Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

Foundation Models

Yuewei Lin Group lead

Foundation model group
Computational science Initiative
Brookhaven National Laboratory

May 7, 2024

@BrookhavenLab

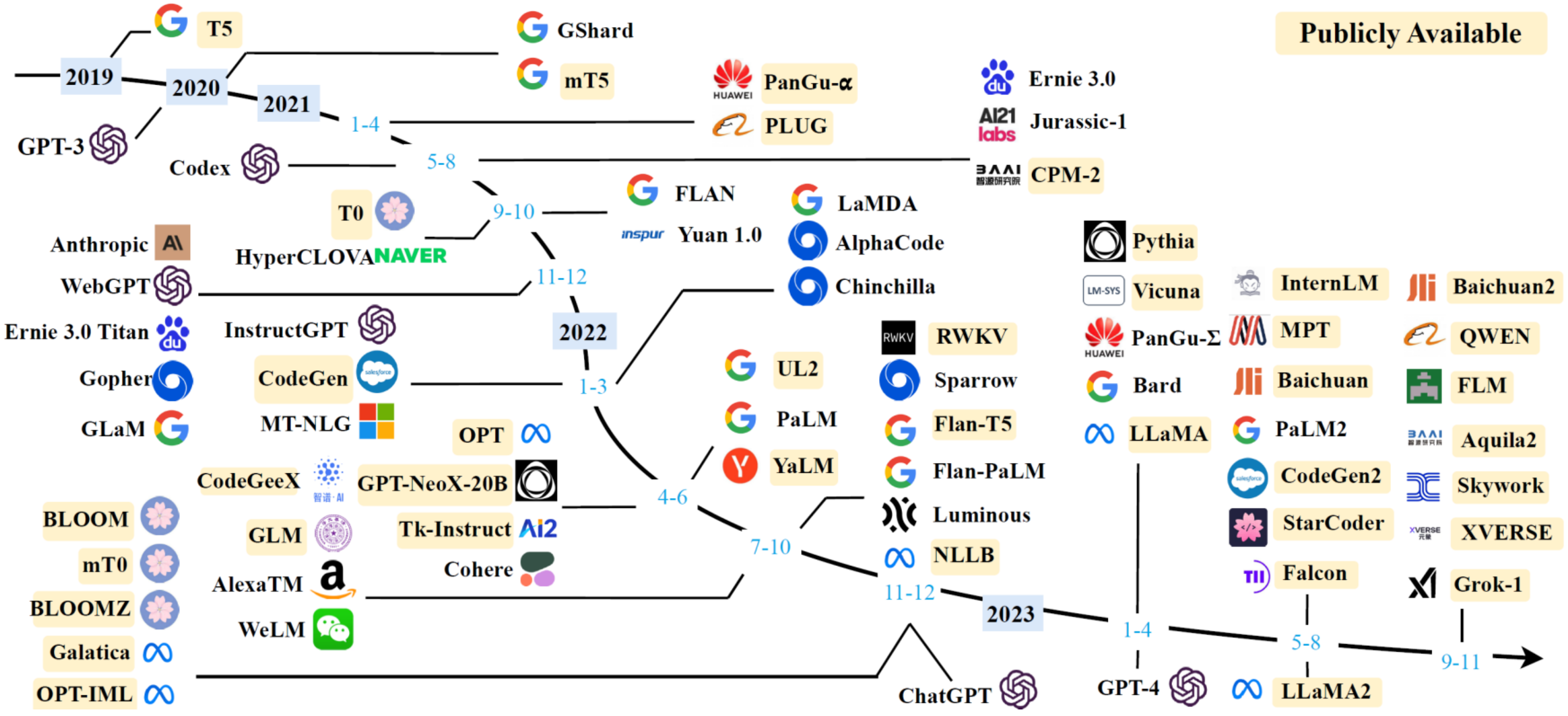
Outline

- Introduction
- Foundation models in Scientific Domains:
  - Foundation Models for Cell Segmentation
  - Foundation Models for Structure Biology
  - Foundation Models for Fusion Energy
  - Generalizability of Foundation Models
  - Multimodal Foundation Models for Plant Disease Diagnose
  - Multiple LLM Agents for Scientific Discovery
- Future Works

Introduction for Foundation Models

Foundation models are those trained on broad data (generally using self-supervision) and can be adapted to a wide range of downstream tasks.

A timeline of existing large foundation models (larger than 10B) in recent years.



Brookhaven National Laboratory

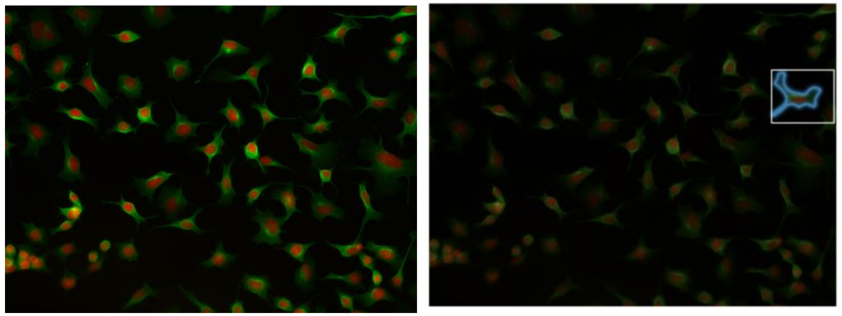
X. Zhao, et al. "A survey of large language models." *arXiv preprint arXiv:2303.18223*, 2023.

