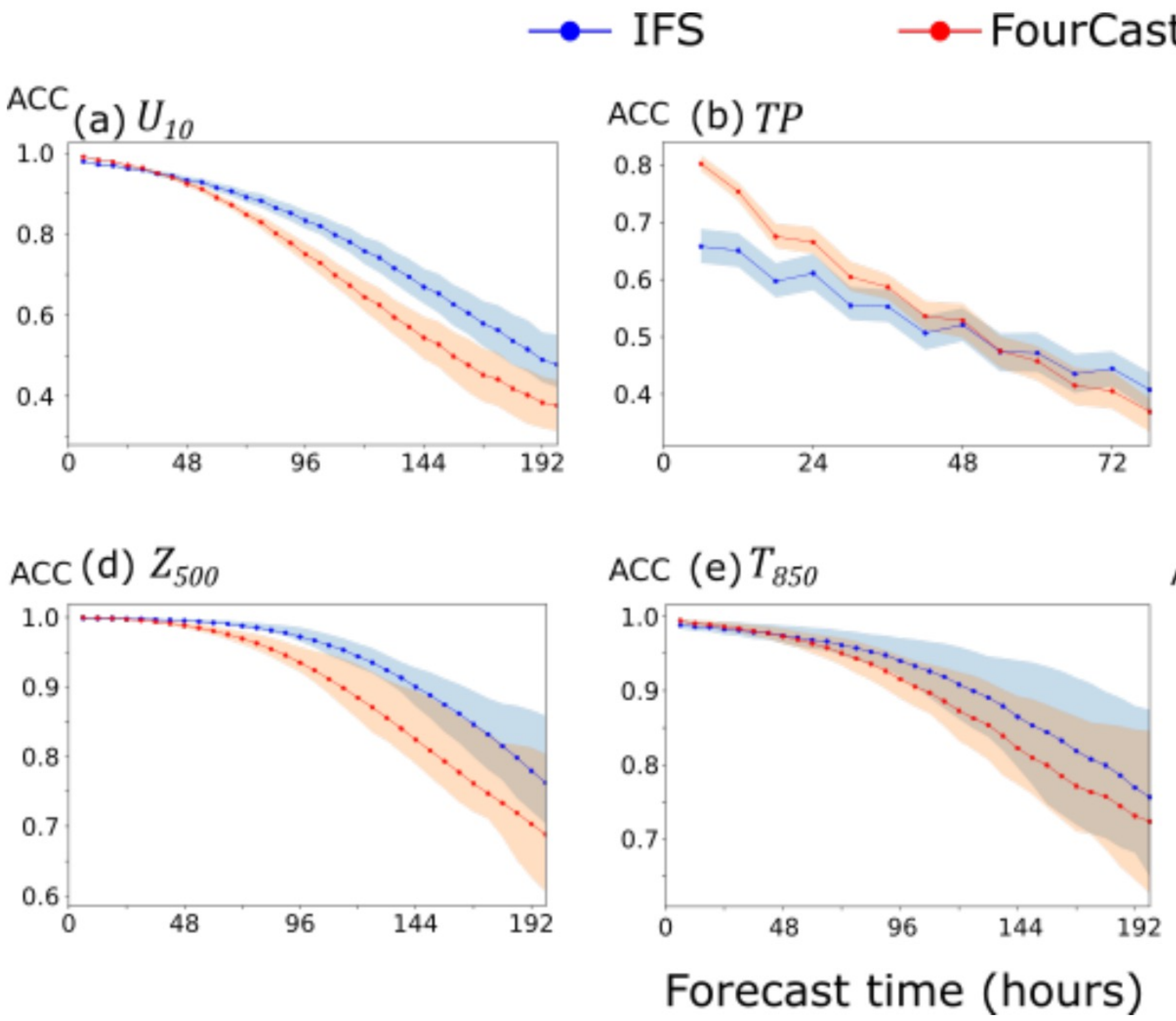



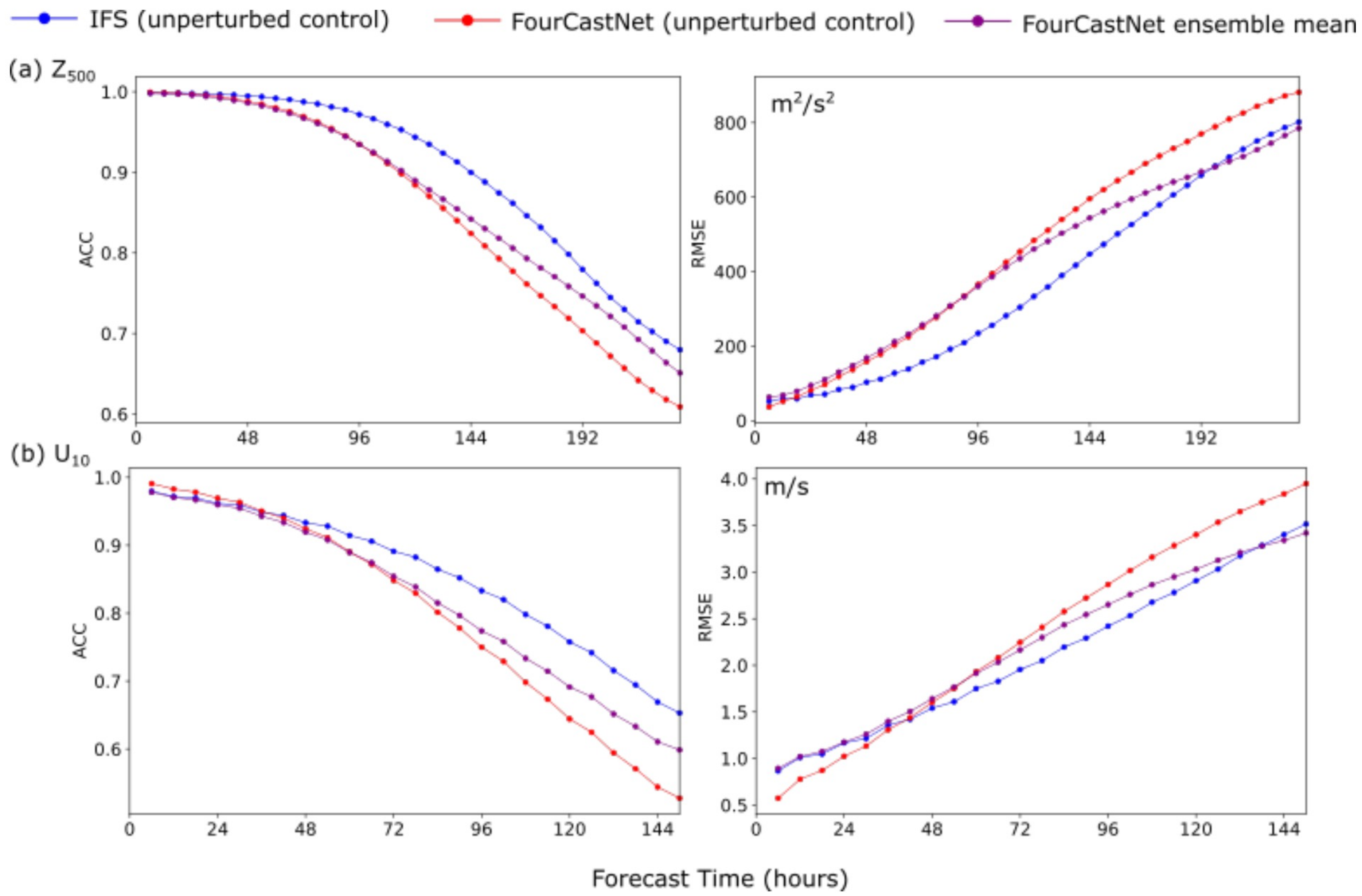



# FourCastNet

https://arxiv.org/abs/2202.11214 , Feb. 2022.

NVIDIA, Lawrence Berkley National Lab, University of Michigan, Rice University, Caltech, Purdue University



# FOURCASTNET

https://arxiv.org/abs/2202.11214 , Feb. 2022.

NVIDIA, Lawrence Berkley National Lab, University of Michigan, Rice University, Caltech, Purdue University

| Accurate | Fast | Efficient |
|---|---|---|
| **< 48 HRs** | **45,000x** | **12,000x** |
| Better than ECMWF-IFS for precipitation prediction | Less than 1 sec to generates a week-long forecast. | Less energy to generate a forecast than the IFS model |

Node-to-Node comparison
