



Extreme Weather with Earth-2

Jay Chen | Solution Architect | May 2024.

Extreme weather events have become more frequent and more severe

Mediterranean Heatwaves 2023



<https://abcnews.go.com/International/wildfires-rage-amid-heat-wave-greece-promoting-evacuations/story?id=101511245>

Hurricane Ian 2022



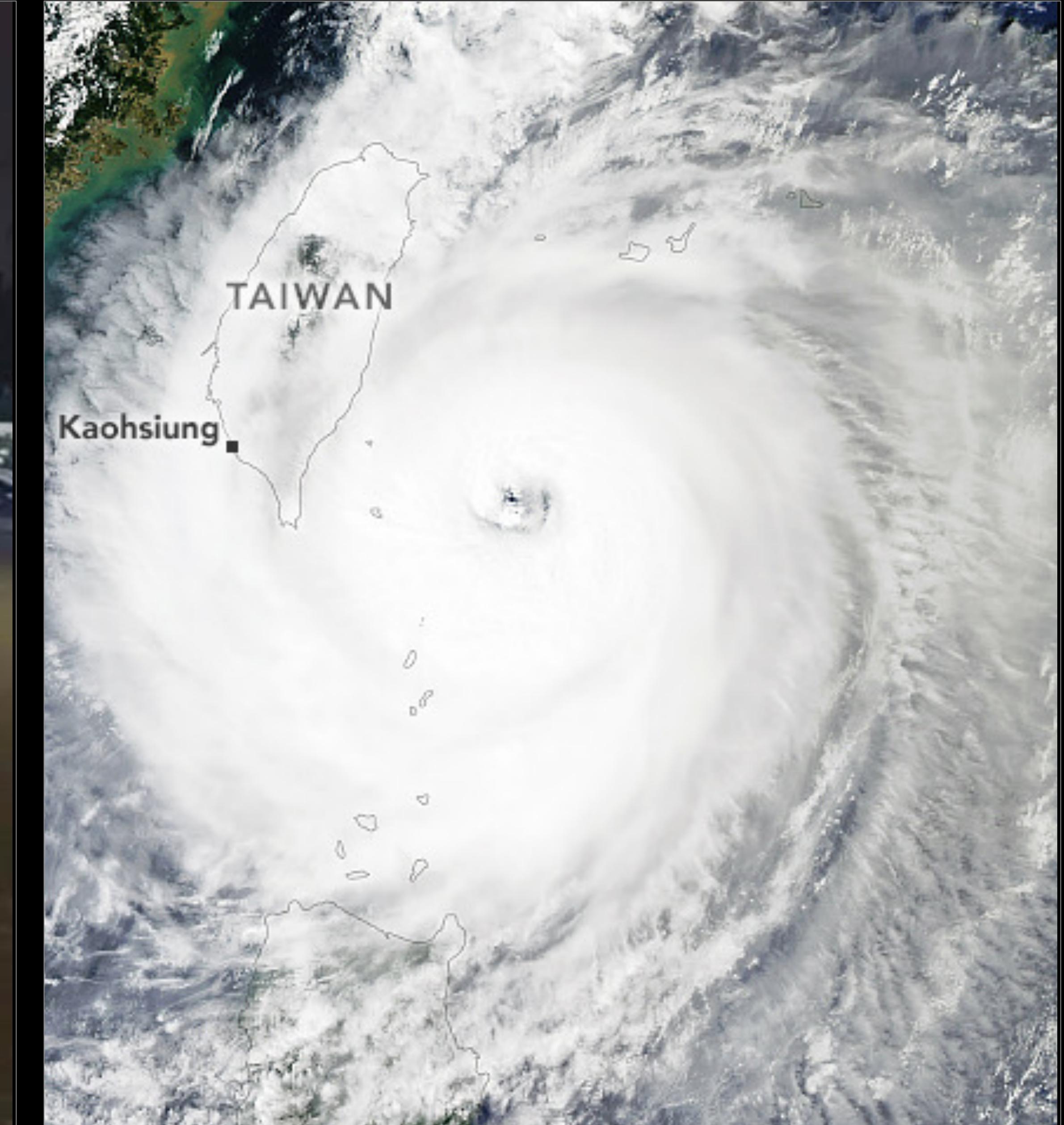
<https://commons.wikimedia.org/w/index.php?curid=123606695>

Pakistan Floods 2022



https://commons.wikimedia.org/wiki/File:Flood_in_Pakistan_2022.png

Disasters Cost NT\$20B Yearly



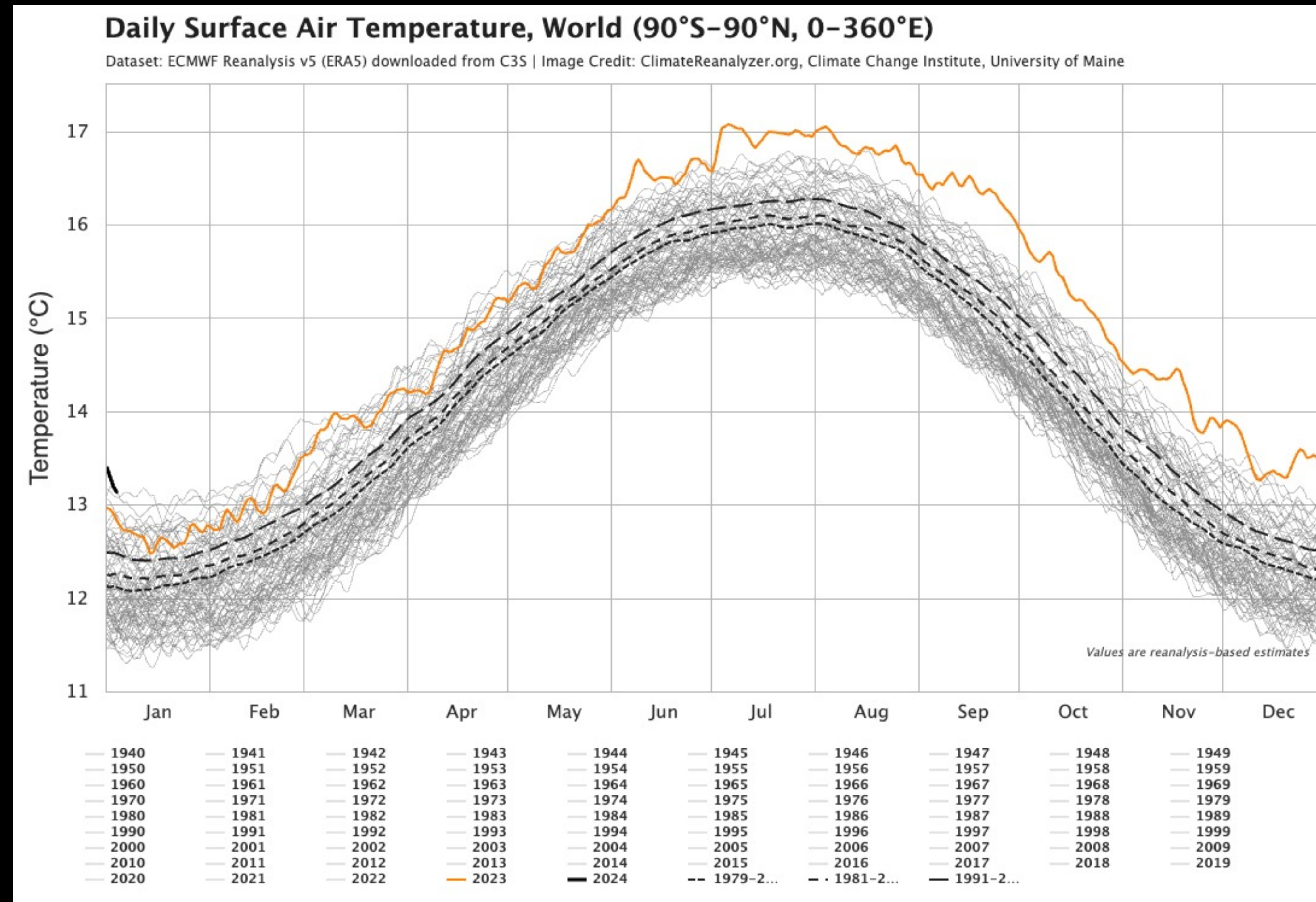
Taiwan Drought
2020-2021

Extremely Heavy Rain
2023/09

Typhoon Koinu
2023/10

Our future climate will be very different from the past

Global Surface Temperatures



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



... Answer Questions about Climate Change's Impacts

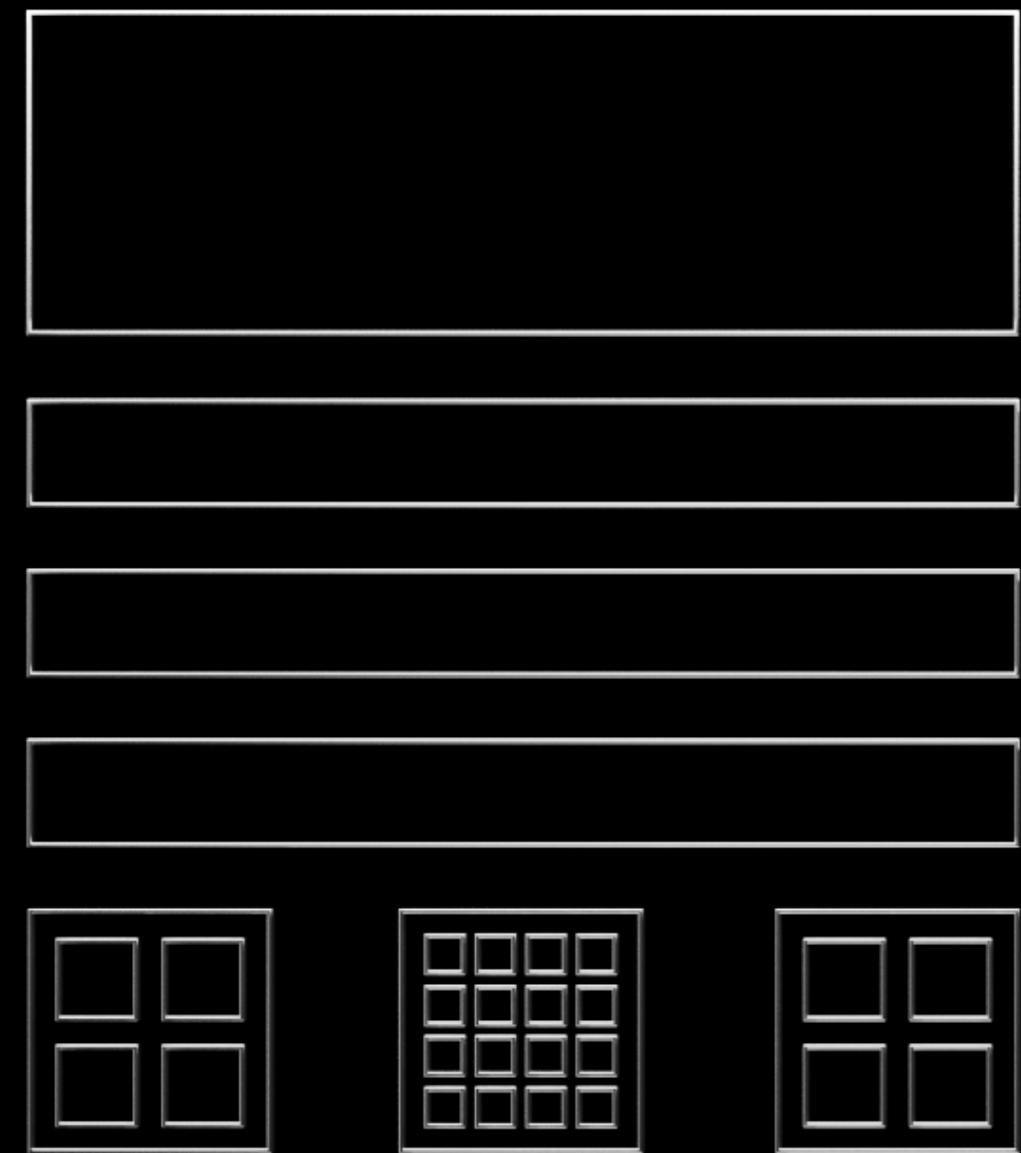
On Food, Health, Infrastructure, Energy systems, and more...



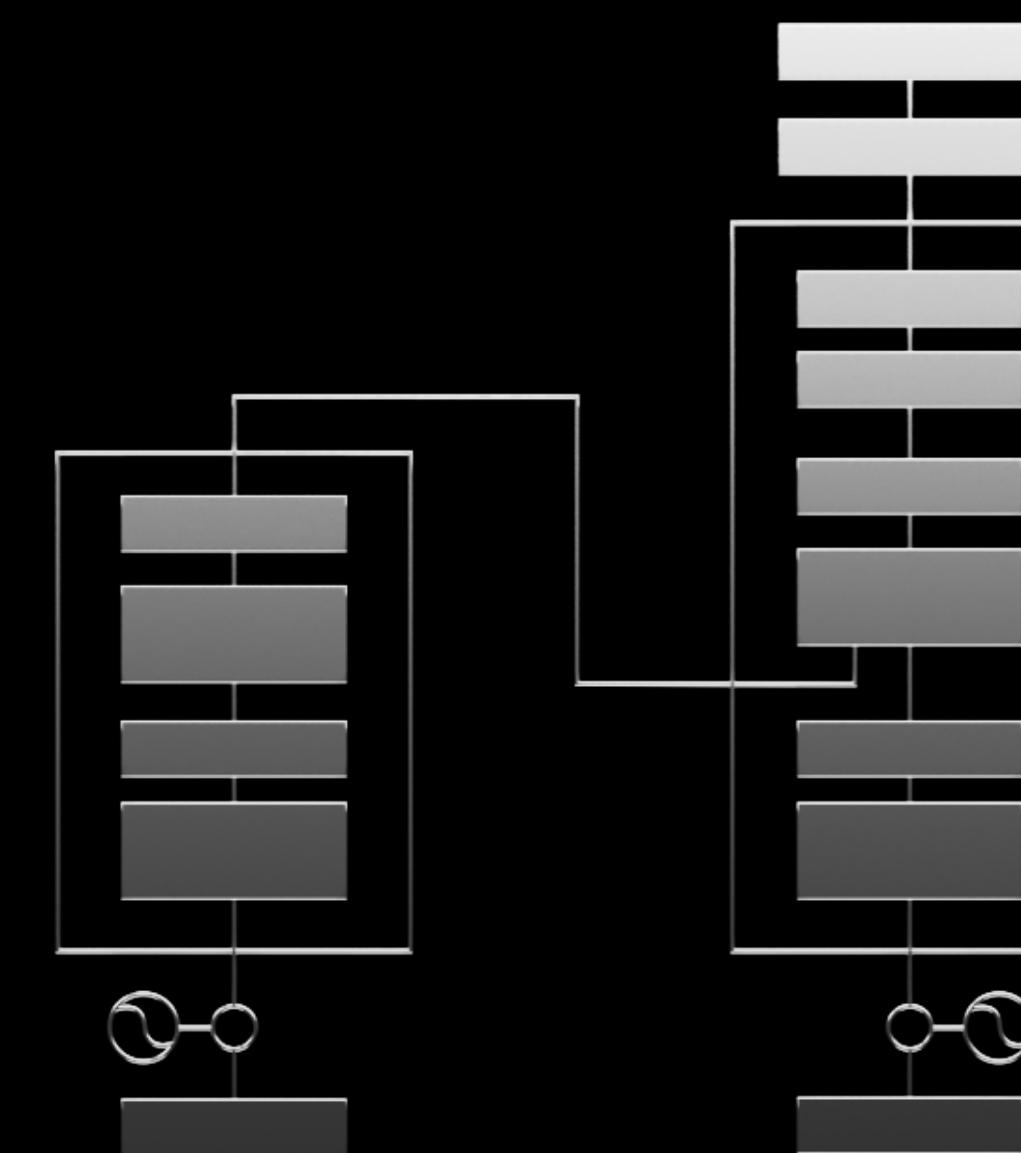
Earth-2: An NVIDIA initiative to boost climate science & climate tech

Using accelerated simulation, visualization, artificial intelligence & informatics.

SIMULATION



AI

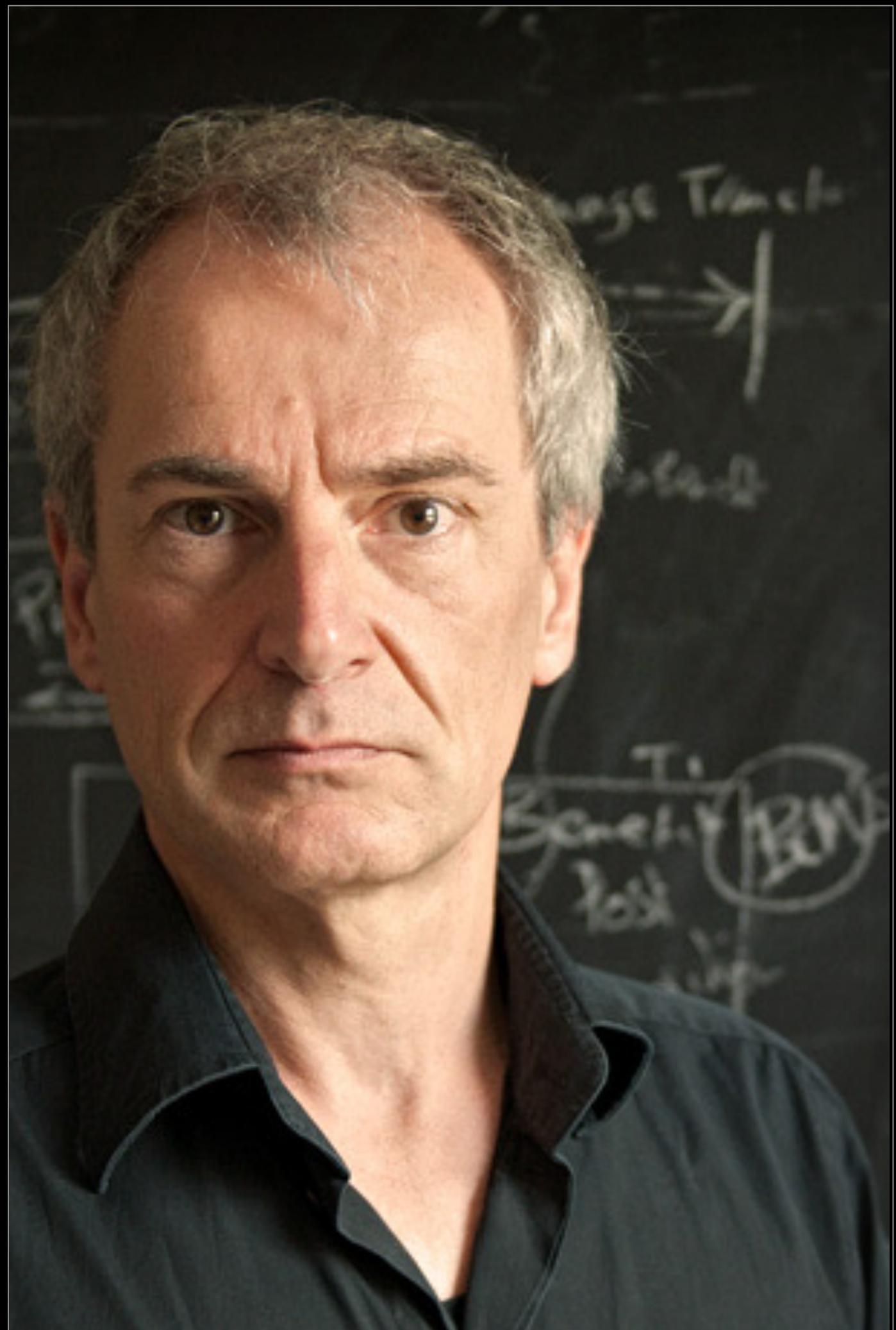


VISUALIZATION



Earth-2 is advised by world-leading climate scientists

Thought-leaders & champions of interactive AI digital twins of our planet.