Foundation Models for Cell Segmentation

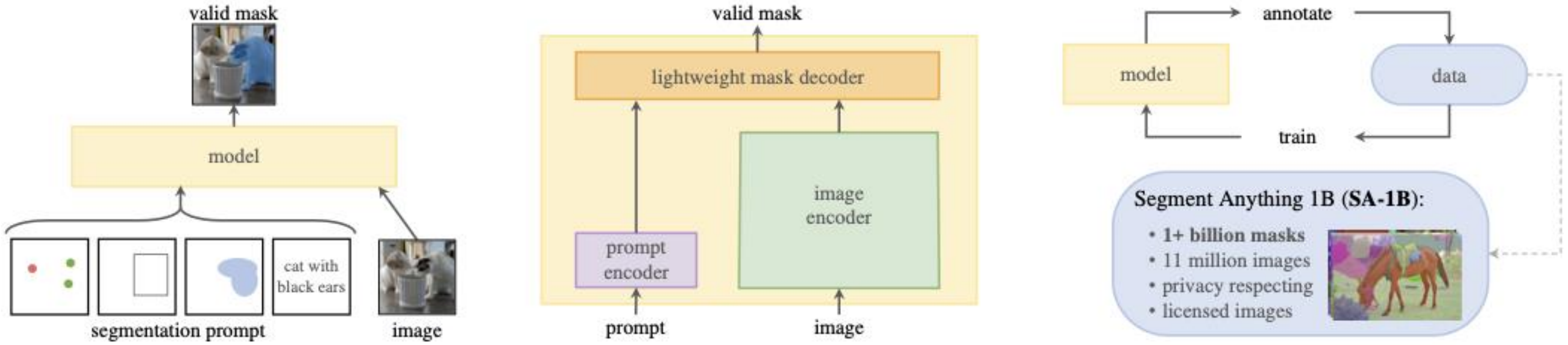
Task: 3D Cell segmentation

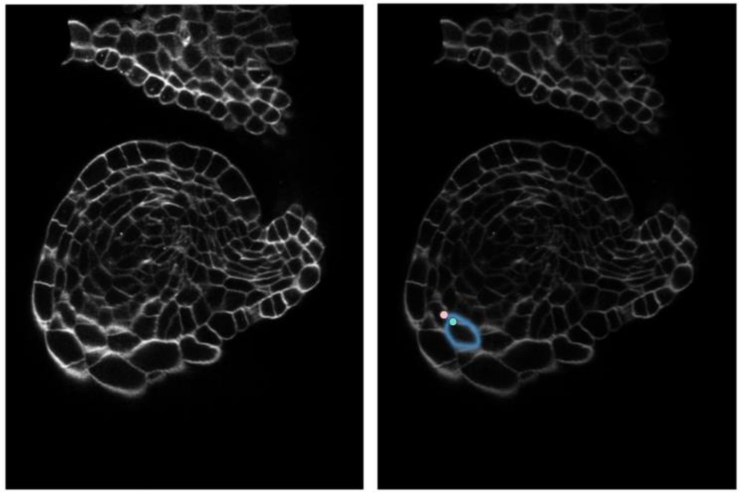
Model: Segment Anything Model (SAM)

- **400M** image-text pairs, **1B** masks from **11M** images
- Zero shot generalization ability
- Data engine: model-in-the-loop annotation