for a **24-hours 100-member 18km** ensemble forecast

CPU node (24 cores) vs GPU node (4 x A100s)

# A NOTE: Math behind Neural Operator

Anima Anandkumar - Neural operator: A new paradigm for learning PDEs
Neural Operator: Learning Maps Between Function Spaces
Nickola K., Zongyi L. et. al., Caltech.
https://zongyi-li.github.io/neural-operator/
Green Function

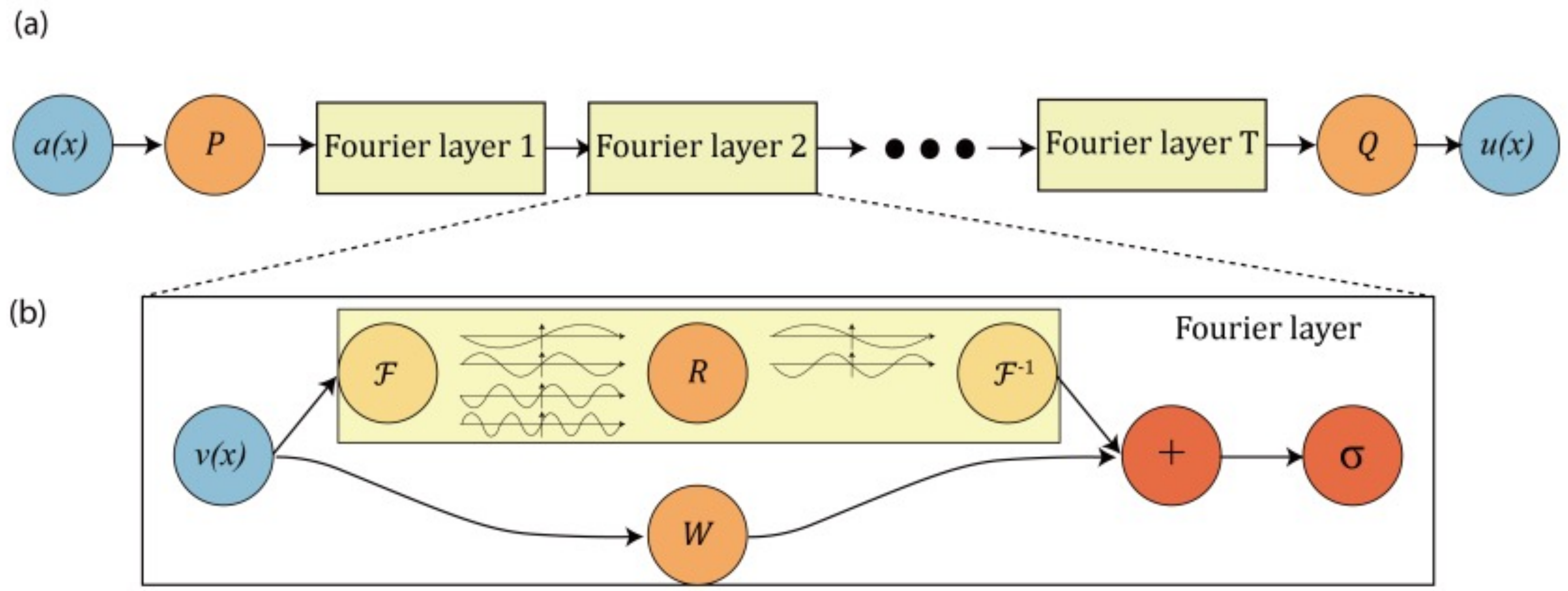
$$\begin{aligned} Goal: \quad & Lu = f \\ Def: \quad & L^{\dagger}G = \delta(x-\xi) \\ \langle Lu, G\rangle &= \langle f, G\rangle \\ \langle u, L^{\dagger}G\rangle &= \langle f, G\rangle \\ \langle u, \delta(x-\xi)\rangle &= \langle f, G\rangle \\ u(\xi) &= \langle f, G\rangle \end{aligned}$$