Peter Bauer
ECMWF



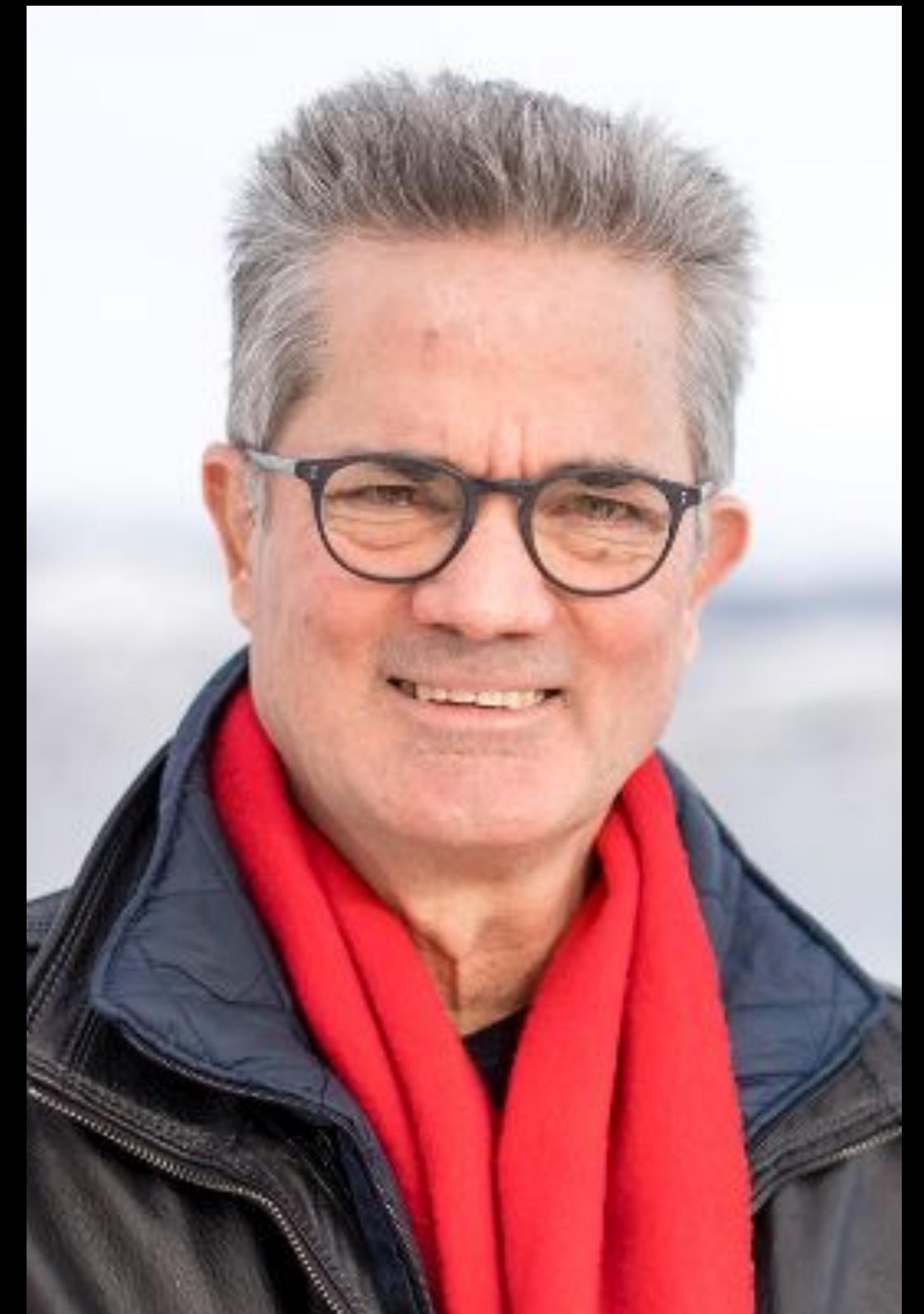
Peter Dueben
ECMWF



Bjorn Stevens
MPI-Hamburg



Francisco Doblas-Reyes
Barcelona Supercomputing Center



Thomas Schulthess
ETHZ and CSCS

Digital twins

Highly Interactive Weather and Climate Information Systems, High Resolution, that Serve Society...

Earth-2 Mission #1

Enable Next-Gen Weather &
Climate Predictions

Earth-2 Mission #2

Interacting with Predictions at
Low Latency

Earth-2's mandate: Achieve "3 miracles" -- Jensen Huang,

Remarks to Berlin Summit for Climate Simulation, Berlin, July 2023.

MIRACLE #1

Km-scale Simulations
30,000 SYPD
30 MW

TO

Generate hi-fidelity data
energy efficiently

MIRACLE #2

AI Emulation of the Full State Vector
Any Region
Any Time Period

TO

To enable low-latency interactivity
with exabytes of data

MIRACLE #3

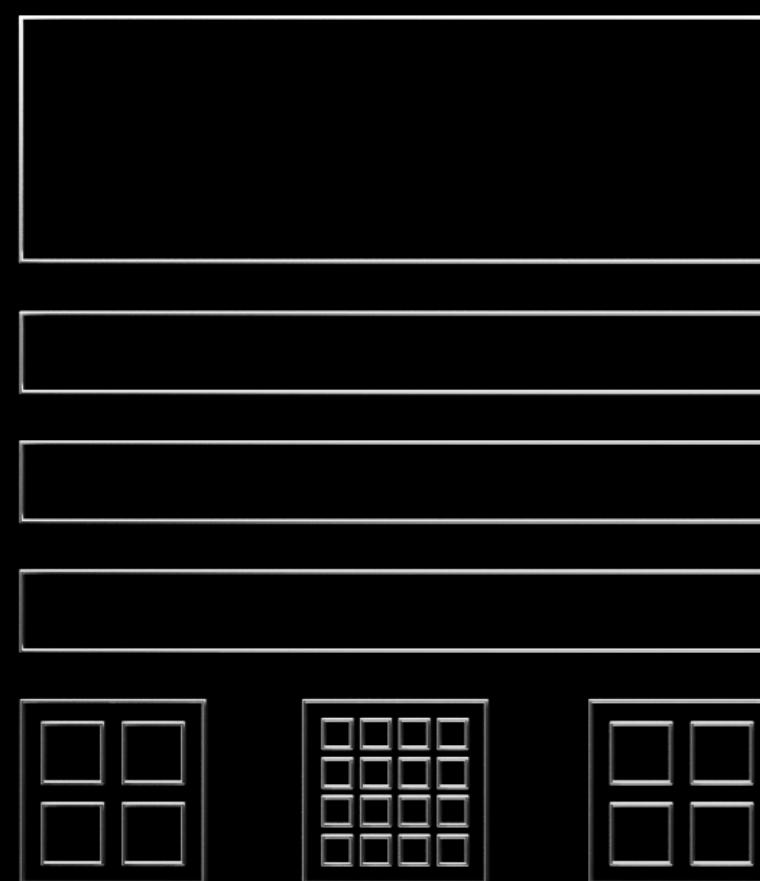
Full State Vector Visualization
From Cloud

TO

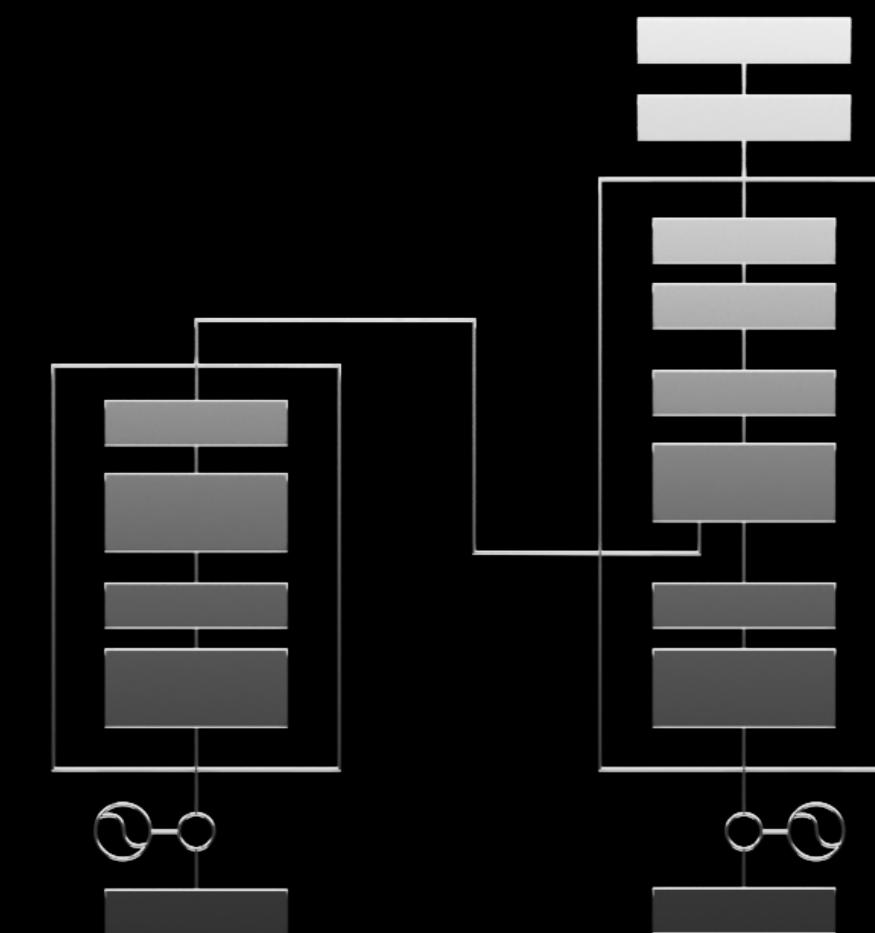
Put climate information on your
fingertips

EARTH-2 — 3 miracles

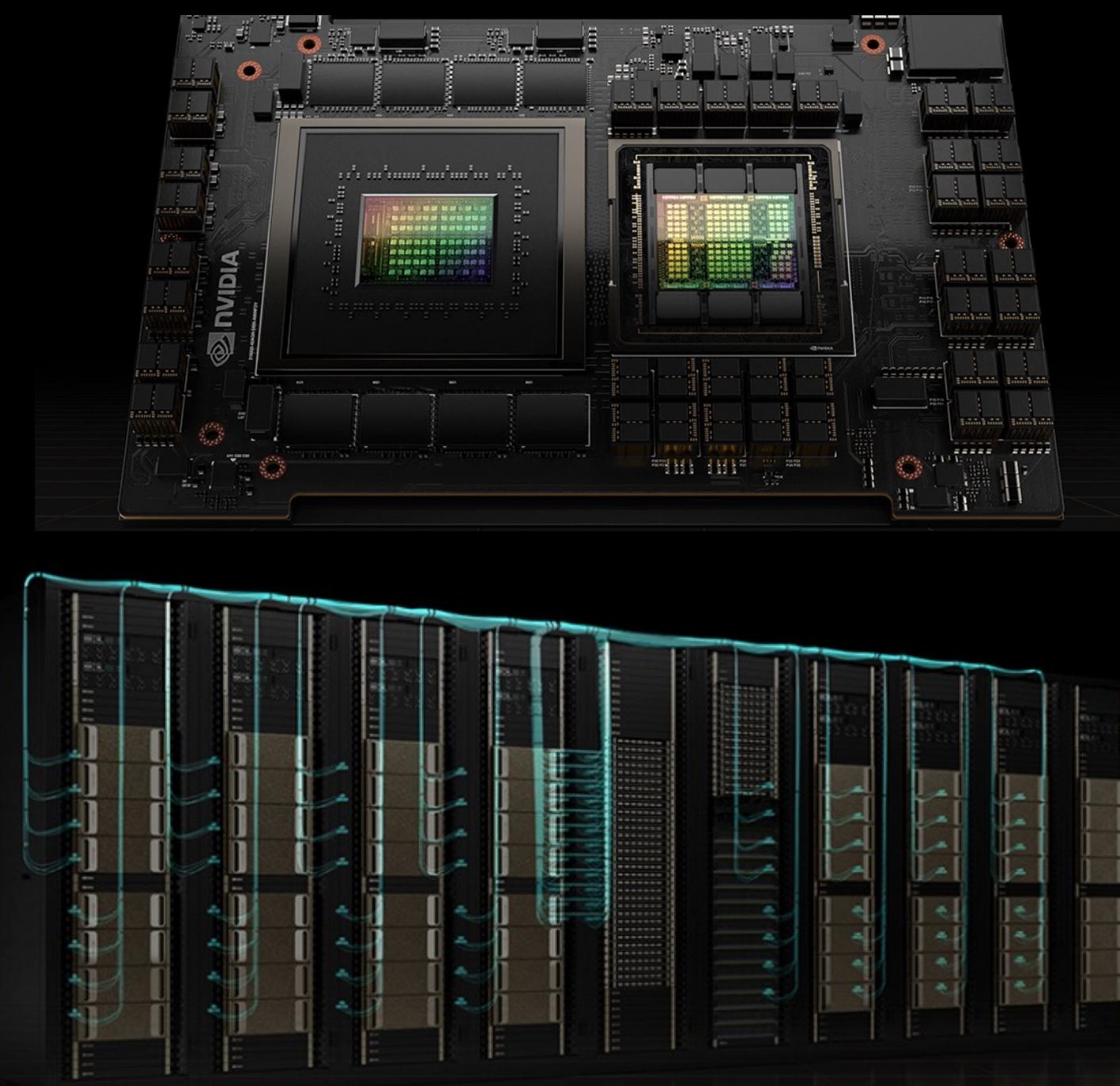
NVIDIA ACCELERATED COMPUTING



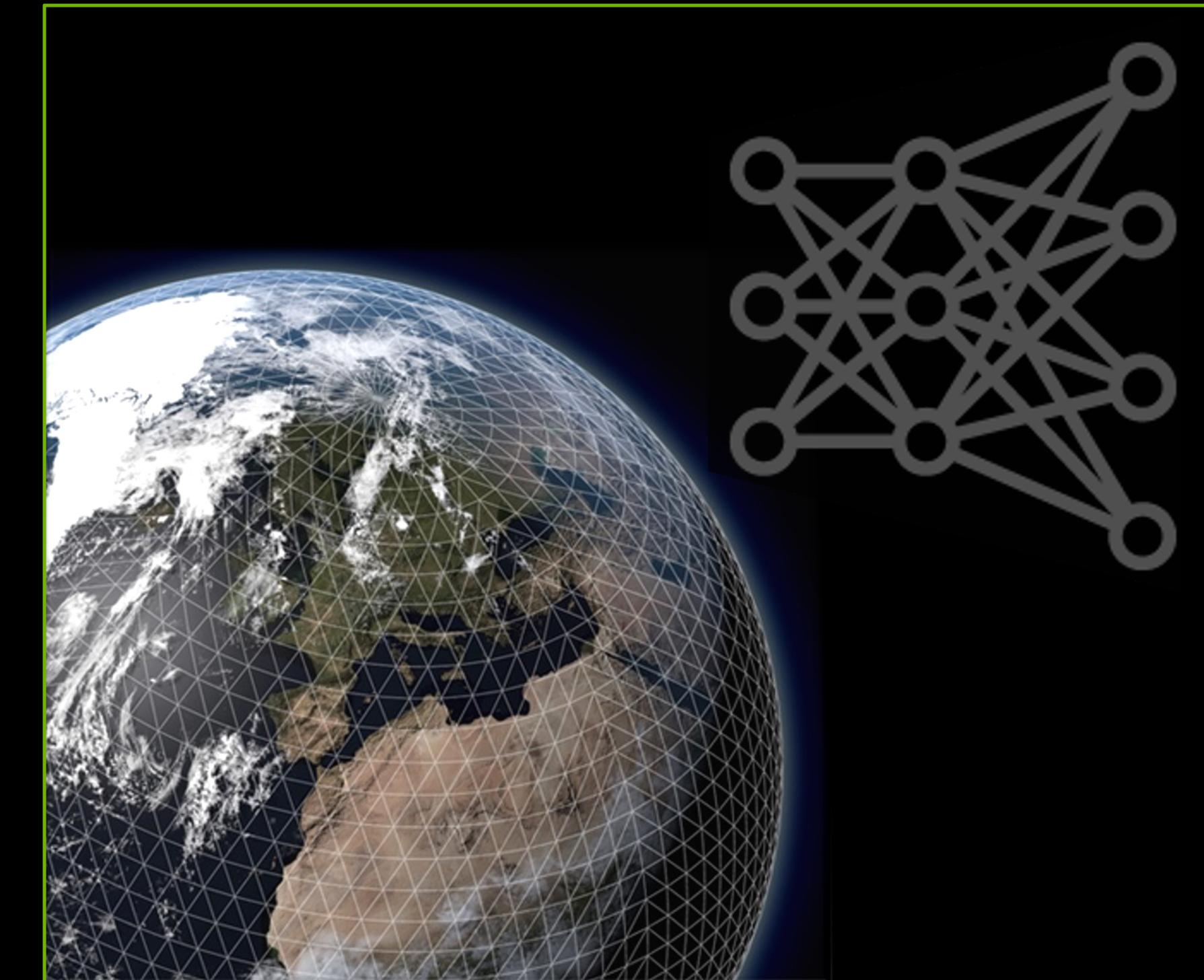
ACCELERATED DATA PROCESSING AI WEATHER MODELS