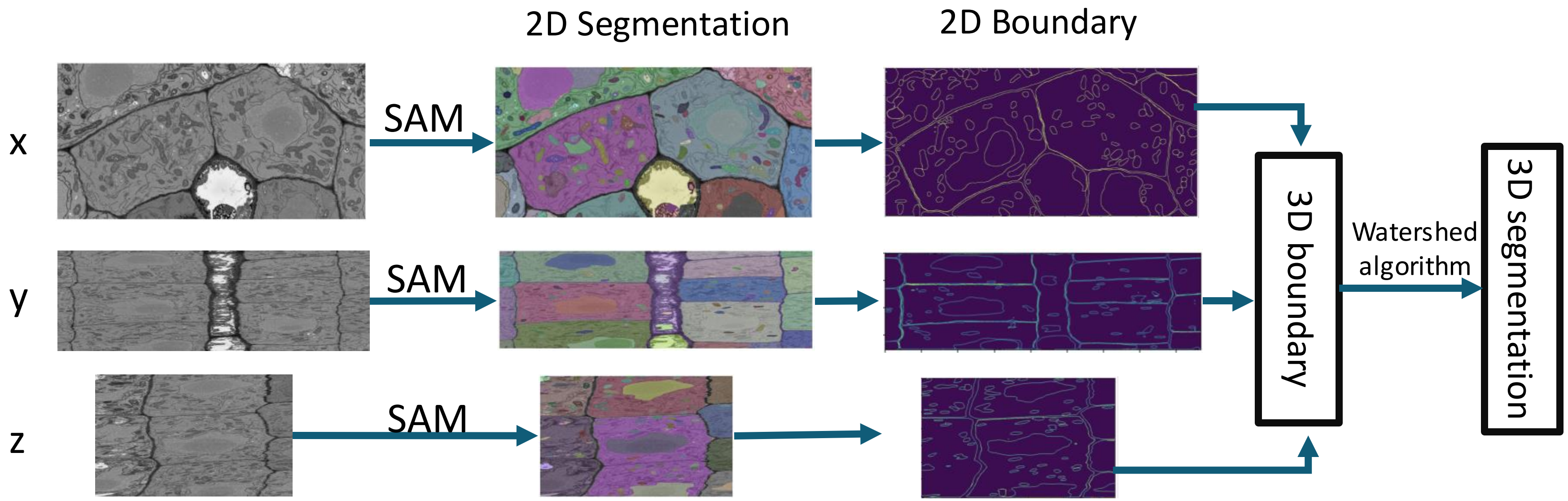
Box prompt





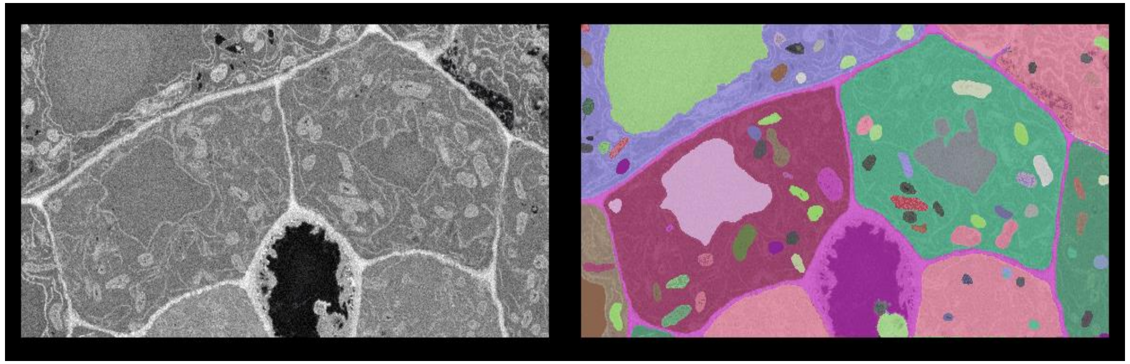
Point prompt
(positive and negative)

Foundation Models for Cell Segmentation



Advantages:

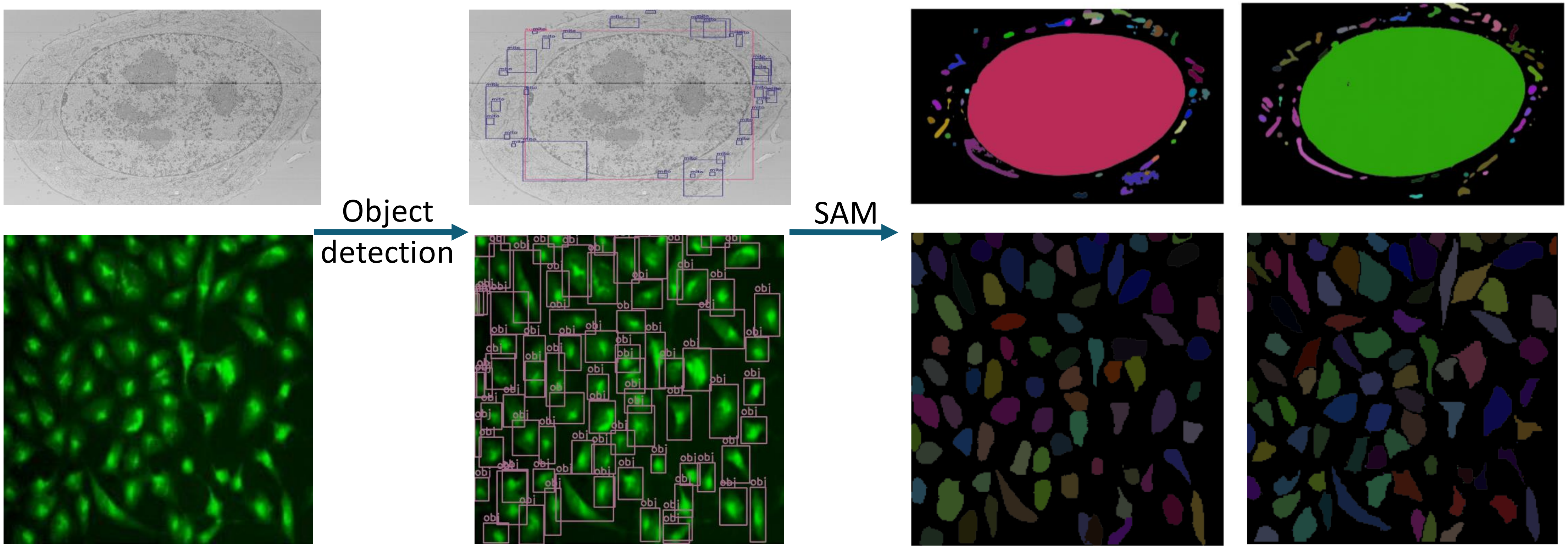
- No training needed
- No annotations needed



3D image segmentation result

Foundation Models for Cell Segmentation

- Train an object detection model to provide the bounding boxes.
- Use the bounding boxes as prompt for SAM.