- Note
  - $\langle , \rangle$: function space inner product, project one function to another, which is the integral of multiplying two functions.
  - $L$: a linear operator.
  - $L^{\dagger}$: a "adjoint" operator of $L$, if $L = L^{\dagger}$ then they should has the same boundary conditions, named "self-adjoint operator".
  - $\delta$: a `dirac function` use to `silter` function `u` at value $\xi$.
  - $u$: target solution. and $\xi$ is just a dummy variable, actually we could simply change it here $u(\xi) \to u(x)$, we get the final solution with proper form.

# Fourier Neural Operator (FNO)

https://arxiv.org/abs/2010.08895, ICLR 2021.
FNO: FFT + NO, Zongyi L. et. al., Caltech+NVIDIA

- Contribution
  - Learned an entire family of PDEs instead of solving only one instance such as FEM, FDM, PINN.
  - Resolution-invariant, can do Zero-Shot super-resolution.
  - Outperformed all existing DL methods with 30%+ lower error rate.
  - Sped up more than 440x comparing to Spectral Method.




- Method

**Definition 1 (Iterative updates)** *Define the update to the representation $v_t \mapsto v_{t+1}$ by*

$$v_{t+1}(x) := \sigma\Big(W v_t(x) + \big(\mathcal{K}(a;\phi)v_t\big)(x)\Big), \qquad \forall x \in D \tag{2}$$

**Definition 2 (Kernel integral operator $\mathcal{K}$)** *Define the kernel integral operator mapping in (2) by*

$$\big(\mathcal{K}(a;\phi)v_t\big)(x) := \int_D \kappa\big(x, y, a(x), a(y); \phi\big)v_t(y)\mathrm{d}y, \qquad \forall x \in D \tag{3}$$

**Definition 3 (Fourier integral operator $\mathcal{K}$)** *Define the Fourier integral operator*

$$\big(\mathcal{K}(\phi)v_t\big)(x) = \mathcal{F}^{-1}\Big(R_\phi \cdot (\mathcal{F}v_t)\Big)(x) \qquad \forall x \in D \tag{4}$$

# Exciting times: More industrial AI weather models rapidly increasing in skill.

We are living in the time when AI modeling outperforms deterministic numerical weather prediction.

## Pangu-Weather: A 3D High-Resolution Model for Fast and Accurate Global Weather Forecast

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, Qi Tian

In this paper, we present Pangu-Weather, a deep learning based system for fast and accurate global weather forecast. For this purpose, we establish a data-driven environment by downloading 43 years of hourly global weather data from the 5th generation of ECMWF reanalysis (ERA5) data and train a few deep neural networks with about 256 million parameters in total. The spatial resolution of forecast is $0.25^\circ \times 0.25^\circ$, comparable to the ECMWF Integrated Forecast Systems (IFS). More importantly, for the first time, an AI-based method outperforms state-of-the-art numerical weather prediction (NWP) methods in terms of accuracy (latitude-weighted RMSE and ACC) of all factors (e.g., geopotential, specific humidity, wind speed, temperature, etc.) and in all time ranges (from one hour to one week). There are two key strategies to improve the prediction accuracy: (i) designing a 3D Earth Specific Transformer (3DEST) architecture that formulates the height (pressure level) information into cubic data, and (ii) applying a hierarchical temporal aggregation algorithm to alleviate cumulative forecast