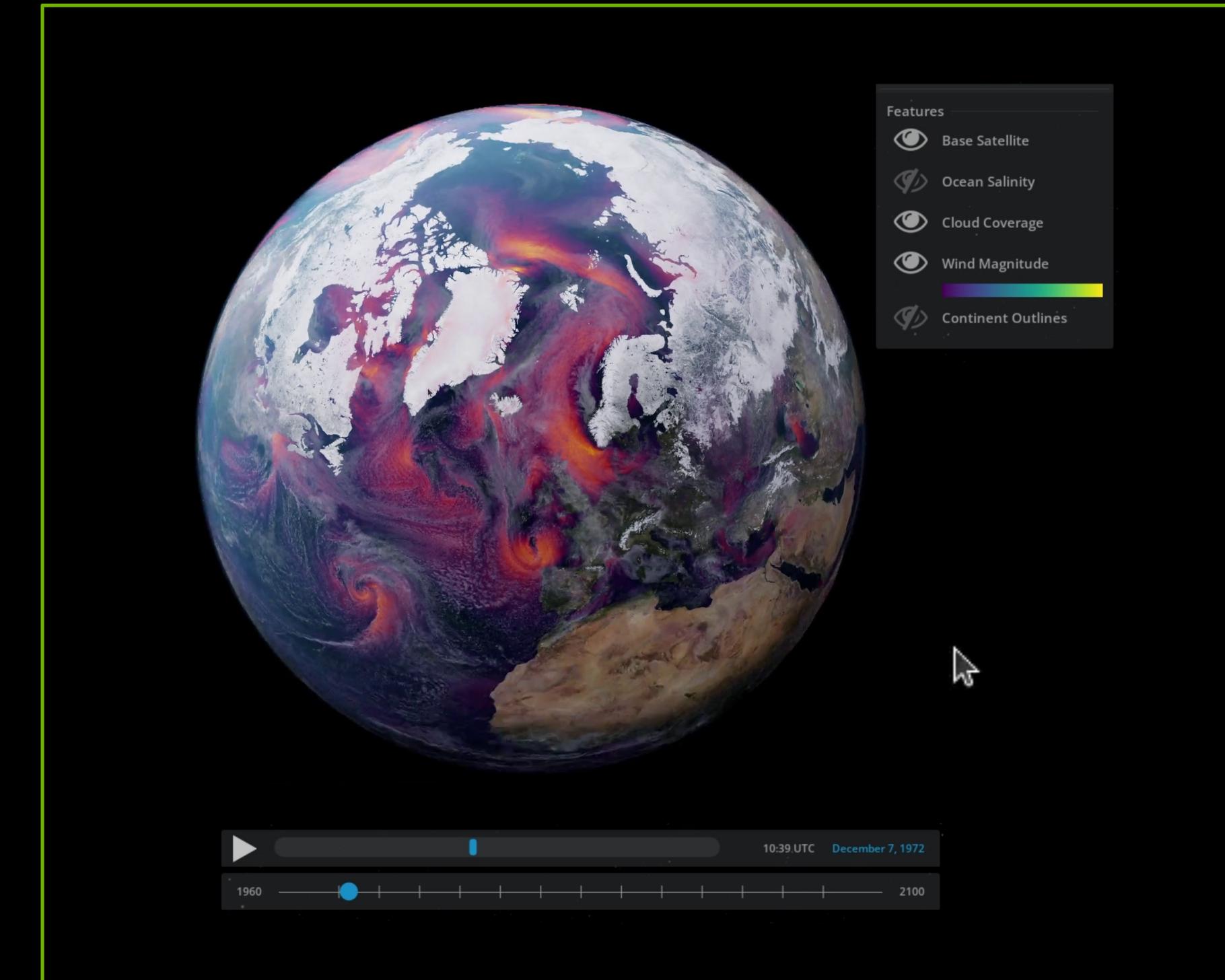
INTERACTIVE DIGITAL TWINS



GPUs, PODs & SUPER-PODs
DGX CLOUD



NVIDIA RAPIDS,
NVIDIA MODULUS, EARTH-2 STUDIO



NUCLEUS
NVIDIA OMNIVERSE

Earth-2's mandate: Achieve "3 miracles" -- Jensen Huang,

Remarks to Berlin Summit for Climate Simulation, Berlin, July 2023.

MIRACLE #1

Km-scale Simulations

30,000 SYPD

30 MW

TO

Generate hi-fidelity data

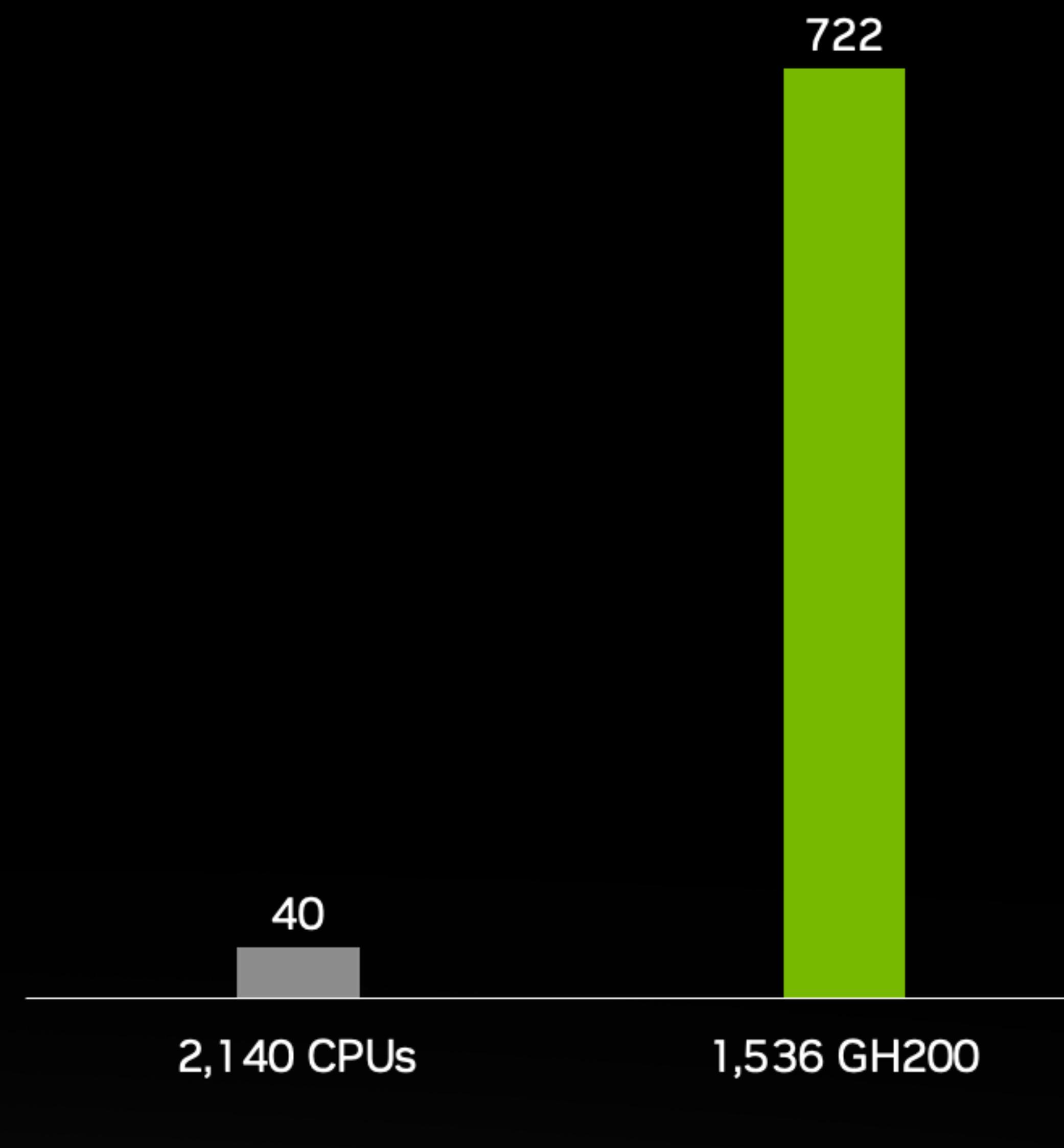
energy efficiently

Jensen Huang, [Summit for EVE Keynote](#)



Accelerating the ICON global cloud resolving model on Grace-Hopper

ICON 2.5km Simulation at the 1 MW
Simulations Days per Day (SDPD)
Coupled Atmosphere-Ocean @20 sec Time Step



Grace-Hopper Accelerates Simulations 20X

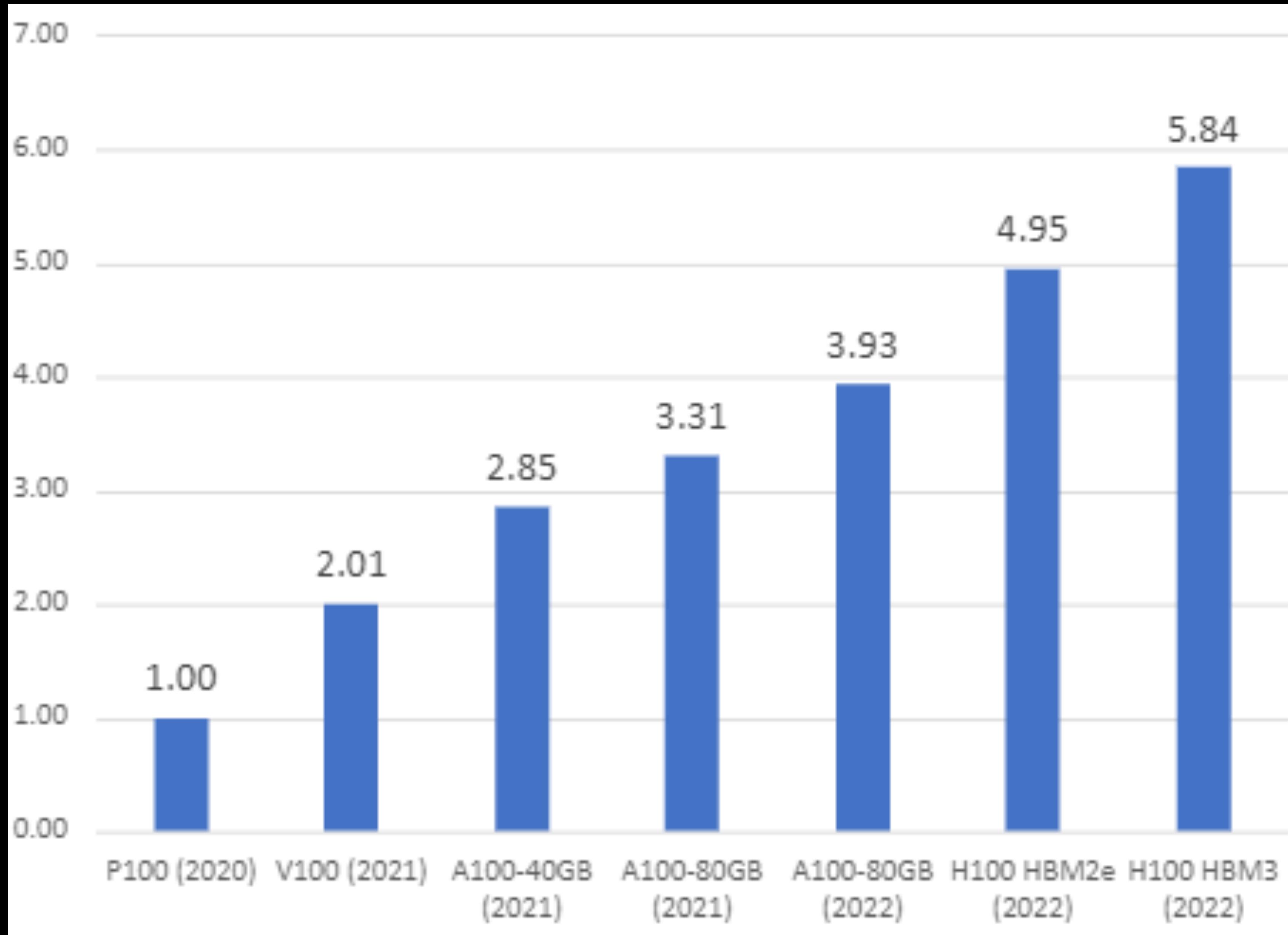
30,000 SYPD < 64,000 GH200 < 40 MW

18X Energy-Efficiency from Accelerated Computing



ICON GPU Improvement ~6x Over 2 Year Span

CSCS, DKRZ, MPI-Met, MeteoSwiss



ICON ~6x from P100 to H100

- Spans 4 generations
- Single GPU results
- Coarse 160km, 80v
- H100 preliminary

Source: NVIDIA, March 2021 Dr. Dmitry Alexeev

GTC On-Demand

<https://www.nvidia.com/en-us/on-demand/>

The screenshot shows a video player interface for the NVIDIA On-Demand platform. At the top, the NVIDIA logo is visible, along with a 'Log Out' button and language selection ('EN'). Below the header, the 'NVIDIA On-Demand' logo is displayed, followed by links to 'Featured Playlists', 'My Channel', 'FAQ', and 'Advanced Search'. The main content area features a video player window. In the video, a man (Thomas Schulthess) is speaking on stage at the ETH Zurich GTC event. He is wearing a white polo shirt and red shorts, gesturing with his hands. The background includes a green screen with the 'GTC' logo and a large blue graphic of a brain with the text 'ALPS'. The video player interface includes standard controls like play/pause, volume, and a progress bar indicating 12:00 / 56:40. To the right of the video, a transcript window is open, showing a list of subtitles corresponding to the video content. The transcript entries are as follows:

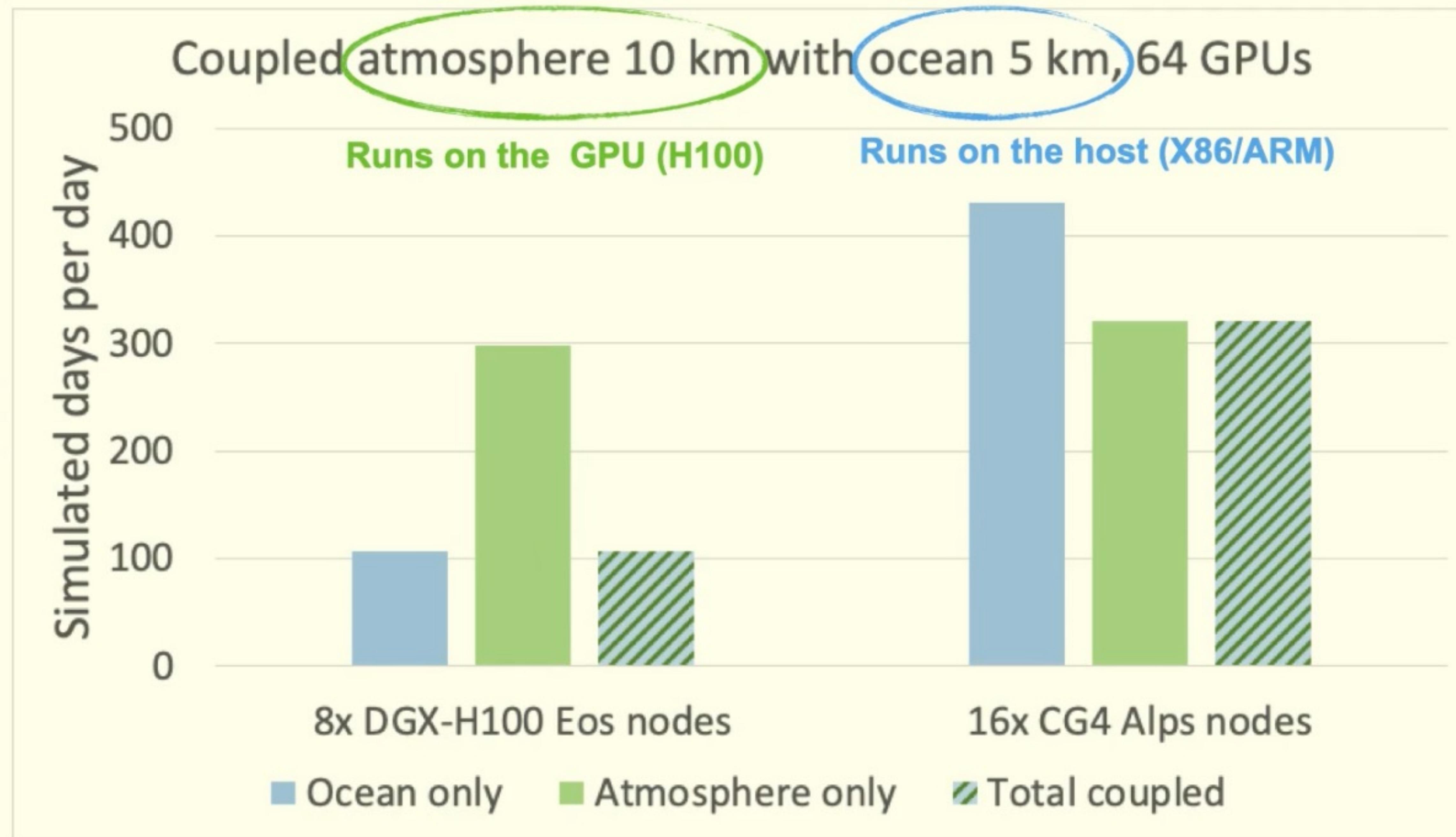
Time	Text
00:05	>> good morning. Thank you for attending.
00:08	My role is a program manager. I am pleased and honored to
00:24	introduce Thomas today. He holds a chair.
00:35	As a director, he is interested in competing infrastructure.
00:46	We have had a long history and collaboration with Thomas
00:53	working with different models. I am very interested to hear the
01:00	update he has to share. Thank you.
01:02	>> OK. It is hard to see.