Original image | Bounding box prediction | Prediction | Ground truth

Foundation Models for Structure Biology

- **Self-supervised learning** models trained on **vast amounts** of biological data
  - Protein sequence database (~1 billion of unlabeled sequences)
  - Protein structure database (~200k experimental structures)
- **Representation learning**: capture fundamental properties and patterns
  - Encode complex relationships between sequence, structure, and function
- **Efficient and scalable** prediction of biological properties and facilitate protein engineering
  - Guide rational protein design and optimization
  - Accelerate the discovery of novel proteins with desired characteristics

Foundation Models in Structure Biology

ESM (Evolutionary Scale Modeling): Protein Foundation Models

- Large protein language model with up to 15B parameters
- Trained on over 617 million diverse protein sequences
- Enhances the efficiency and accuracy of protein structure prediction
- Serves as a powerful tool for various downstream applications

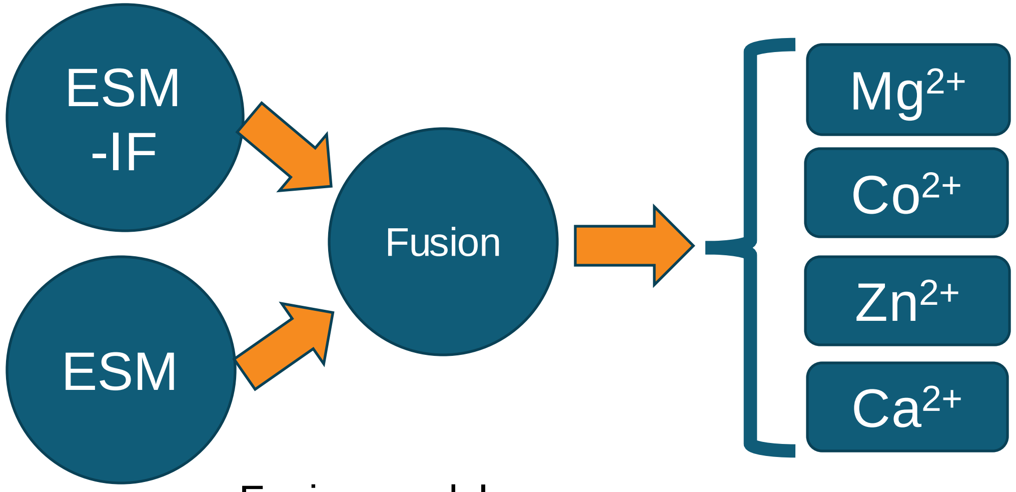
ESM-IF (Inverse Folding): Predict protein sequence based on backbone structure

- Enables protein design and optimization:
- Training data: 16K experimental structures and 12M predicted structures from AlphaFold2

Foundation Models in Structure Biology

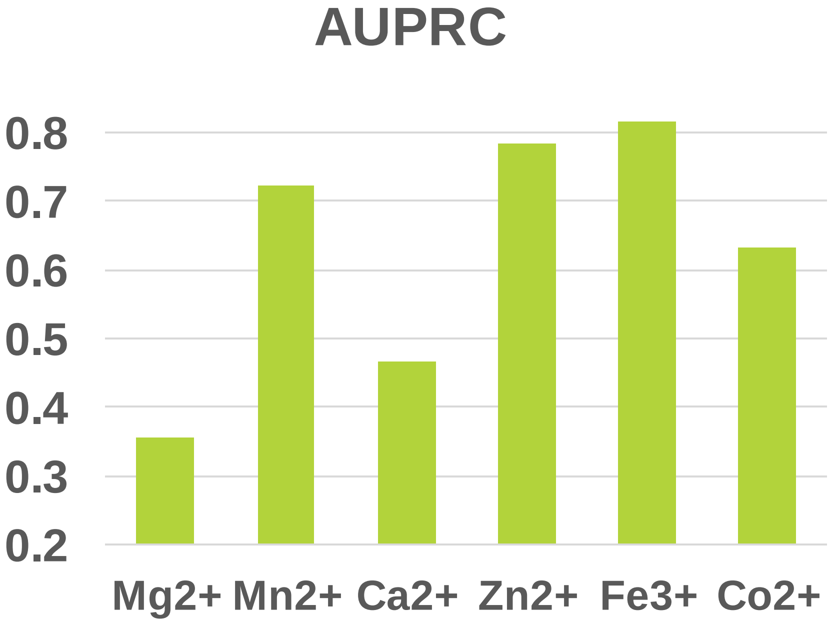
ESMBind: Protein-metal ion binding prediction

- ESM: Sequence embedding
- ESM-IF: Structure embedding
- Ion-specific models share the same upstream fused input