Huawei, Nov. 2022
https://arxiv.org/abs/2211.02556

## GraphCast: Learning skillful medium-range global weather forecasting

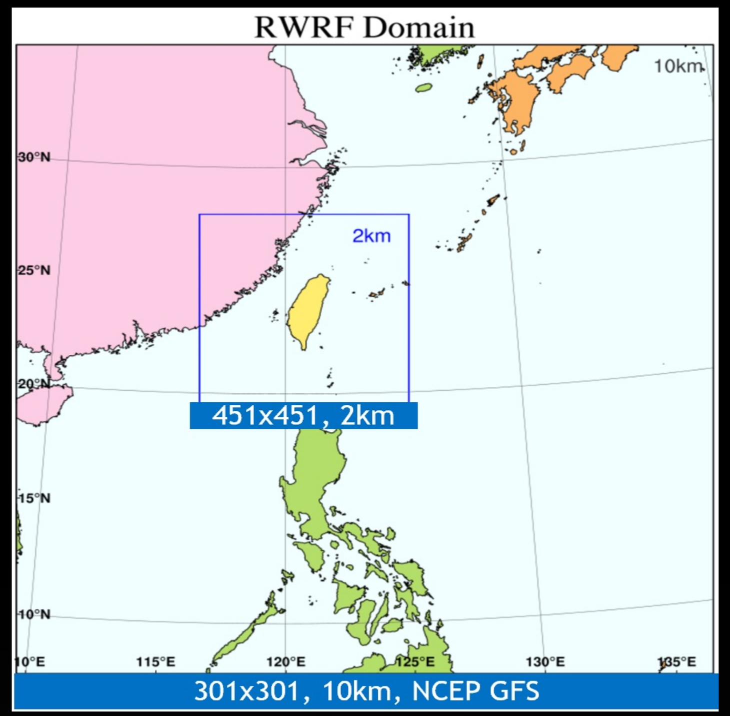
Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Alexander Pritzel, Suman Ravuri, Timo Ewalds, Ferran Alet, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Jacklynn Stott, Oriol Vinyals, Shakir Mohamed, Peter Battaglia

We introduce a machine-learning (ML)-based weather simulator--called "GraphCast"--which outperforms the most accurate deterministic operational medium-range weather forecasting system in the world, as well as all previous ML baselines. GraphCast is an autoregressive model, based on graph neural networks and a novel high-resolution multi-scale mesh representation, which we trained on historical weather data from the European Centre for Medium-Range Weather Forecasts (ECMWF)'s ERA5 reanalysis archive. It can make 10-day forecasts, at 6-hour time intervals, of five surface variables and six atmospheric variables, each at 37 vertical pressure levels, on a 0.25-degree latitude-longitude grid, which corresponds to roughly 25 x 25 kilometer resolution at the equator.

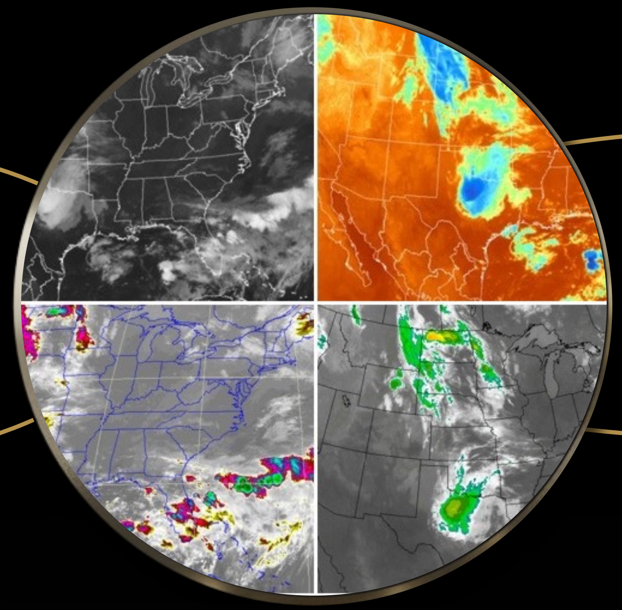
DeepMind, Dec. 2022
https://arxiv.org/abs/2212.12794

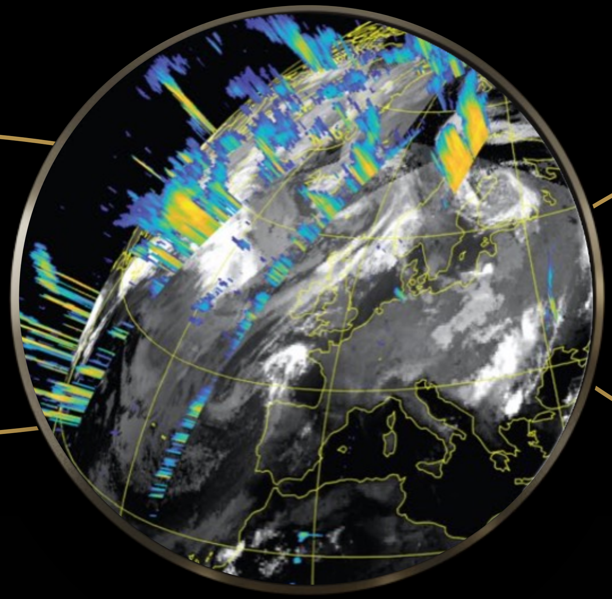
# EARTH-2/ TAIWAN-2 PROOF-OF-CONCEPT @ **CWB**