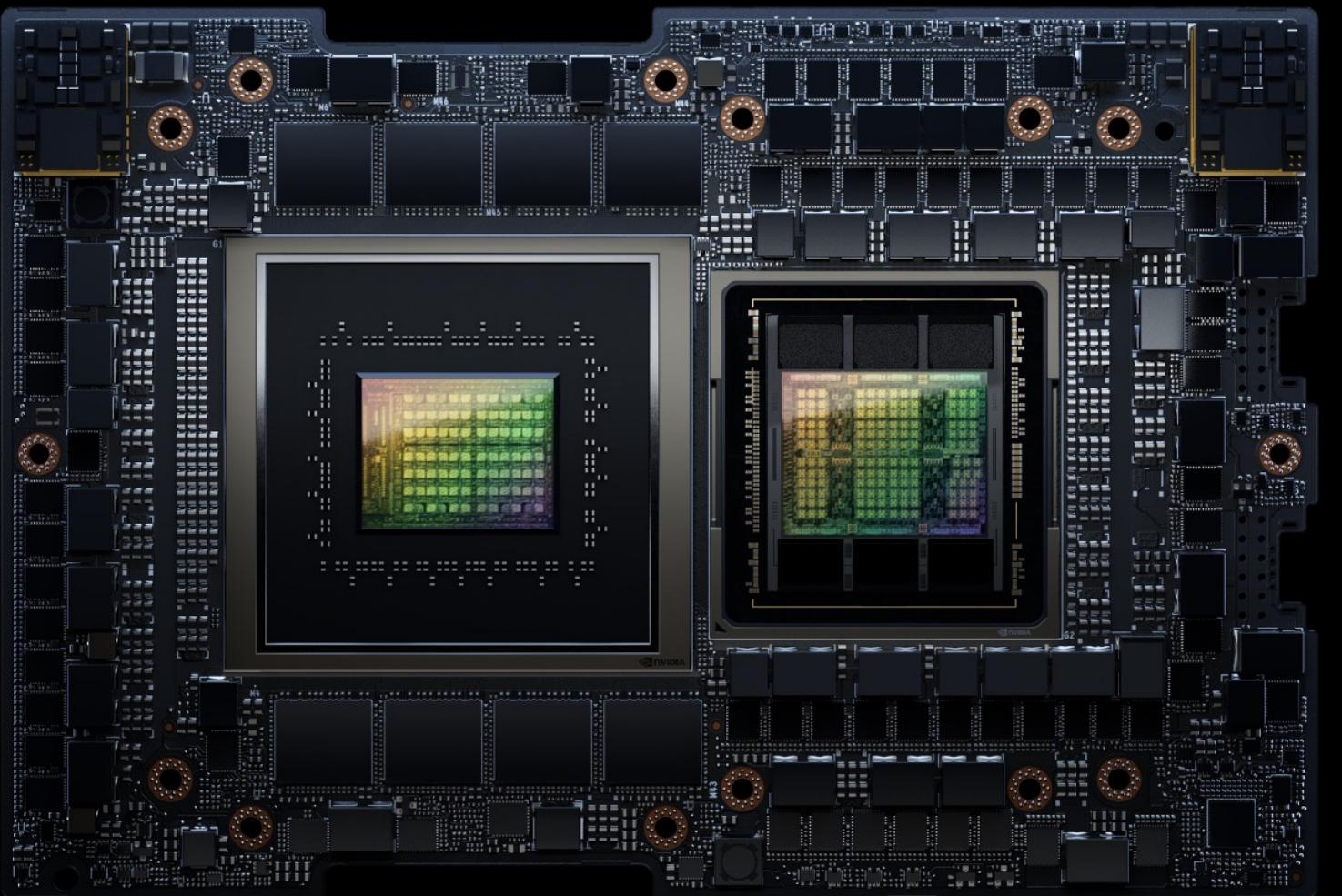
Below the video player, the title 'Early Science with Grace Hopper at Scale on Alps' is displayed, along with the speaker's name, 'Thomas Schulthess, Director, ETH Zurich/The Swiss National Supercomputing Center (CSCS)'. There are also 'Share', 'Favorite', and 'Add to list' buttons.

Early Science with Grace Hopper at Scale on Alps - Thomas Schulthess, Director, ETH Zurich/CSCS

<https://www.nvidia.com/en-us/on-demand/session/gtc24-s62157/>

The latest results from this morning (03/18/2024)

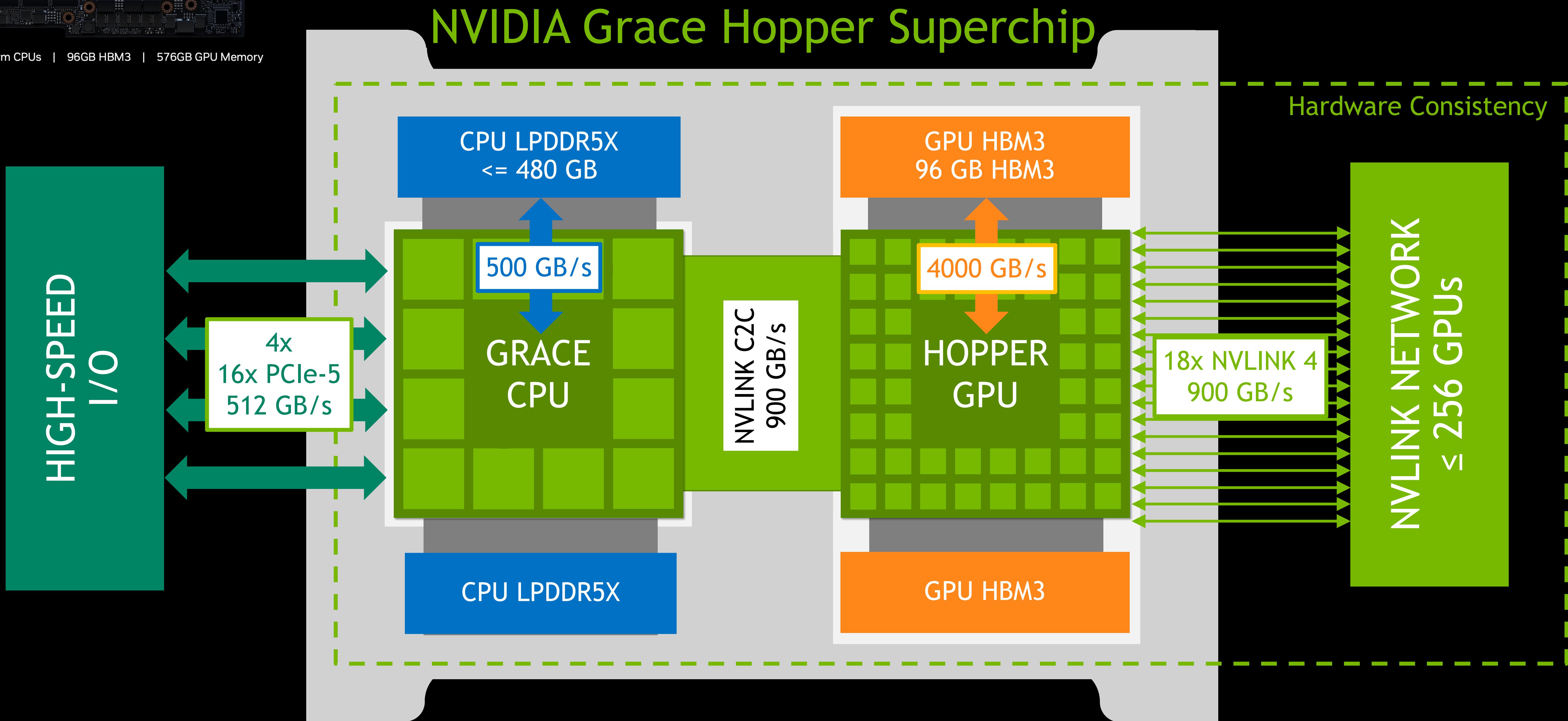




Grace Hopper Superchip

GPU can access CPU memory at CPU memory speeds

4 PetaFLOPS TE | 72 Arm CPUs | 96GB HBM3 | 576GB GPU Memory



NVIDIA has a long history of accelerating physical climate simulation codes

Examples of collaborations with atmospheric modeling groups worldwide

Global:



Model
E3SM, MMF,
HOMEXX, SCREAM

Organizations
US DOE: ORNL, SNL

Funding Source
E3SM, ECP



Model
MPAS-A
FV3 Dycore

Organizations
NCAR, UWyo, IBM
NOAA, AI2

Funding Source
WACA II
SENA, AI2



Model
NUMA/NEPTUNE
IFS

Organizations
US Naval Res Lab, NPS
ECMWF

Funding Source
ONR
ESCAPE, US DOE



Model
GungHo/LFRic

Organizations
MetOffice, STFC

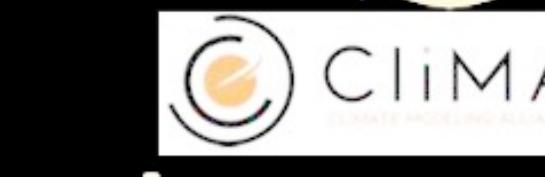
Funding Source
PSyclone



Model
ICON
GEOS-5

Organizations
DWD, MPI-M, CSCS, MCH
NASA GMAO

Funding Source
PASC ENIAC
NASA



Model
CLIMA/NUMA

Organizations
CLIMA (NASA JPL, MIT, NPS)

Funding Source
Private, US NSF



Regional:



Model
AROME
COSMO

Organizations
Meteo France
MCH, CSCS, DWD

Funding Source
MF/CNRS
PASC GridTools



Model
AceCAST-WRF

Organizations
TempoQuest

Funding Source
Venture backed



Earth-2's mandate: Achieve "3 miracles" -- Jensen Huang,

Remarks to Berlin Summit for Climate Simulation, Berlin, July 2023.

MIRACLE #2

AI Emulation of the Full State Vector
Any Region
Any Time Period

TO

To enable low-latency interactivity
with exabytes of data

Jensen Huang, [Summit for EVE Keynote](#)



2023 was a remarkable year for AI weather prediction

Global, data-driven, 25-km forecasting has exited its infancy



DeepMind's "GraphCast"

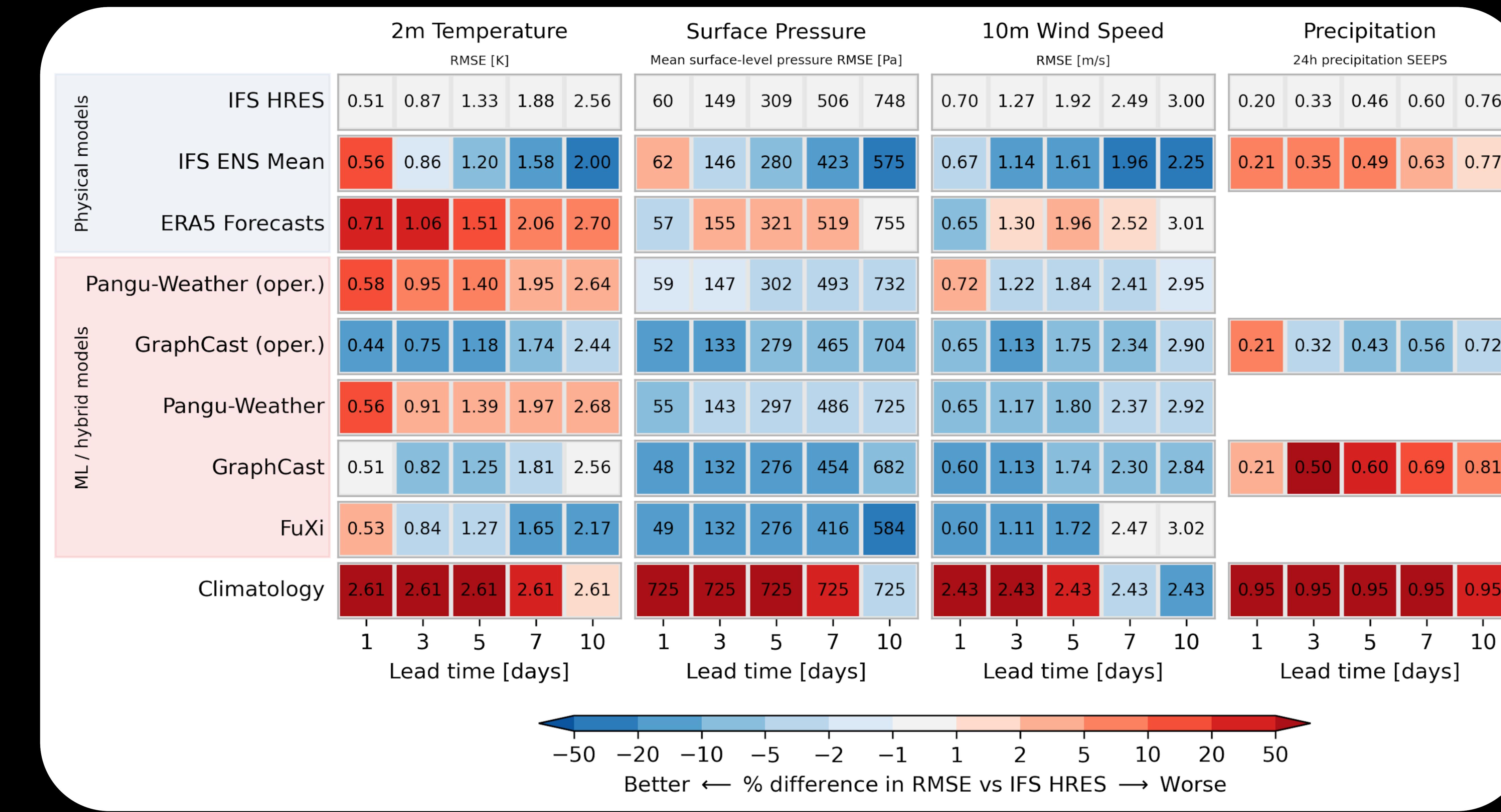
A screenshot of the ECMWF (European Centre for Medium-Range Weather Forecasts) website. The header features the ECMWF logo and navigation links for "Who we are", "What we do", "Jobs", "Media centre" (which is highlighted in bold), "Suppliers", and "Location". The main content area has a dark background with a digital interface overlay showing data grids and numbers. A prominent headline reads "ECMWF unveils alpha version of new ML model". Below the headline, the date "13 October 2023" and the author "The AIFS team" are listed. A link "◀ View all AIFS blog posts" is provided. A detailed description of the launch follows: "ECMWF is today launching a newborn companion to the IFS (Integrated Forecasting System), the AIFS, our Artificial Intelligence/Integrated Forecasting System (one "I" covering both Intelligence and Integrated)".

ECMWF (European Centre for Medium-Range Weather Forecasts), the world's #1 weather prediction center, just launched their AI weather model in October, 2023.

ECMWF makes its own AI weather model

Intriguing Deterministic Skill on Medium-Range Timescales

Source: [Weather Bench 2.0 / April 2024](#)

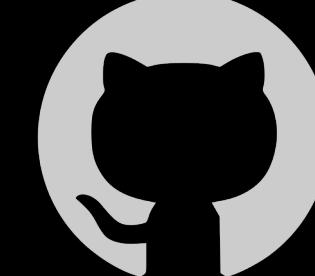
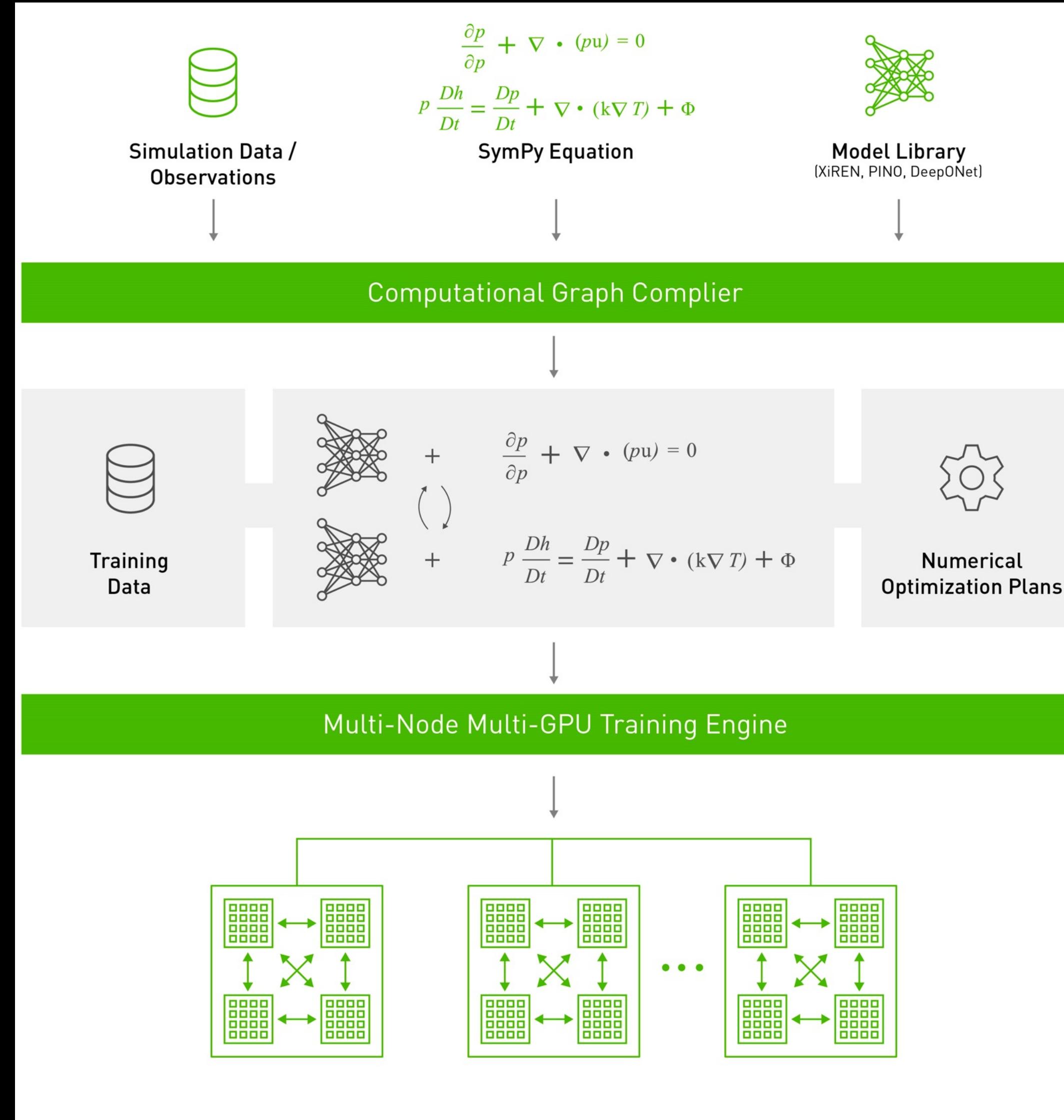