Fusion module: learns a unified feature from sequence and structure

ion-specific networks: binding prediction at residue level



SOTA performance on common metal ions in protein

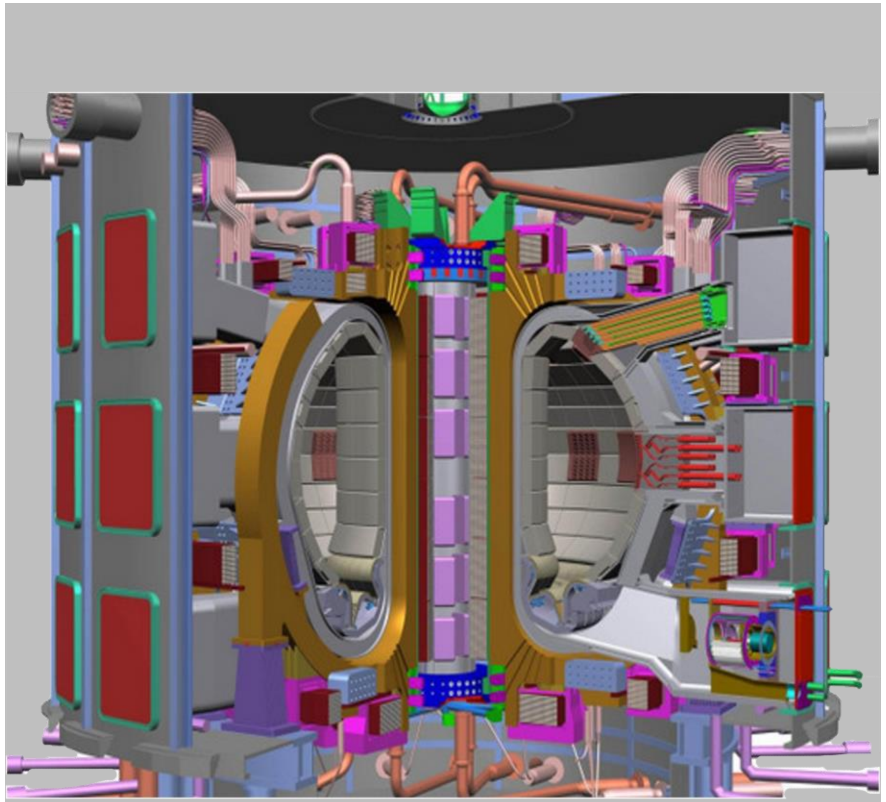
Foundation Models for Fusion Energy

Task: Predict Disruptions in Fusion Device

Challenge: Disruptions have proven very difficult to predict with classic simulation tools, especially in real time.

Goal: Develop foundation models on exascale computers to predict disruption events.

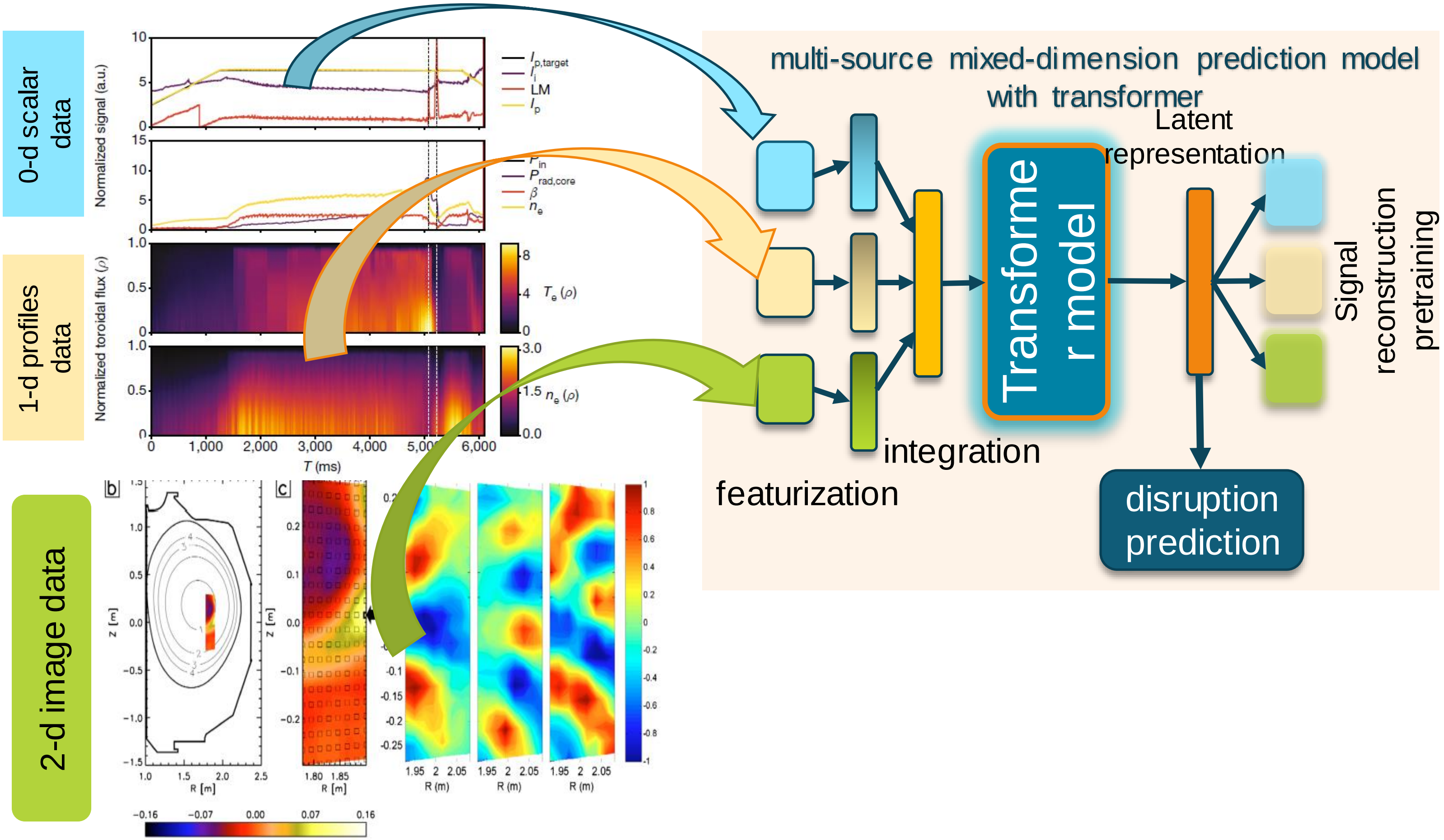
Real-time models are being developed and trained to predict the disruption and take appropriate action.