交通部中央氣象局
Central Weather Bureau



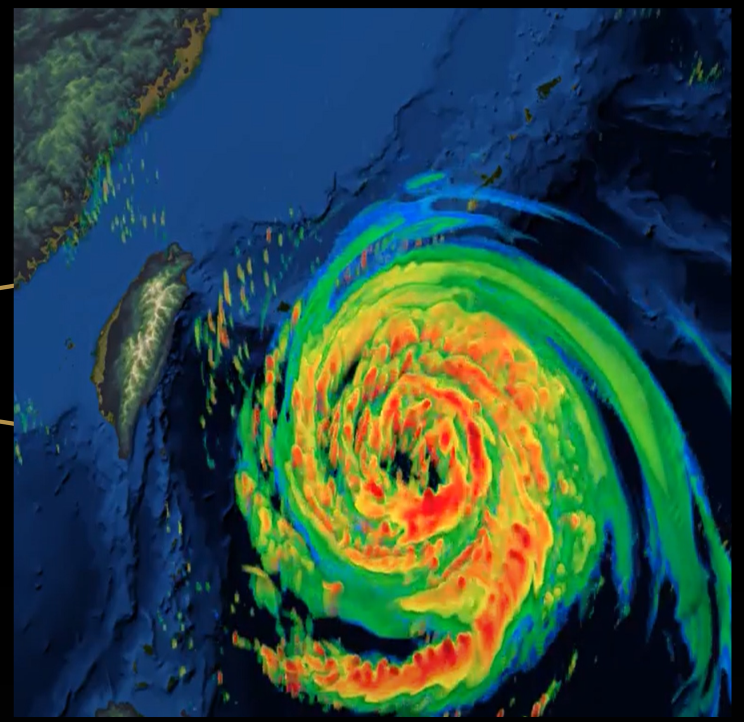

CWB RWRF (2km)





Fourier Neural Operator

AI Taiwan EMULATOR



Accuracy

Efficiency

Speed

RAPIDS

Modulus

Omniverse

**DATA PROCESSING**

**AI-SIMULATION**

**VISUALIZATION**

# CWB-RWRF-FCN

## CWB RWRF Data Status

| CWB-FCN | | |
|---|---|---|
| Past 6 hours | → | Next hour prediction. |

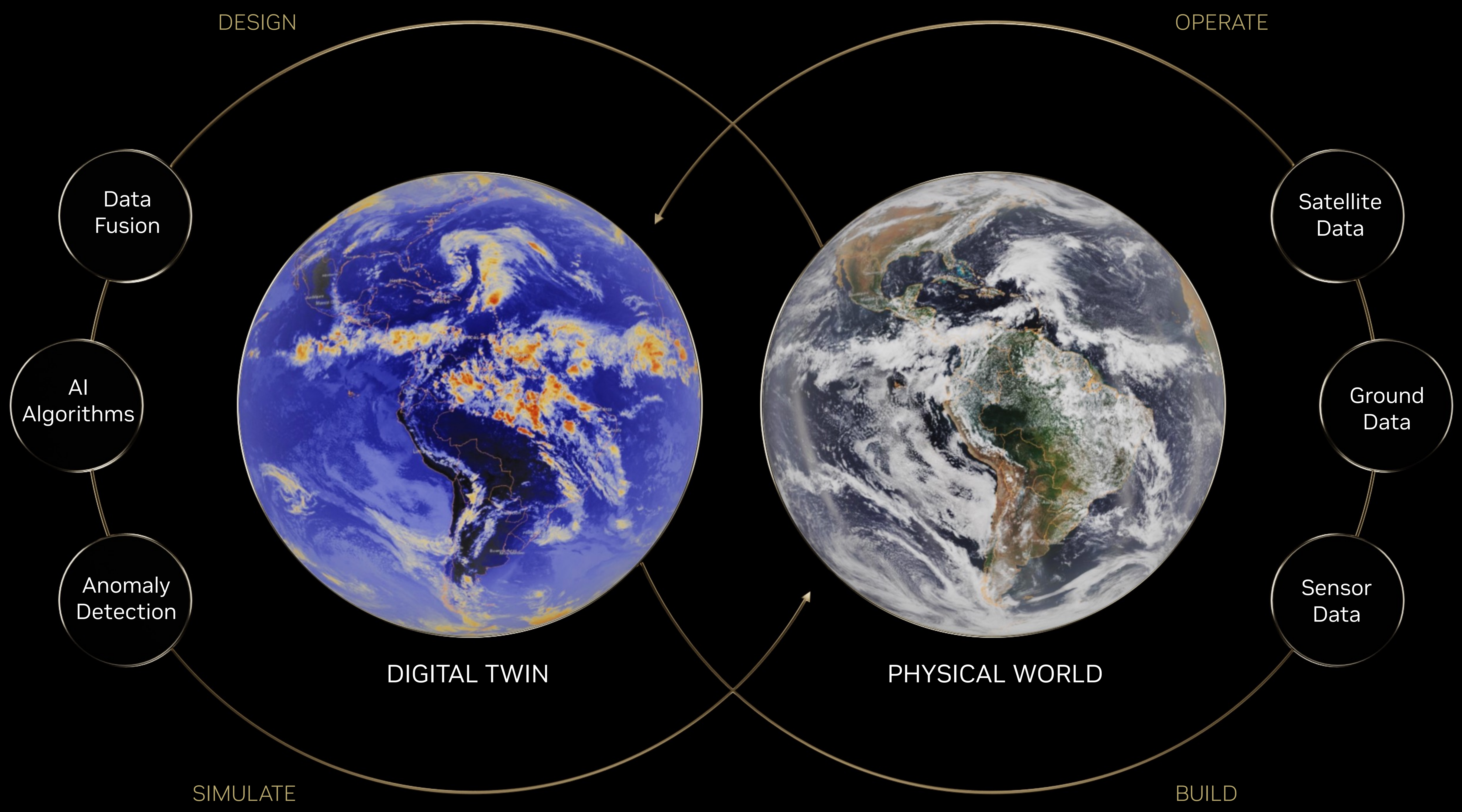
| CWB RWRF | |
|---|---|
| **Vertical Levels** | **Variables** |
| Observations | Radar |
| Surface | T2M, U10, V10 |
| 500 hpa | T, U, V, Z |
| 700 hpa | T, U, V, Z |
| 850 hpa | T, U, V, Z |
| 925 hpa | T, U, V, Z |