Better ← Δ% difference in RMSE vs IFS HRES → Worse
Scores for 2020 evaluated against IFS analysis or ERA5

NVIDIA Modulus

Open-Source Platform for Developing Physics-Based Machine Learning

Training Neural Networks using both
Data and Governing Equations



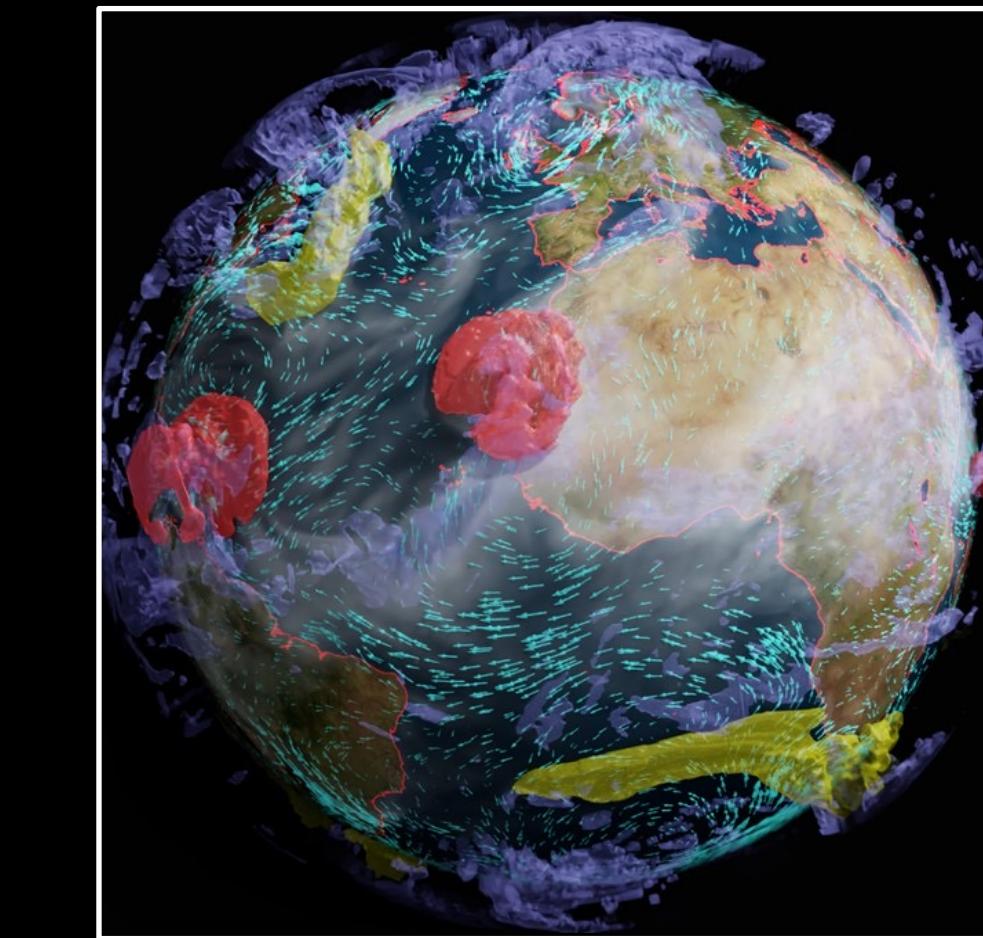
<https://github.com/NVIDIA/modulus>

Advancing Scientific Discovery With Modulus

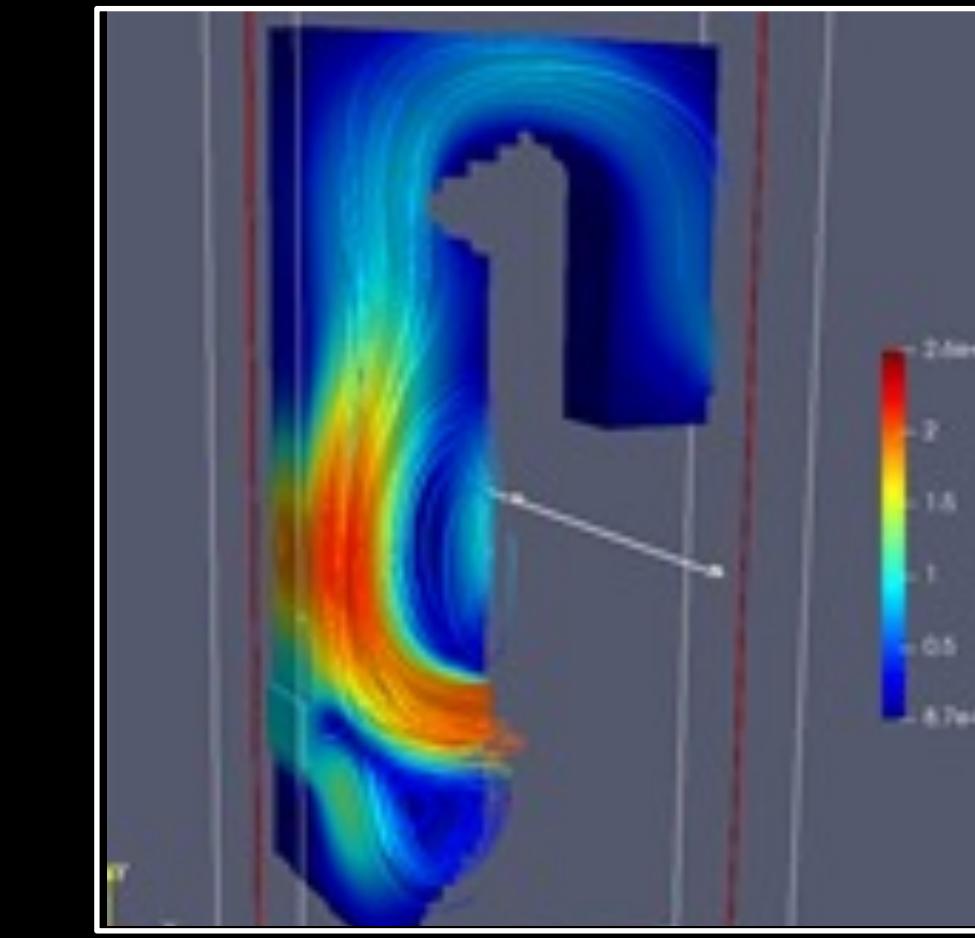
Renewable Energy
Siemens Gamesa: 4000X Faster
wind turbine wake optimization



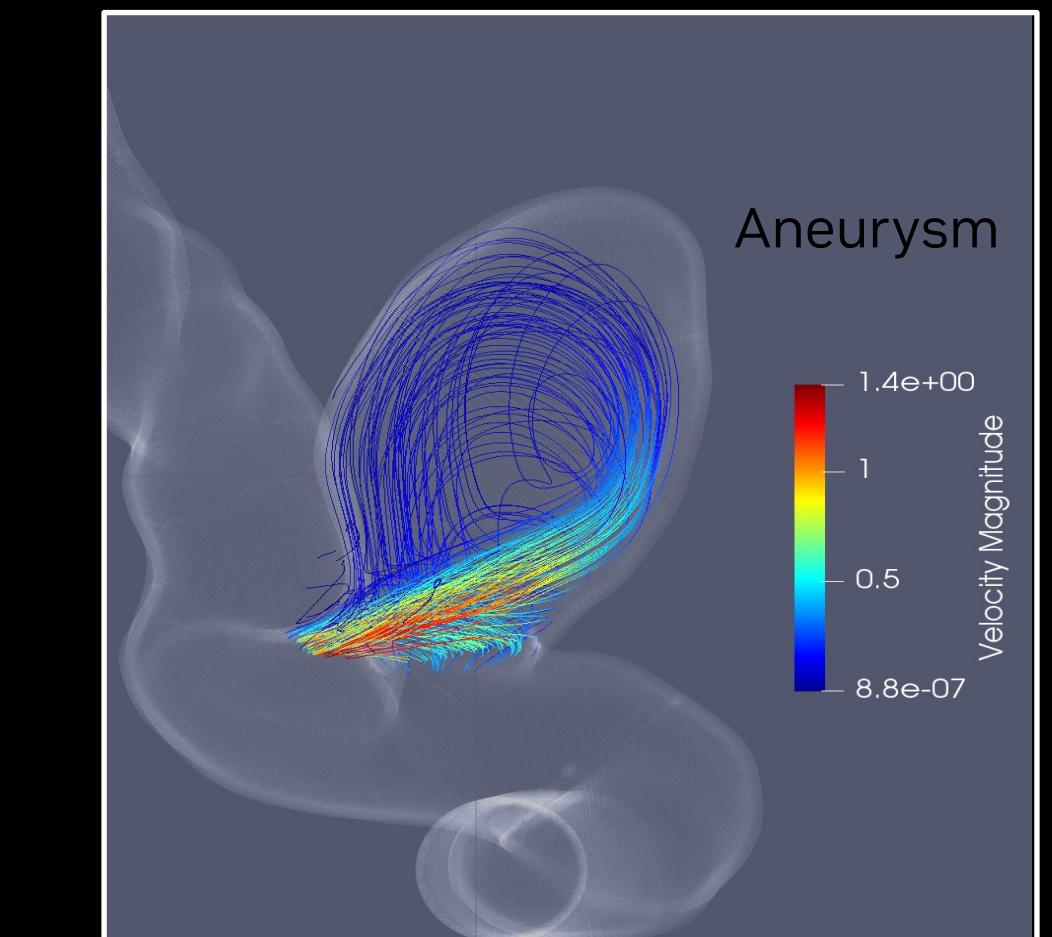
Climate Change
45,000X Faster extreme weather
prediction with AI Weather Models



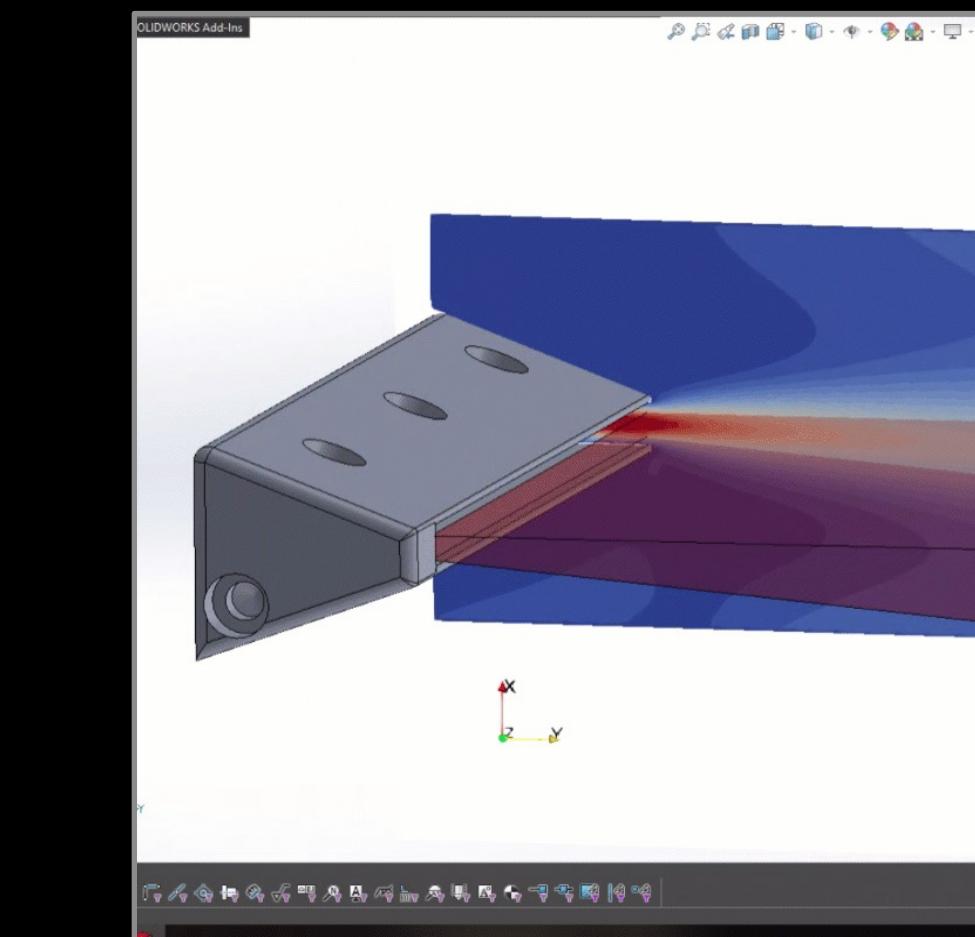
Industrial HPC
NETL: 10,000X Faster build of high-
fidelity surrogate models



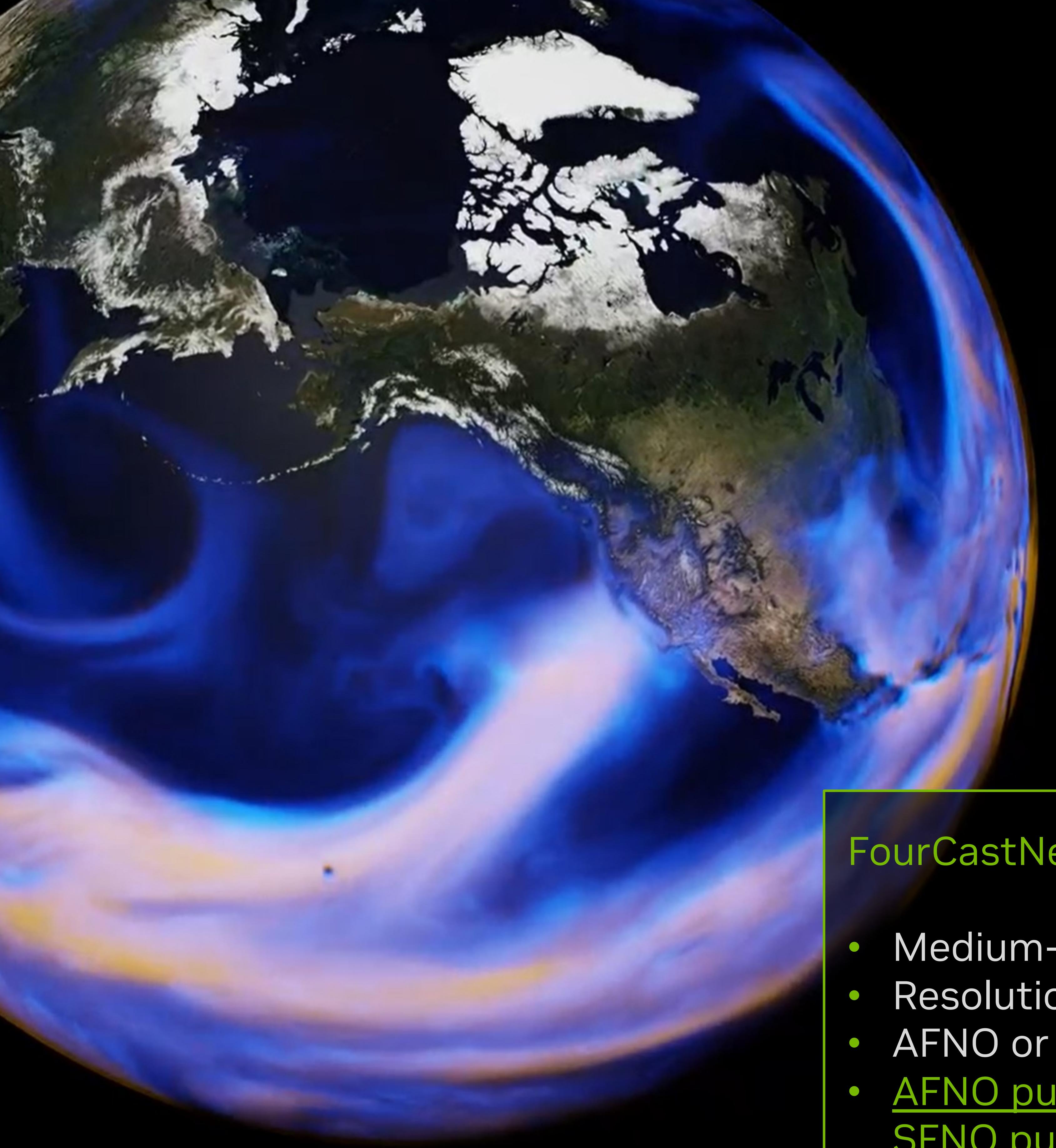
Healthcare
High-fidelity results faster for
blood flow in inter-cranial aneurysm



Digital Twins
Kinetic Vision: Design optimization
using parameterized models



Support by NVIDIA AI Enterprise / Science and Engineering Teaching Kit available now.