DIII-D Tokamak Fusion Reactor

Foundation Models for Fusion Energy

- Train a 1.5B parameter GPT-like foundation model using ~10 TB training data collected from the DIII-D tokamak Fusion Reactor.
- Data fusion of 2D ECEi data along with 0-d scalar and 1-d profile diagnostic data

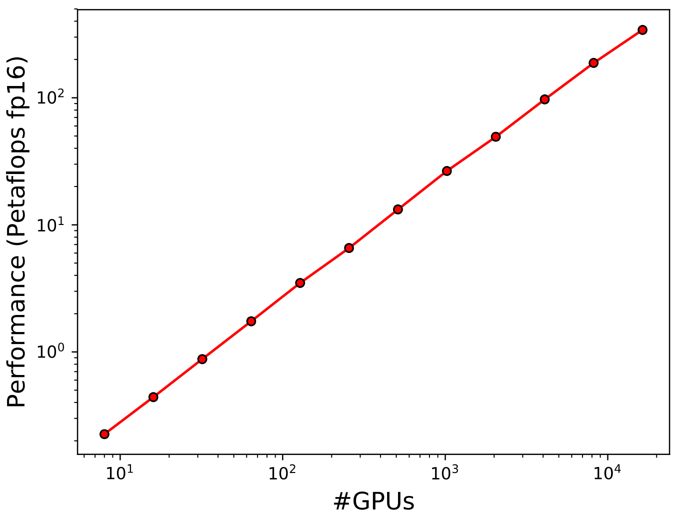


Electron Cyclotron Emission Imaging (ECEI)

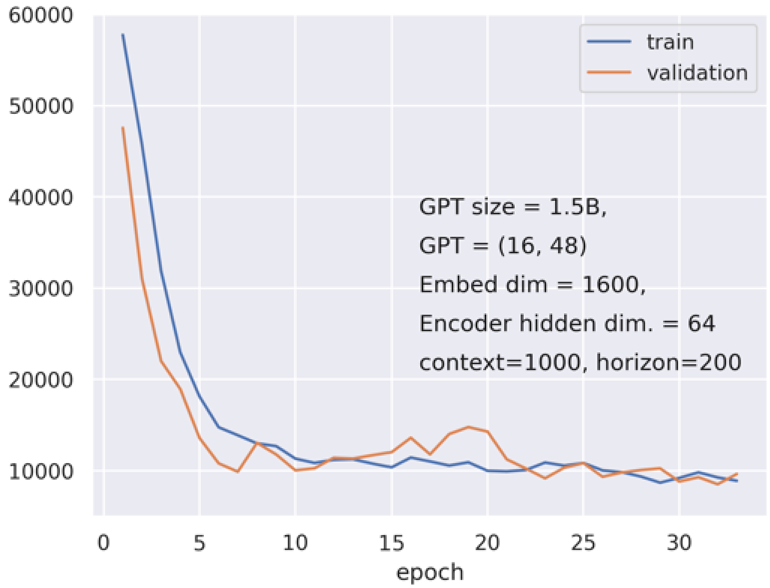
Foundation Models for Fusion Energy

Science Accomplishment and Impact

- Run the foundation model on Frontier cluster at ORNL with good scalability
- Improve predictive accuracy for disruption forecasting over previous ML tools



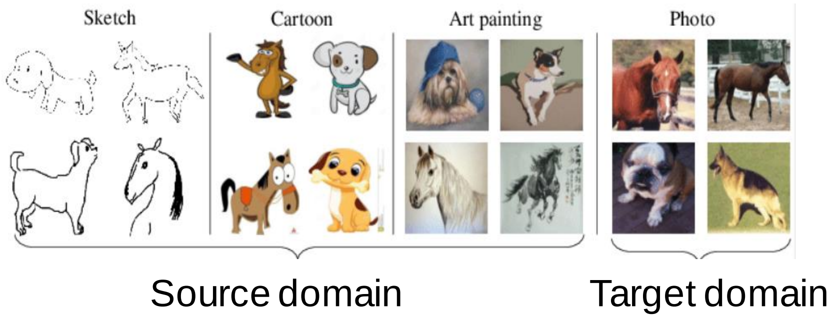
Scaling of model training on Frontier cluster at ORNL with 16,384 GPUs. Achieved 0.4 fp16 ExaFlops.