| | ECMWF ERA5 | CWB RWRF |
|---|---|---|
| **Data** | ERA5 (30km) | RWRF (2km) |
| **Data Shape** | 1440x720x20 | 450x450x20 |
| **Data Size** | 36 Years | 5 Years |
| **Model Type** | Global | Regional |
| **Frequency** | 6-hour | 1-hour |

# FOURCASTNET ON CWB RWRF

| Accurate | Fast | Efficient |
| --- | --- | --- |
| **< 5 HRs** | **3,600x** | **47,000x** |
| Show strong potential to compensate CWB RWRF short-term forecasting. | Less than 1 sec to generates a 13 hour-long forecast. | Less energy to generate a forecast than the RWRF. |
| 2020/08/26 11:00 Southwesterly Event, FCN outperforms RWRF for 1~5 hours radar reflectivity forecasting. (Reference: CWB Researcher) | **13-hours 32-member ensemble** forecast 912 CPU cores vs 1 GPU V100, Yet Optimized (Reference: 2021 CWB RFP Operation Settings) | 200W per 32 CPU cores vs 3500W per 8 GPU V100 (Reference: Wiki SPARC64 Xifx node) |

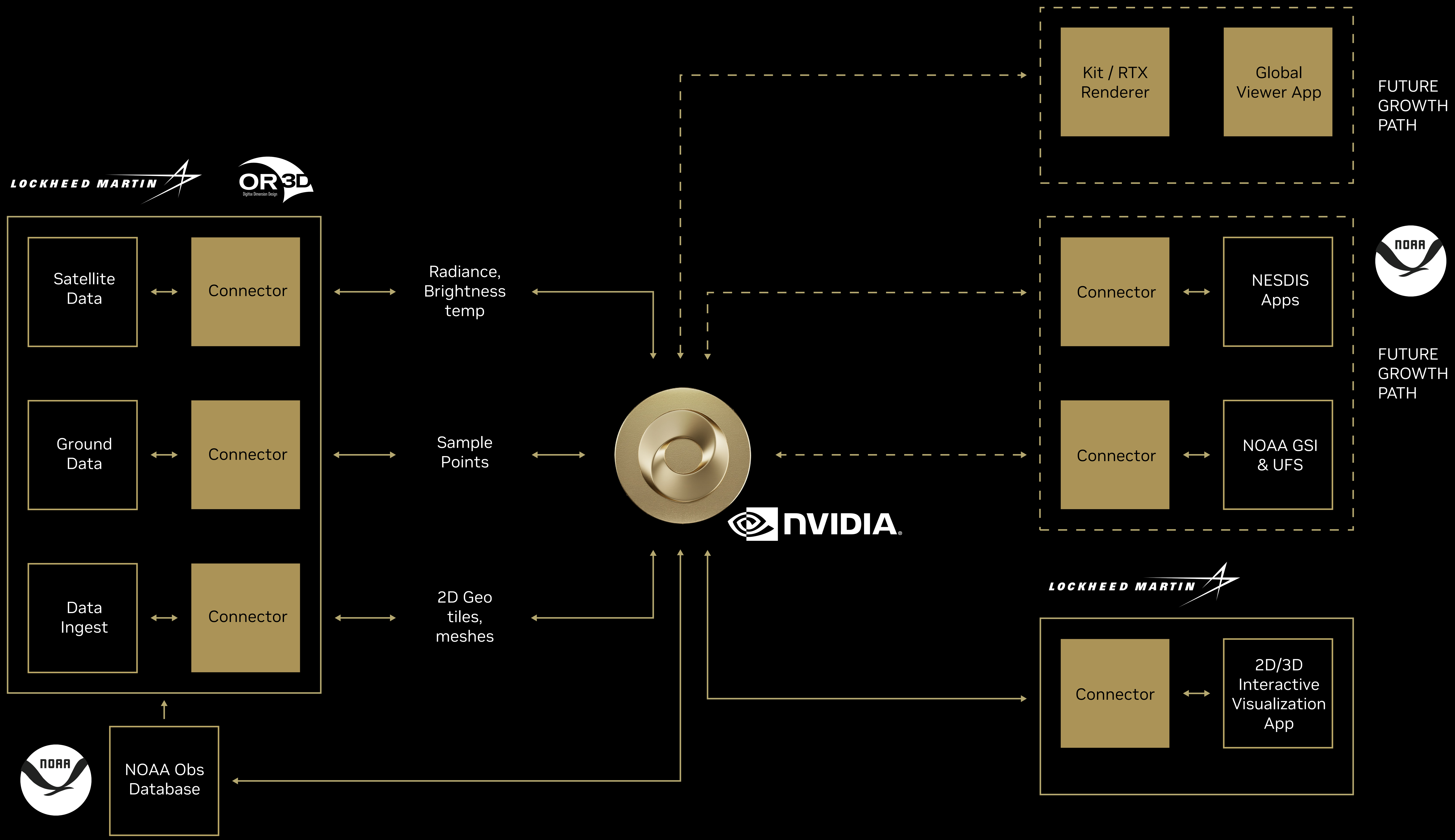
# Earth-2 Framework



NVIDIA, LOCKHEED MARTIN TEAM UP WITH STATE AND FEDERAL FOREST SERVICES TO FIGHT WILDFIRES WITH AI

# Earth-2 Framework

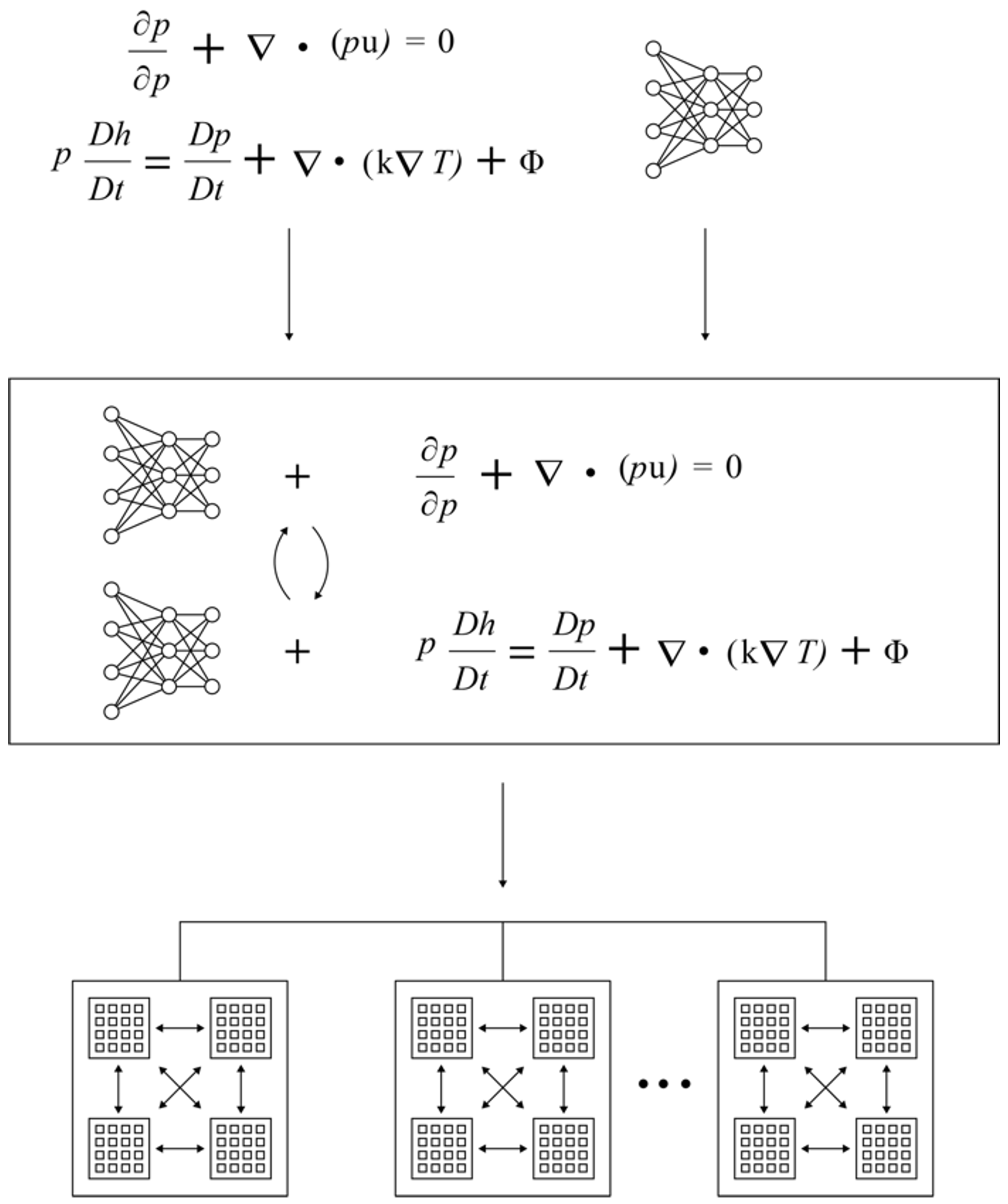
## Lockheed Martin & NVIDIA Selected By NOAA For Climate Science Digital Twin