NVIDIA Modulus: Access to Various AI Weather Prediction Models

With Complexity, Open Frameworks are Vital

- Open-source
- Support for multi-GPU and multi-node training
- Includes pipelines to train on ERA5 and similar data sets
- Many architectures already implemented, optimized for GPUs
- Training workflow examples and notebooks

FourCastNet

- Medium-range
- Resolution: 0.25°
- AFNO or SFNO
- [AFNO publication](#)
[SFNO publication](#)

Modulus-GraphCast

- Medium-range
- Resolution: 0.25°
- GNN
- [Link to publication](#)

DLWP

- Medium-range & S2S
- Resolution: 1.4°
- CNN + U-Net
- [Link to publication](#)



<https://github.com/NVIDIA/modulus/tree/main/examples/weather>



Global Weather Models

FourCastNet SFNO

- Medium-range
- Resolution: 0.25°
- SFNO
- [Link to publication](#)
- [Model Checkpoint](#)

PanguWeather

- Medium-range
- Resolution: 0.25°
- Vision Transformer
- [Link to publication](#)
- [Model Checkpoint](#)

FourCastNet AFNO

- Medium-range
- Resolution: 0.25°
- AFNO
- [Link to publication](#)
- [Model Checkpoint](#)

FengWu

- Medium-range
- Resolution: 0.25°
- Transformer
- [Link to publication](#)
- [Model Checkpoint](#)

DLWP

- Medium-range & S2S
- Resolution: 1.4°
- CNN + U-Net
- [Link to publication](#)
- [Model Checkpoint](#)

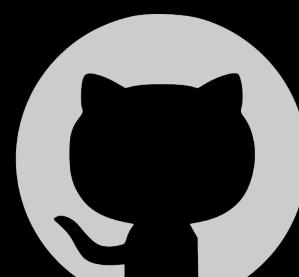
Modulus-GraphCast (coming soon)

- Medium-range
- Resolution: 0.25°
- GNN

AI Weather Model Inference through Earth-2 Studio

High-level API for Rapid Experimentation and Testing

- Open-source
- Building blocks and examples for deterministic and ensemble-based predictions
- Supports multiple data sources CDS (ERA5), ARCO, GFS, HRRR, and on-prem
- Bring your own model ...or select a pre-trained model
- AI-on-top models for increasing set of variables and spatiotemporal resolution



<https://github.com/NVIDIA/earth2studio>

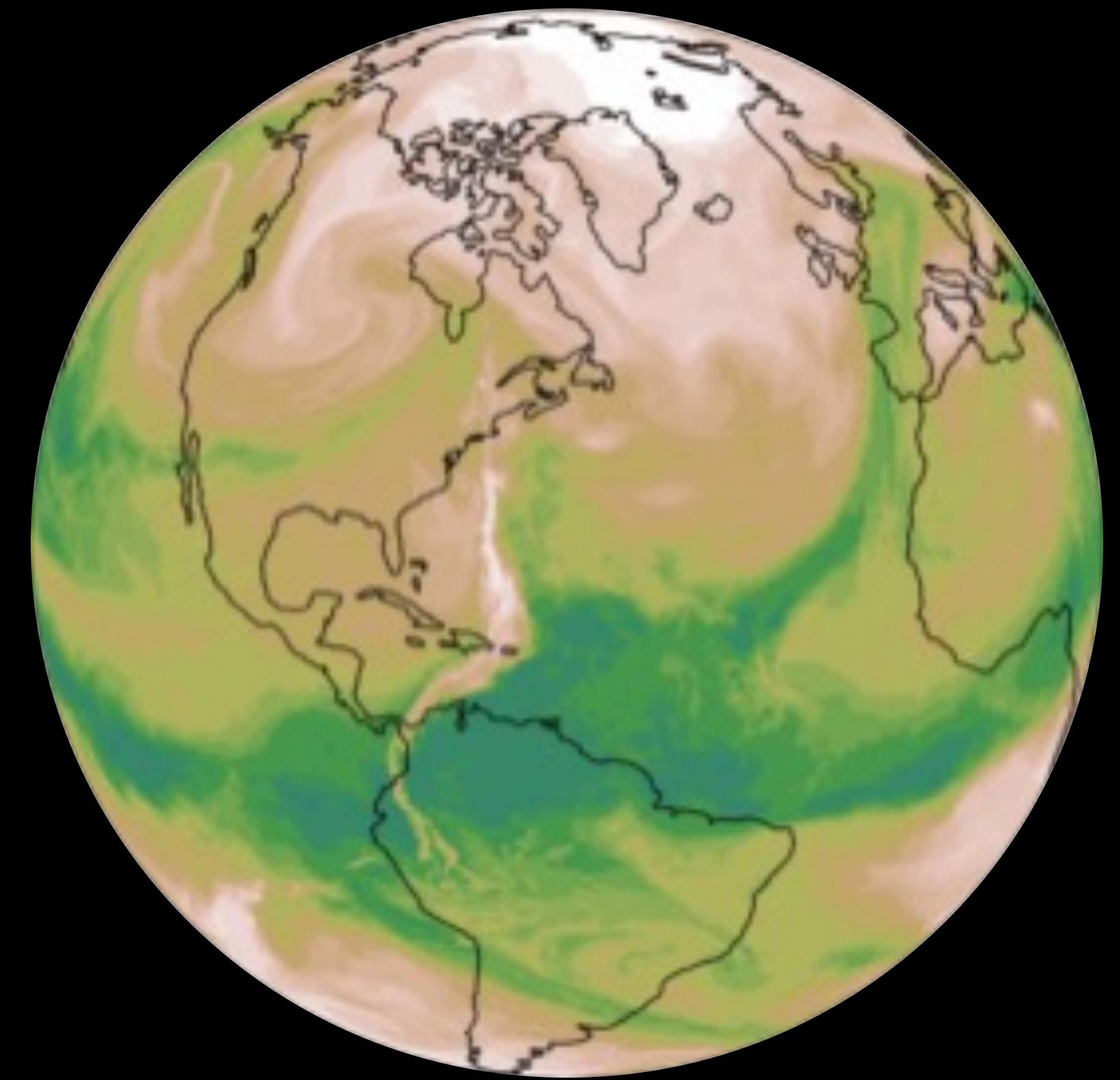


Produce an AI Weather Forecast with a few lines of code

<https://github.com/NVIDIA/earth2studio>

```
from earth2studio.models.px import FCN
from earth2studio.data import GFS
from earth2studio.io import ZarrBackend
import earth2studio.run as run

# Load FourCastNet pretrained model
model = FCN.load_model(FCN.load_default_package())
# Create the data source
data = GFS()
# Create a Zarr IO Backend
io = ZarrBackend()
# Run 20 steps of inference
output_datastore = run.deterministic(["2024-01-01"], 20, model, data, io)
```



Total column water vapour field
at last time step of the forecast

ANALYZING AND EXPLORING TRAINING RECIPES FOR LARGE-SCALE TRANSFORMER-BASED WEATHER PREDICTION

Paper: <https://arxiv.org/abs/2404.19630>, 30th April 2024.

Github: https://github.com/NERSC/swin_v2_weather

ANALYZING AND EXPLORING TRAINING RECIPES FOR LARGE-SCALE TRANSFORMER-BASED WEATHER PREDICTION

JARED D. WILLARD,^a PETER HARRINGTON,^a SHASHANK SUBRAMANIAN,^a ANKUR MAHESH,^b TRAVIS A. O'BRIEN,^c WILLIAM D. COLLINS^{b,d}

^a Lawrence Berkeley National Laboratory, National Energy Research Scientific Computing Center

^b University of California, Berkeley, Department of Earth and Planetary Science

^c Indiana University Bloomington, Department of Earth and Atmospheric Sciences

^d Lawrence Berkeley National Laboratory, Earth & Environmental Sciences Area



Berkeley
UNIVERSITY OF CALIFORNIA

This repository contains the code used for "Analyzing and Exploring Training Recipes for Large-Scale Transformer-Based Weather Prediction" [[paper](#)]

The code was developed by the authors of the preprint: [Jared Willard](#), [Shashank Subramanian](#), [Peter Harrington](#), [Ankur Mahesh](#), [Travis O'Brien](#), and [William Collins](#)

SwinV2_Weather is a global data-driven weather forecasting model that provides accurate short to medium-range global predictions at 0.25° resolution using a minimally modified SwinV2 transformer. SwinV2_Weather outperforms the forecasting accuracy of the ECMWF Integrated Forecasting System (IFS) deterministic forecast, a state-of-the-art Numerical Weather Prediction (NWP) model, at nearly all lead times for critical large-scale variables like geopotential height at 500 hPa (z500), 2-meter temperature (t2m), and 10-meter wind speed (u10m).

SwinV2_Weather is based on the original Swin Transformer V2 architecture proposed in Liu et al. [2022], and we adapt the [Hugging Face implementation of the model](#) for this repository.

HPC Job Launching

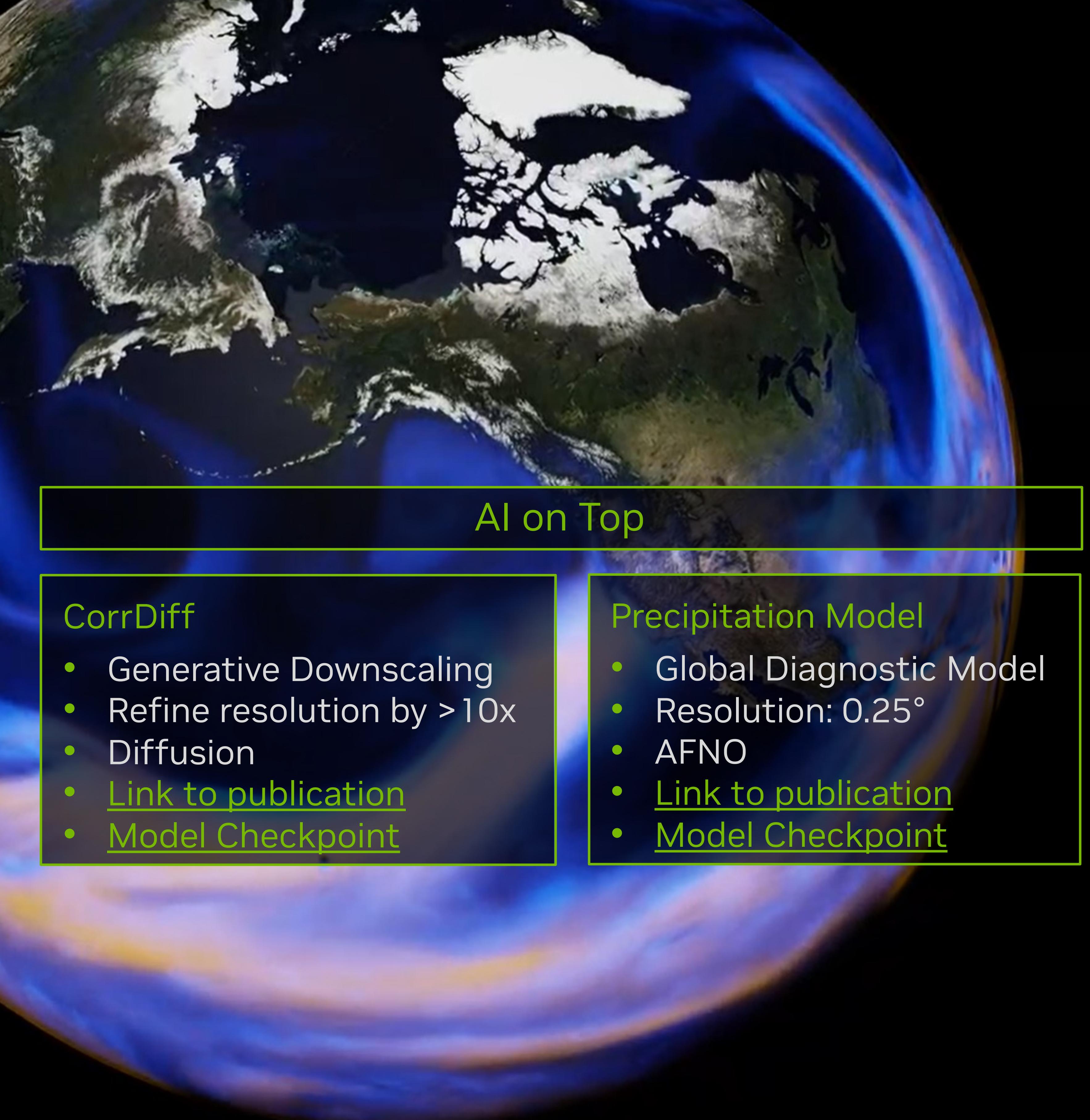
An example launch script for distributed data parallel training on the slurm based HPC cluster perlmutter is provided in [submit_batch.sh](#).

Inference and Scoring:

For inference and scoring we used [Earth2MIP](#). A fork that contains an implementation of our Swin model and directions for inference and scoring can be found at <https://github.com/jdwillard19/earth2mip-swin-fork/>.

If you find this work useful, cite it using:

```
@misc{willard2024analyzing,
    title={Analyzing and Exploring Training Recipes for Large-Scale Transformer-Based Weather Prediction},
    author={Jared D. Willard and Peter Harrington and Shashank Subramanian and Ankur Mahesh},
    year={2024},
    eprint={2404.19630},
    archivePrefix={arXiv},
    primaryClass={cs.LG}
}
```



AI on Top

CorrDiff

- Generative Downscaling
- Refine resolution by >10x
- Diffusion
- [Link to publication](#)
- [Model Checkpoint](#)

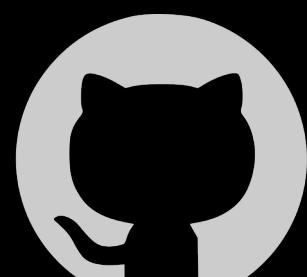
Precipitation Model

- Global Diagnostic Model
- Resolution: 0.25°
- AFNO
- [Link to publication](#)
- [Model Checkpoint](#)

AI Weather Model Inference through Earth-2 Studio

High-level API for Rapid Experimentation and Testing

- Open-source
- Building blocks and examples for deterministic and ensemble-based predictions
- Supports multiple data sources CDS (ERA5), ARCO, GFS, HRRR, and on-prem
- Bring your own model ...or select a pre-trained model
- AI-on-top models for increasing set of variables and spatiotemporal resolution