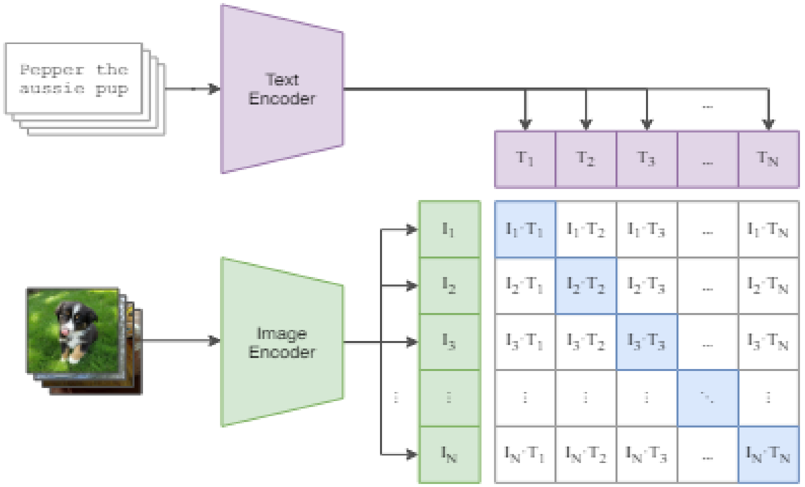
Brookhaven National Laboratory

Generalizability of Foundation Models

- **Domain Generalization (DG)** aims to learn a generalizable model trained on observed source domains, and directly applied on the target domain which is unseen during training.



- **Visual-language model (VLM)** are trained on massive datasets, like CLIP[1] model, trained on 400 million pairs of images and texts. The diverse data that help VLMs demonstrate impressive zero-shot ability.

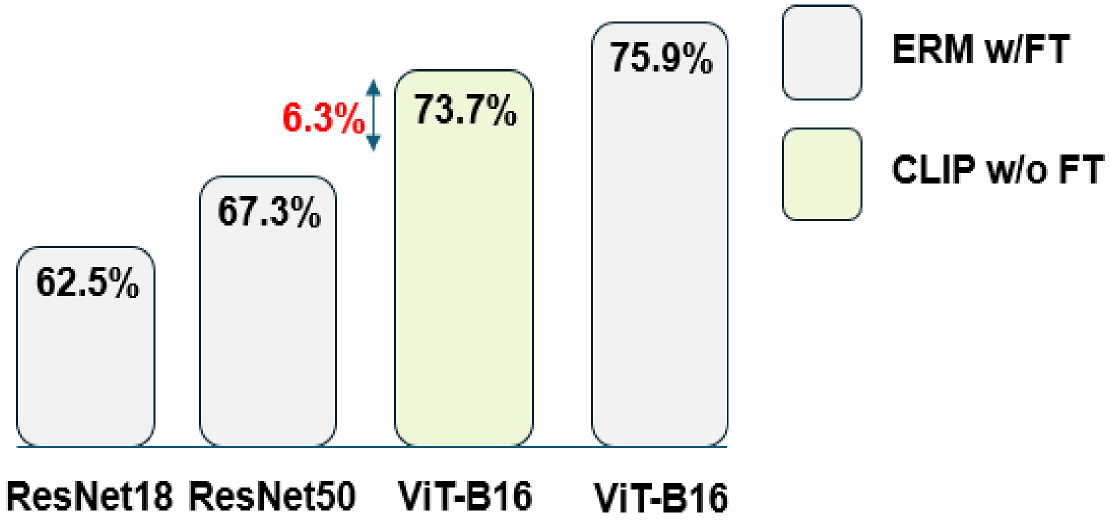


Brookhaven National Laboratory

[1] A. Radford, et al. "Learning transferable visual models from natural language supervision." In *ICML*, 2021.

Generalizability of Foundation Models

Are the large visual language models good and good enough for generalizability?