# NVIDIA Modulus

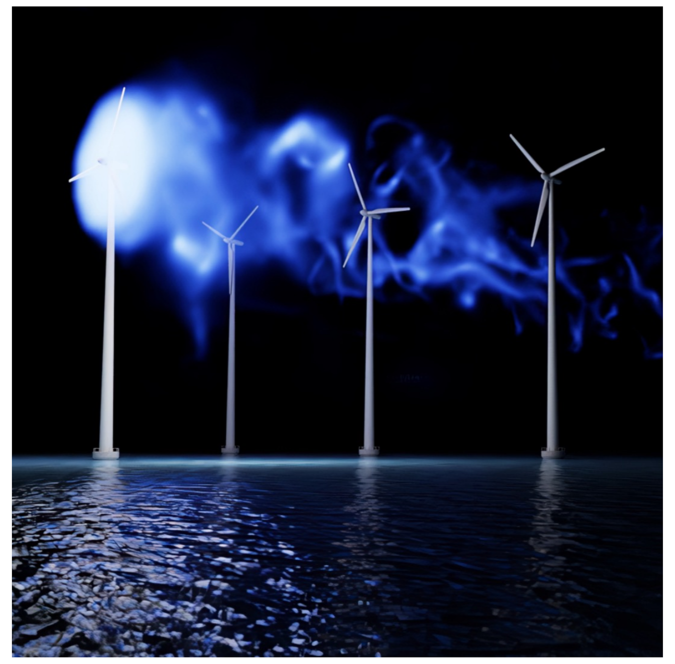
## Physics Machine Learning Platform

### Training Neural Networks with Data And Governing Equations

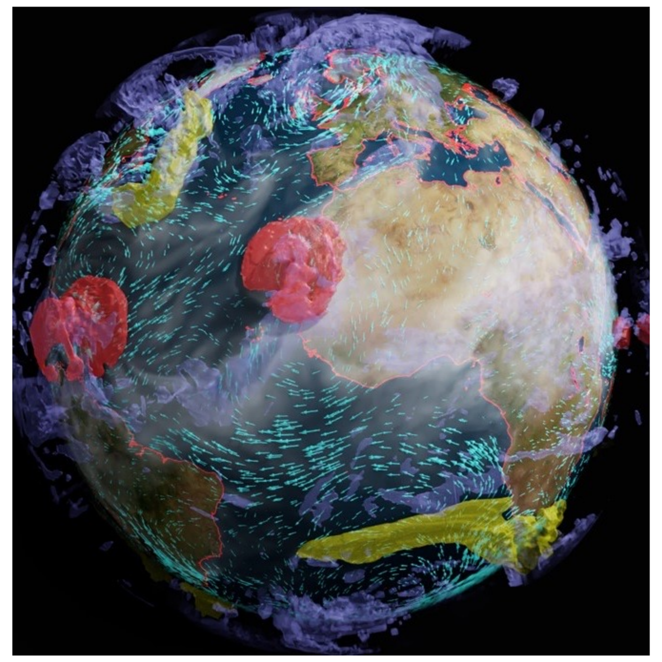



### Advancing Scientific Discovery

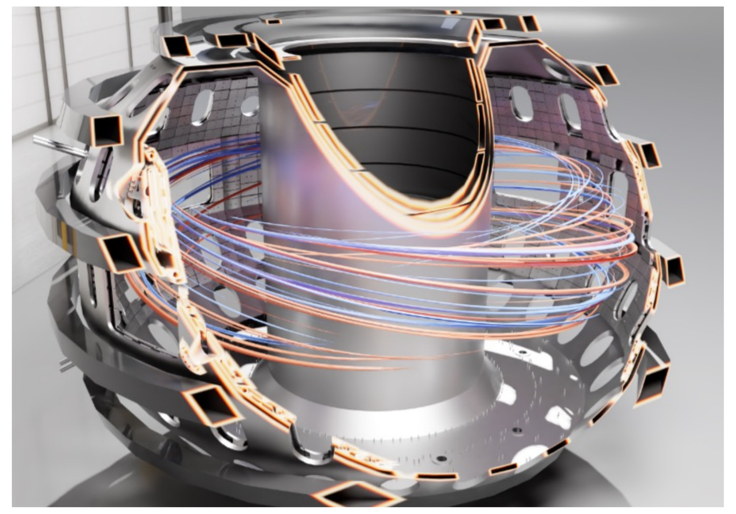
RENEWABLE ENERGY
Siemens Gamesa
Up to 4000X Speedup of Wind Turbine Wake Optimization



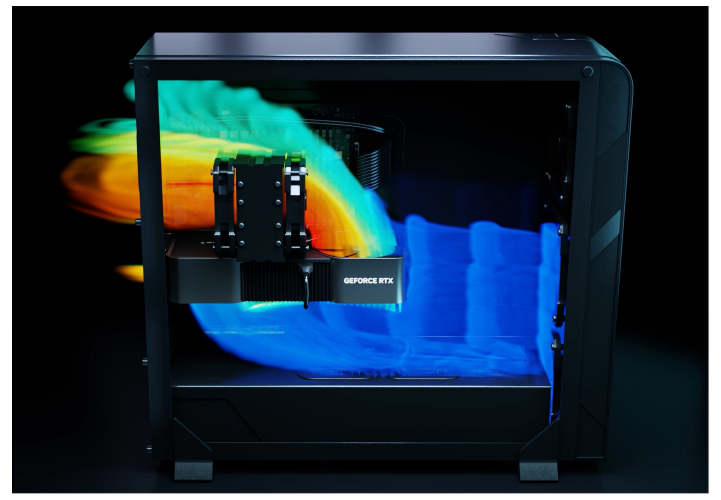
CLIMATE CHANGE
45,000X Speedup of Extreme weather Prediction with FourCastNet



RENEWABLE ENERGY
SGTC Plasma Fusion



INDUSTRIAL HPC
RTX Workstation Digital Twin



NVIDIA