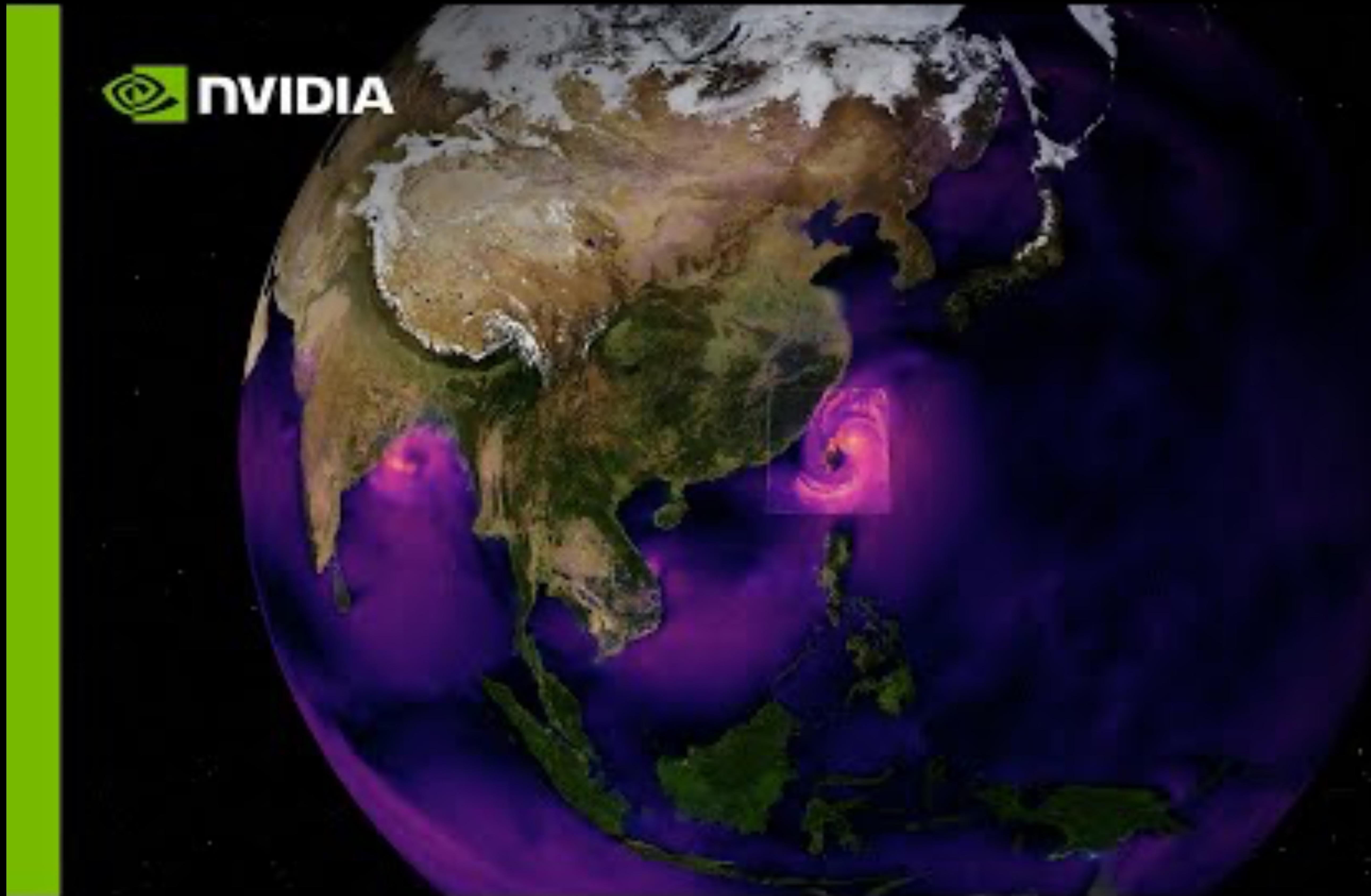


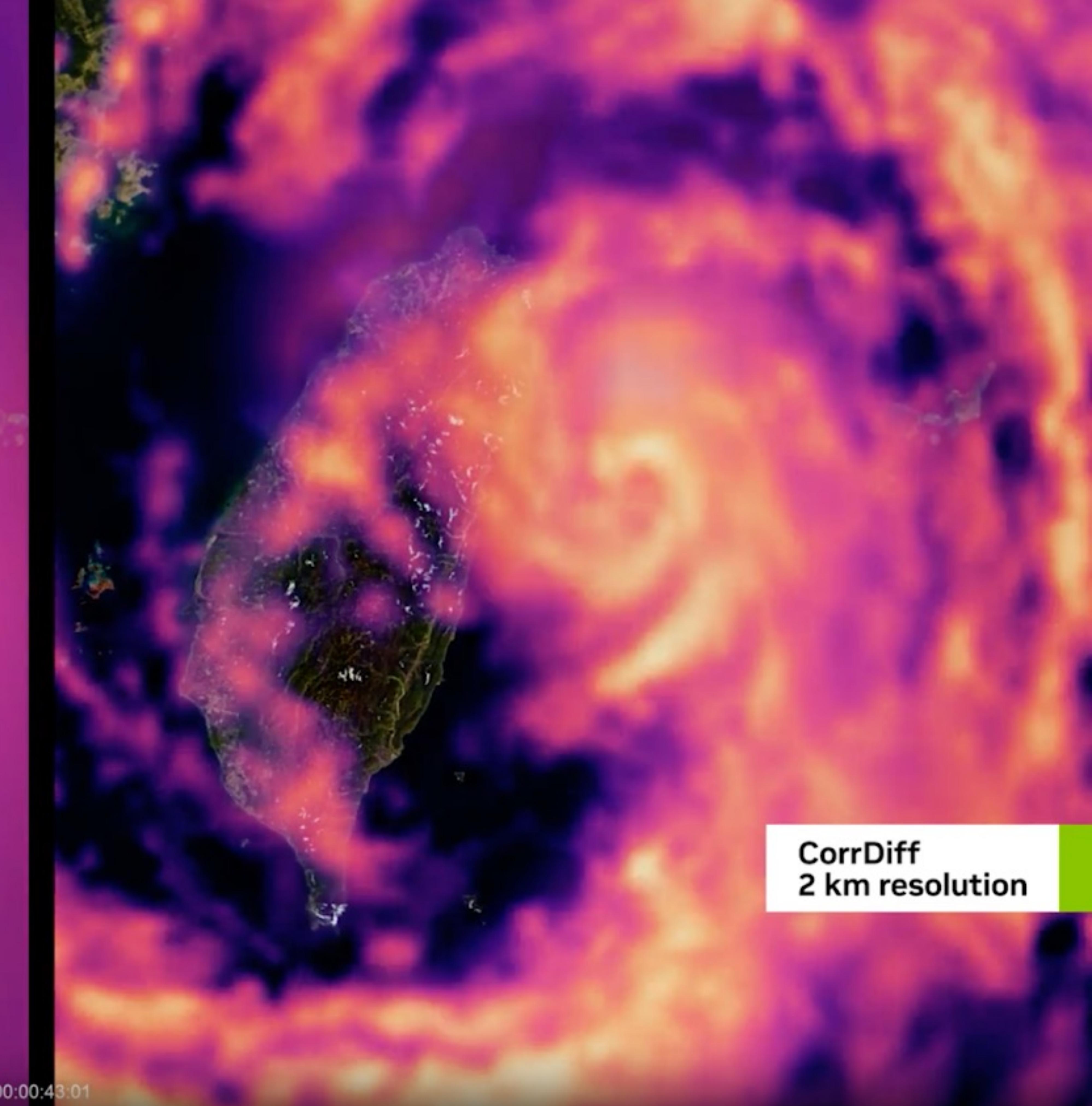
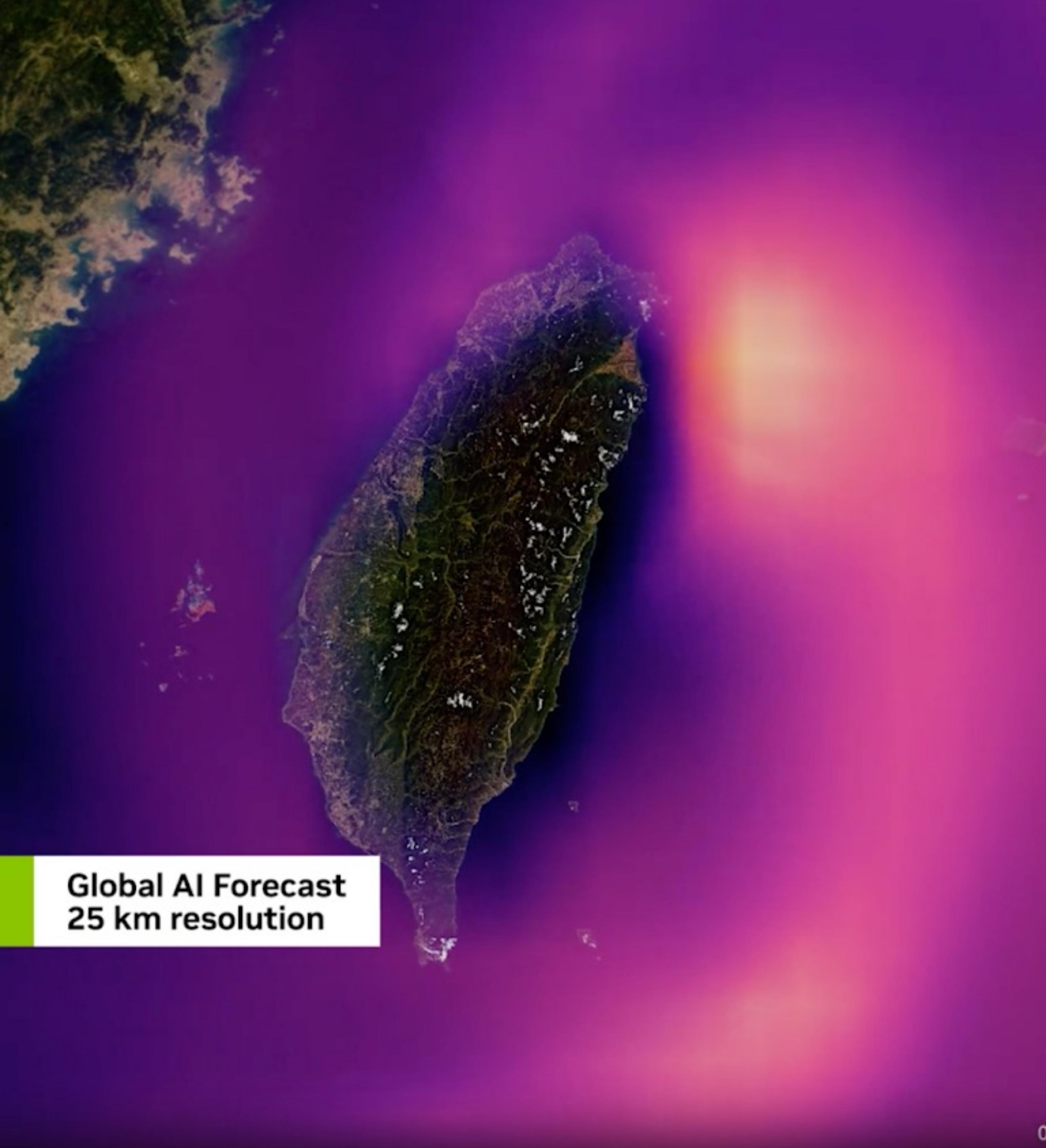
<https://github.com/NVIDIA/earth2studio>



NVIDIA CorrDiff: Resolving Extreme Weather Events With Generative AI

https://www.youtube.com/watch?v=9vEalmsSCrw&ab_channel=NVIDIA



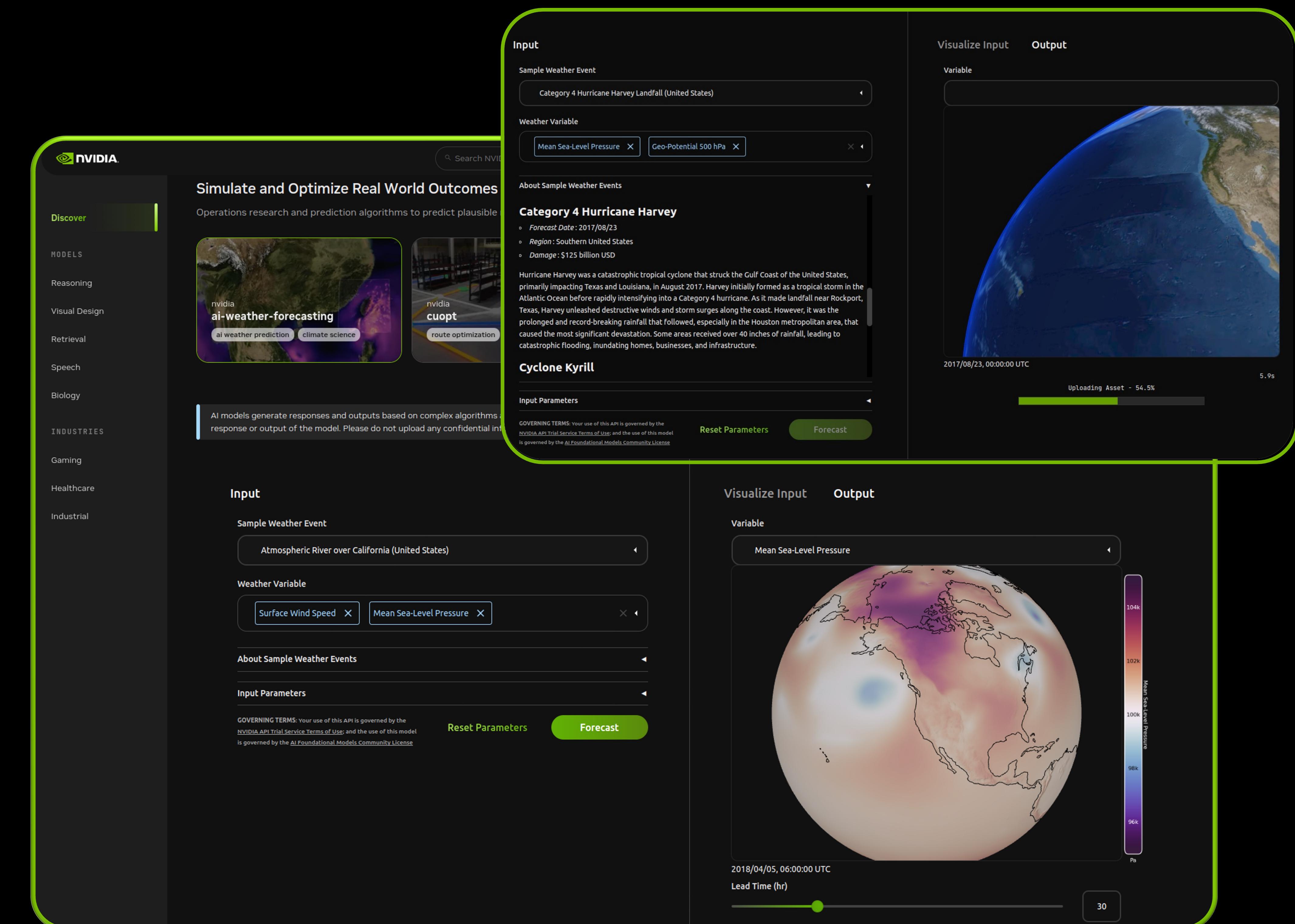
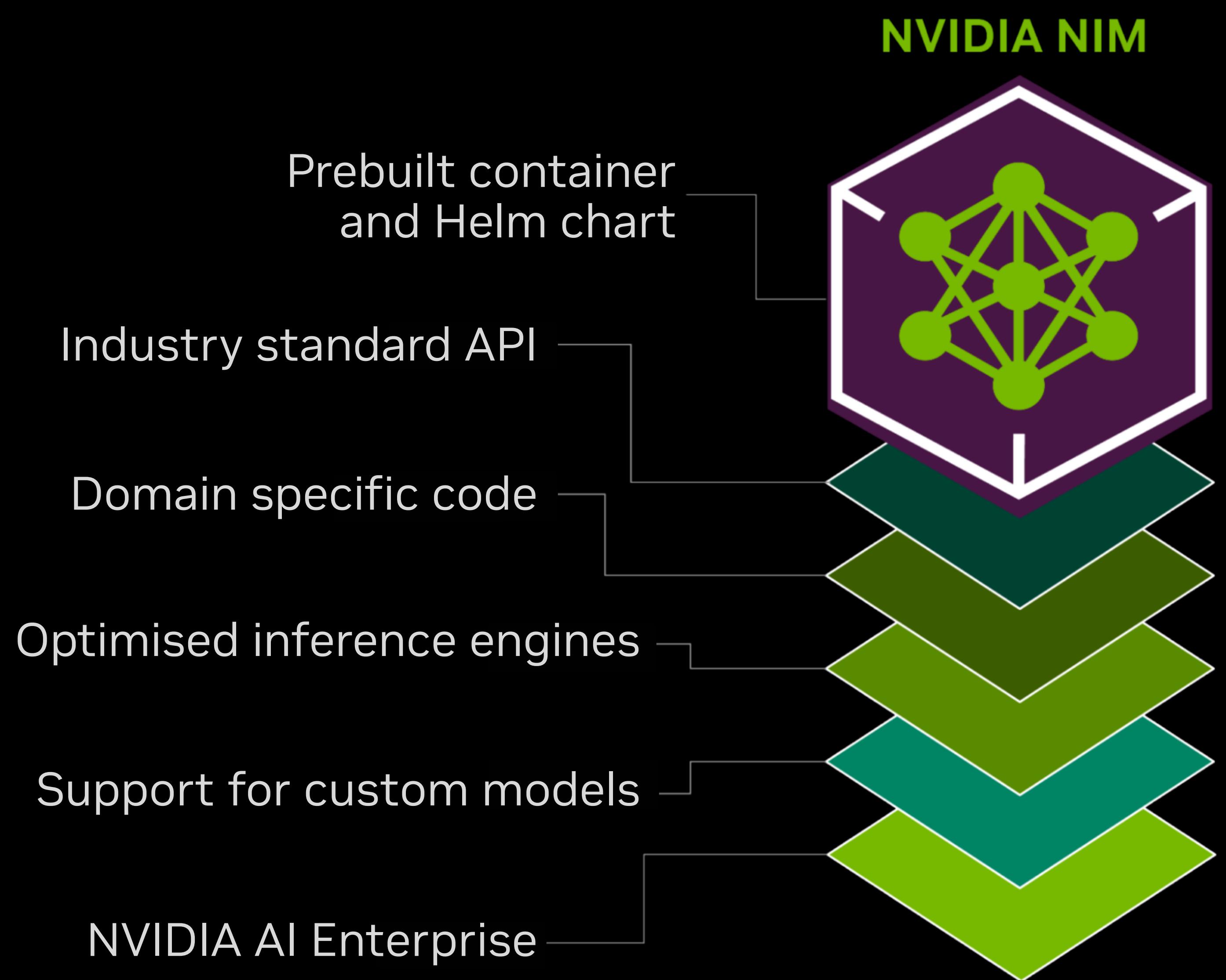


Potencial for Higher Resolution, Larger Ensemble members, Faster and More Efficient.



Production Deployment through NVIDIA Inference Microservices

Deploy Customised Models leveraging Optimised Inference Engines



Click [this link](#) to check out the browser demo!



Earth-2's mandate: Achieve "3 miracles" -- Jensen Huang,

Remarks to Berlin Summit for Climate Simulation, Berlin, July 2023.

MIRACLE #3

Full State Vector Visualization
From Cloud

TO

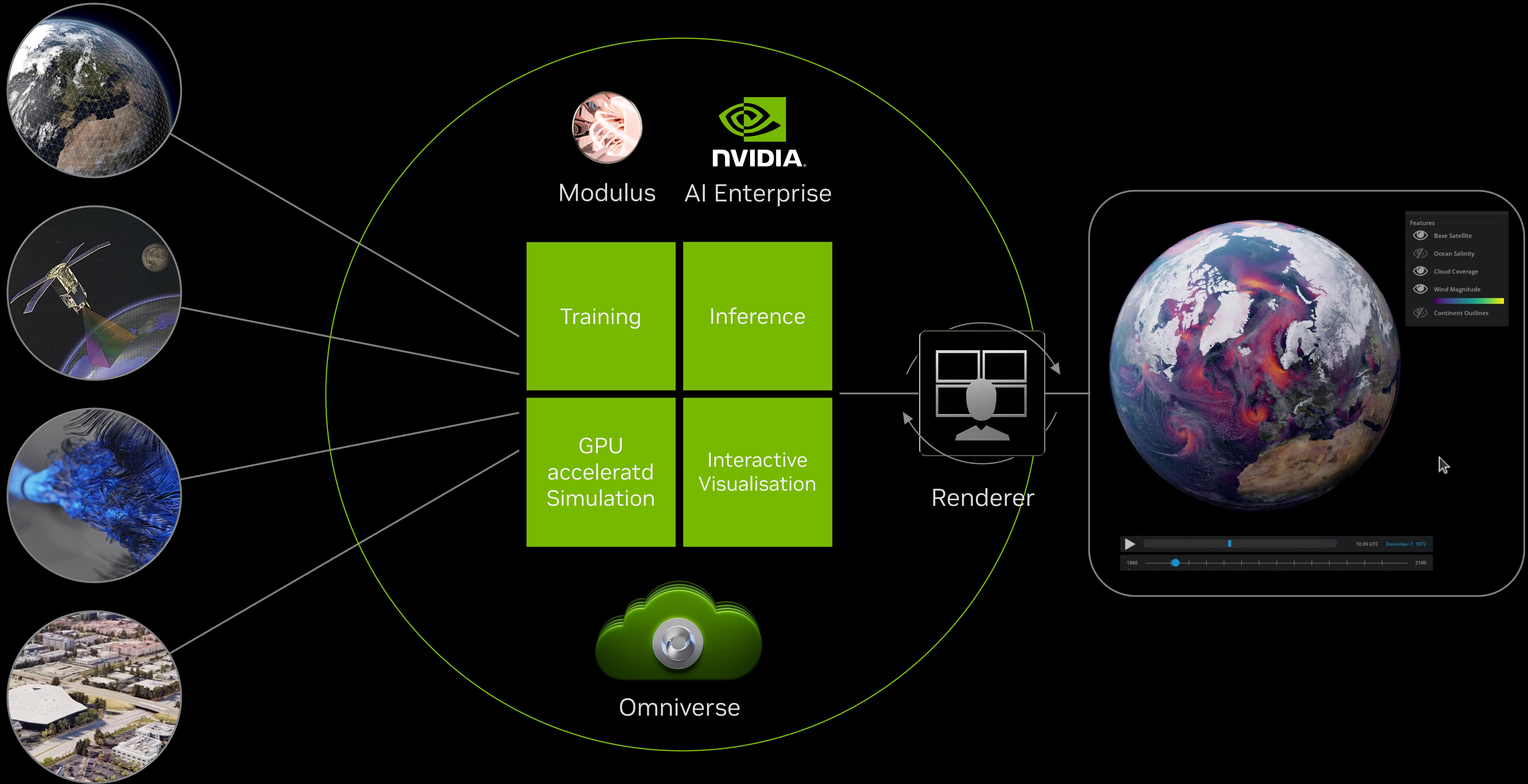
Put climate information on your
fingertips

Jensen Huang, [Summit for EVE Keynote](#)



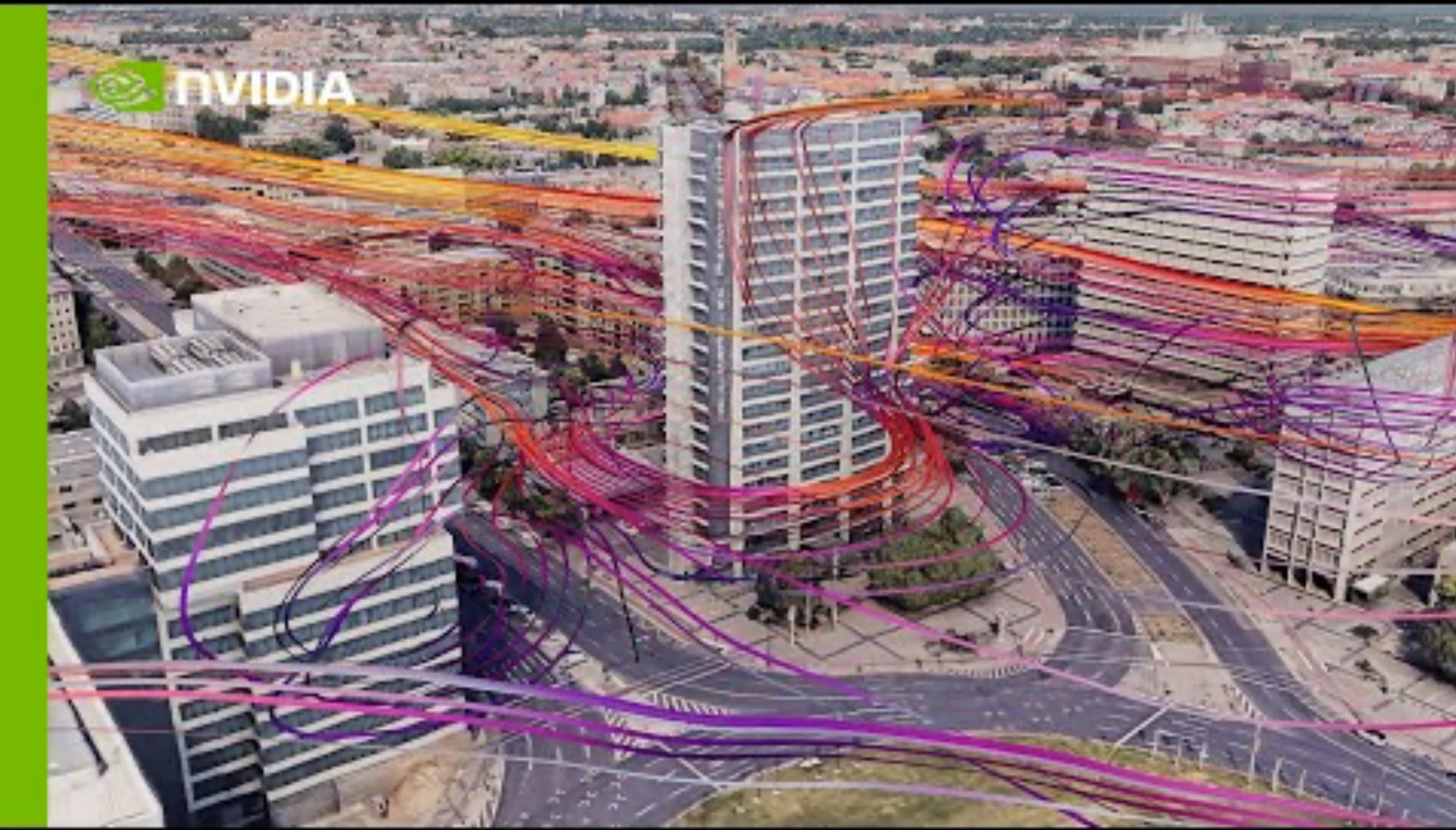
Earth-2 Platform

Connecting Complex Simulation, Data and AI Workflows



Omniverse for Weather and Climate

https://youtu.be/8cQoYcbUG_M



(extend from 2023 Open Hackathon)

NCAR WRF/MPAS Users Workshop 2024

<https://www.mmm.ucar.edu/events/workshops/wrf-mpas>

Advancements in Implementing the MPAS-A Regional Model at the Central Weather Administration

¹Wu, Y.-J., ²W. Wang, ³C.-Y. Chen, ³Y.-L. Chen, ¹S.-L. Huang, ¹B.-S. Lin, ¹L.-F., Hsiao

¹Central Weather Administration, Taiwan

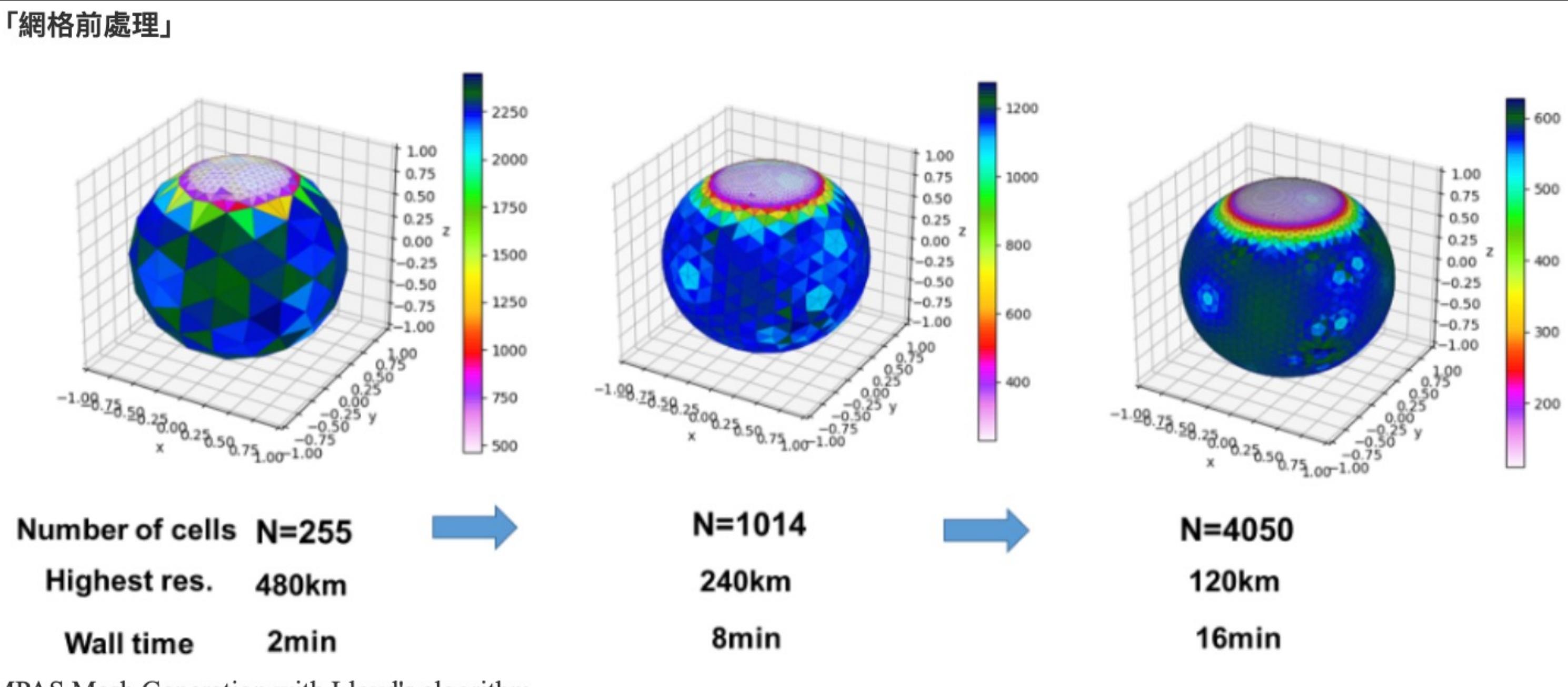
²University Corporation for Atmospheric Research, Boulder, Colorado

³NVIDIA

In our ongoing efforts to enhance weather modeling capabilities at the Central Weather Administration (CWA), we have integrated several key components from the Weather Research and Forecasting (WRF) operational model into the MPAS-A Regional Model. Collaborating with National Taiwan University (NTU), we have developed the cloud physics scheme TCWA1, which has been integrated into the model framework.

One key focus of our work has been comparing the surface wind speeds simulated by MPAS-A with those from the Weather Research and Forecasting (WRF) model. Our analyses have revealed notable discrepancies between the two models, prompting the introduction of the topo_wind option from WRF. This addition aims to mitigate wind speed biases and improve the overall accuracy of our simulations, particularly in complex terrain regions.

The grid generation program plays a critical role in setting up the computational grid for weather simulations. By leveraging GPU acceleration, we achieved significant improvements in the performance of this program. The tailored GPU acceleration techniques, developed in collaboration with the mentors in NVIDIA workshop, allowed for faster data processing and computation, leading to a six-fold speed increase compared to the previous implementation.



MPAS Mesh Generation with Lloyd's algorithm

CWA Mesh Gen 團隊成員來自 氣象署科氣發展組及數值資訊組，將跨尺度預報模式(NCAR MPAS) 網格產製工具加速6.7倍！

* NVIDIA Mentors: Leo Chen, Jay Chen

網格生成是數值模擬的關鍵步驟，製作一個網格需要數個月的時間，使用Fortran 程式進行幾何與質心計算疊代直到收斂。黑客松期間將質心計算的OpenMP區段改為OpenACC。並藉由Nsight Systems + NVTX 工具，幫助團隊快速分析運算瓶頸進行優化。

	gfortran	pgfortran	nvfortran	stripack F77 to F90	stripack: clist ACC
Voronoi computation	0.15	0.15	0.15	0.11	0.03
Centroid computation	0.25	0.24	0.03	0.03	0.03
Total	0.4	0.39	0.18	0.14	0.06
Speedup	1x	1.02x	2.2x	2.85x	6.67x

Acceleration Results

“現在我實在難以想像，如果沒有GPU加速的改進，要完成相同的計算任務需要多長時間，非常感謝 Mentors 幫助” – 氣象署CWA Mesh Gen 團隊

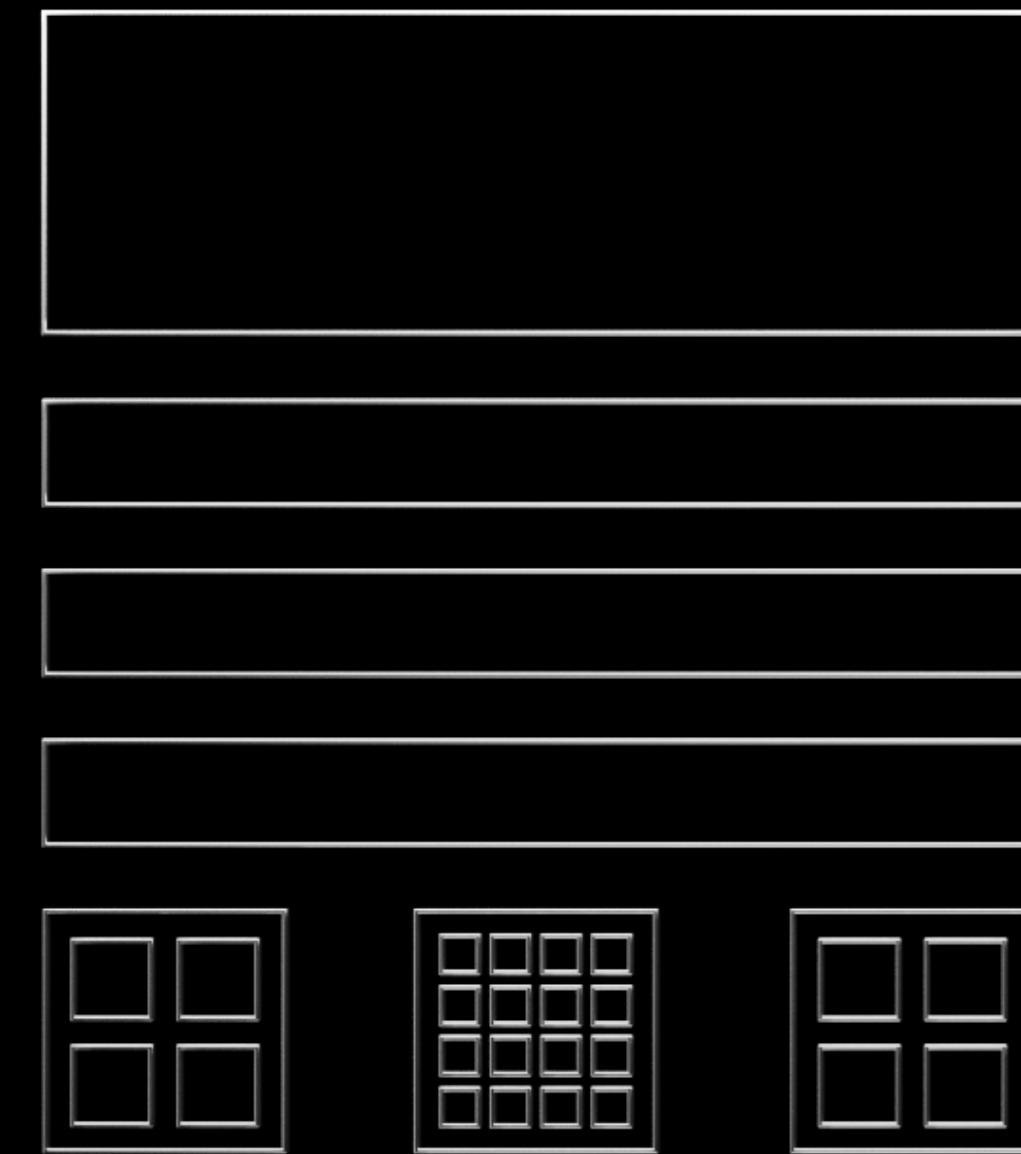
更多資訊請看：

<https://github.com/nqobu/nvidia/raw/main/20231207/Team05.pdf>

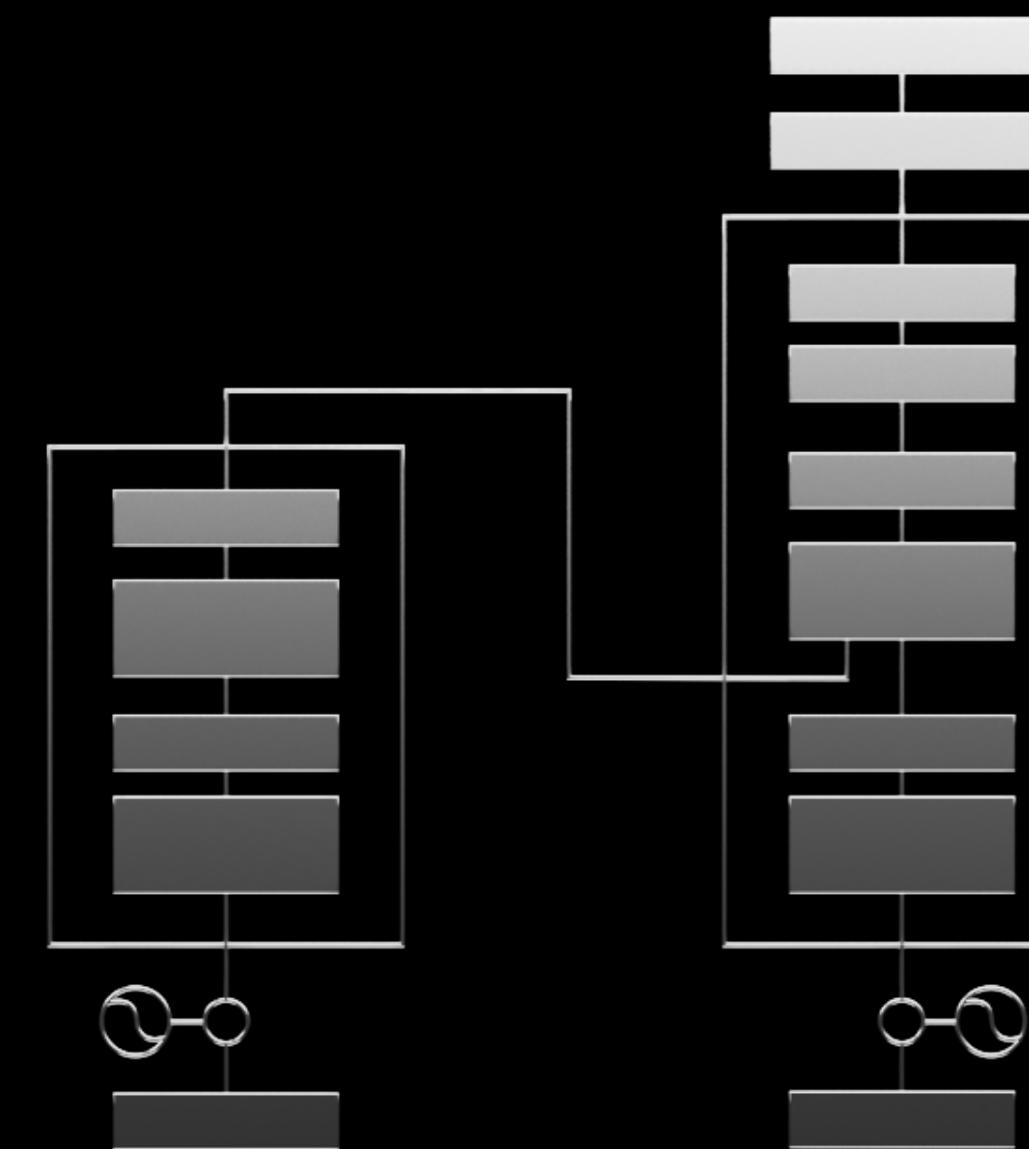


Earth-2 @ Taiwan

SIMULATION



AI



VISUALIZATION



- Energy Efficiency
 - WRF on **Grace CPU Superchip**
- GPU acceleration
 - **MPAS-GPU**, **R-MPAS-GPU**
 - **CWA-GFSTco-GPU**
 - **FV3-GPU**
 - Pre/Post-processing and Verification with **GPU**

- Downscaling
 - **CorrDiff**
- AI Research
 - **Earth2MIP**
 - **Modulus**
 - **Earth2Studion**
 - ...

- 3D Rendering
 - **Paraview+Index**
- Digital Twin
 - **Omniverse**
 - ...



THANK YOU!