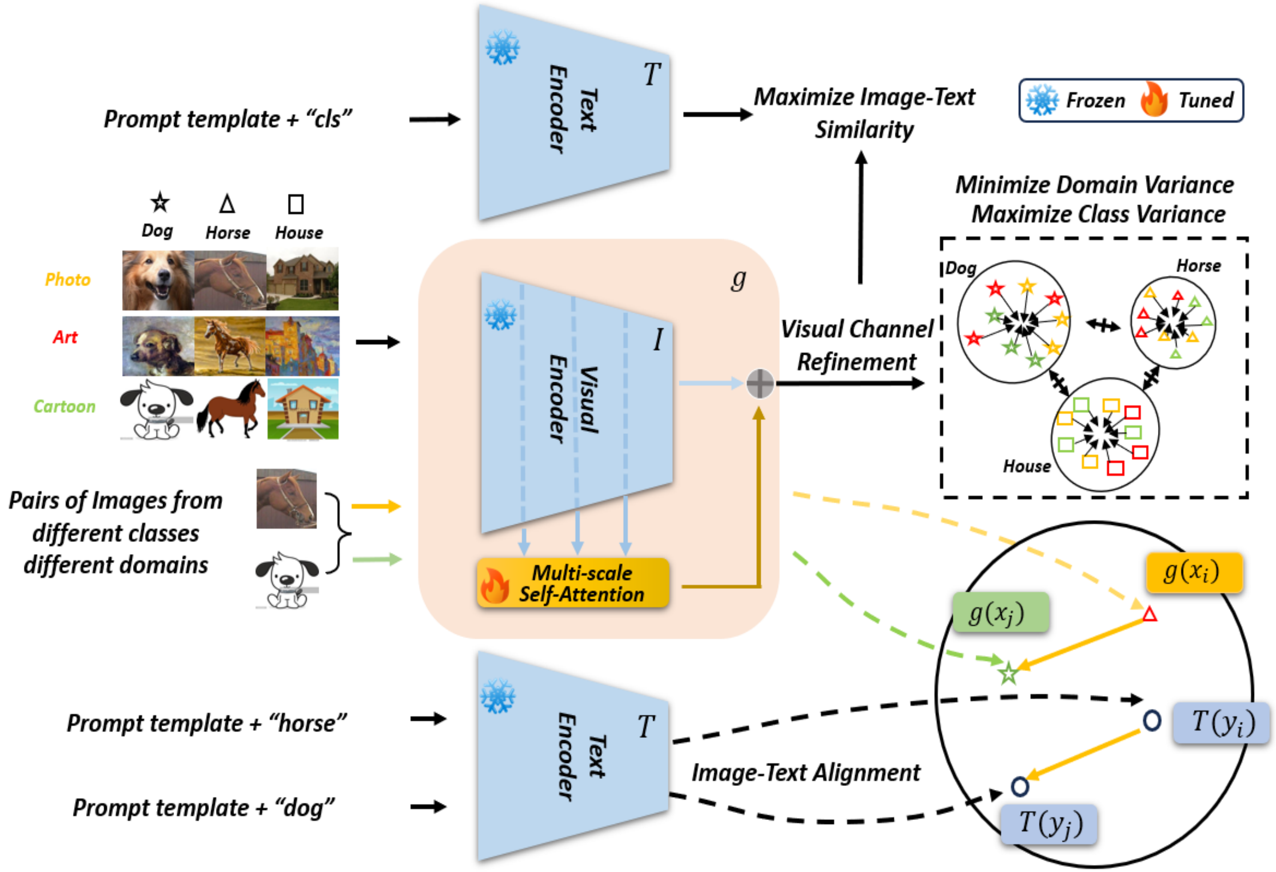


We evaluated the model's generalizability experiments. Even without fine-tuning, CLIP outperforms trained ResNet50 by 6.3%. *Although CLIP has shown impressive performance, there is still some scope for improvement.*

Brookhaven National Laboratory

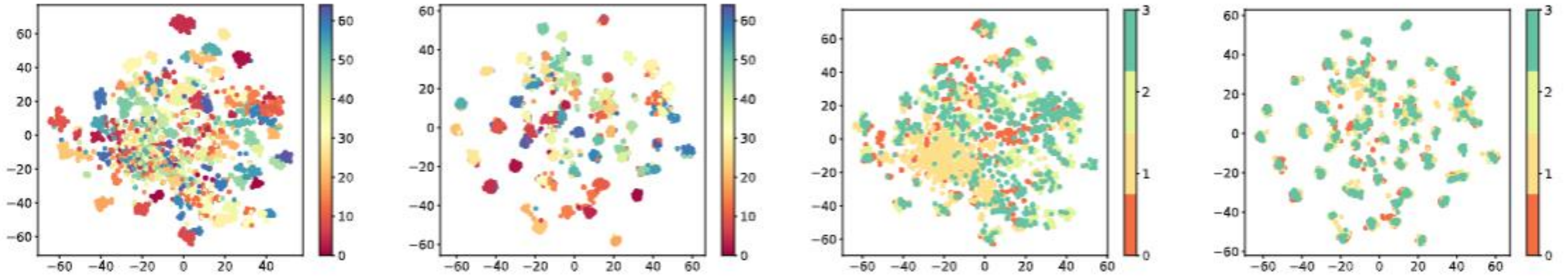
Generalizability of Foundation Models

- CLIPCEIL (Boosting Domain Generalization with **CLIP** by **C**hannel r**E**finement and **I**mage-text a**L**ignment) introduces a lightweight module to fine-tune the CLIP and employ the multi-scale visual features.
- We proposed channel refinement module to ensure each channel contains domain-invariant (minimizing domain variance) and class-relevant (maximizing class variance) information.
- To align the image and text, we maximize the image-text similarity and calculate direction loss using text class descriptions based on data pairs from different classes and domains.



CLIPCEIL

Generalizability of Foundation Models



(a) Zero-shot across classes (b) CLIPCEIL across classes (c) Zero-shot across domains (d) CLIPCEIL across domains

Model	Venue	PACS	VLCS	OfficeHome	TerraInc	DomainNet	Avg
*SAGM [48]	CVPR'23	86.6	80.0	70.1	48.8	45.0	66.1
*DomainDrop [12]	ICCV'23	89.5	78.3	71.8	-	44.4	-
CLIP Zero-Shot	-	96.2	81.7	82.4	33.4	57.5	70.2
Lin.Probing	-	96.5	82.6	80.4	50.2	57.6	73.5
ERM [44]	-	93.7	82.7	78.5	52.3	53.8	72.2
MIRO [4]	ECCV'22	95.6	82.2	82.5	**54.3**	54.0	73.7
CoOp [61]	IJCV'22	96.0	81.1	83.5	47.0	59.8	73.5
CoCoOp [60]	CVPR'22	95.7	83.1	84.3	50.4	60.0	74.7
DPL [54]	Arxiv'22	**97.3**	84.3	84.2	52.6	56.7	75.0
CLIPCEIL	Ours	**97.3**±0.1	**84.7**±0.2	**85.6**±0.1	53.8±0.2	**61.2**±0.1	**76.5**±0.1

Multimodal Foundation Models for Plant Disease Diagnose

Goal:

- Develop a computational framework for plant disease detection, surveillance, and prediction.
- Develop the concept of Digital Twins to connect physical and digital representations of plant diseases, enabling timely decision-making and scalable exploration of disease management strategies.

Challenges:

- Various modalities and scales to accurately detect and predict plant diseases.

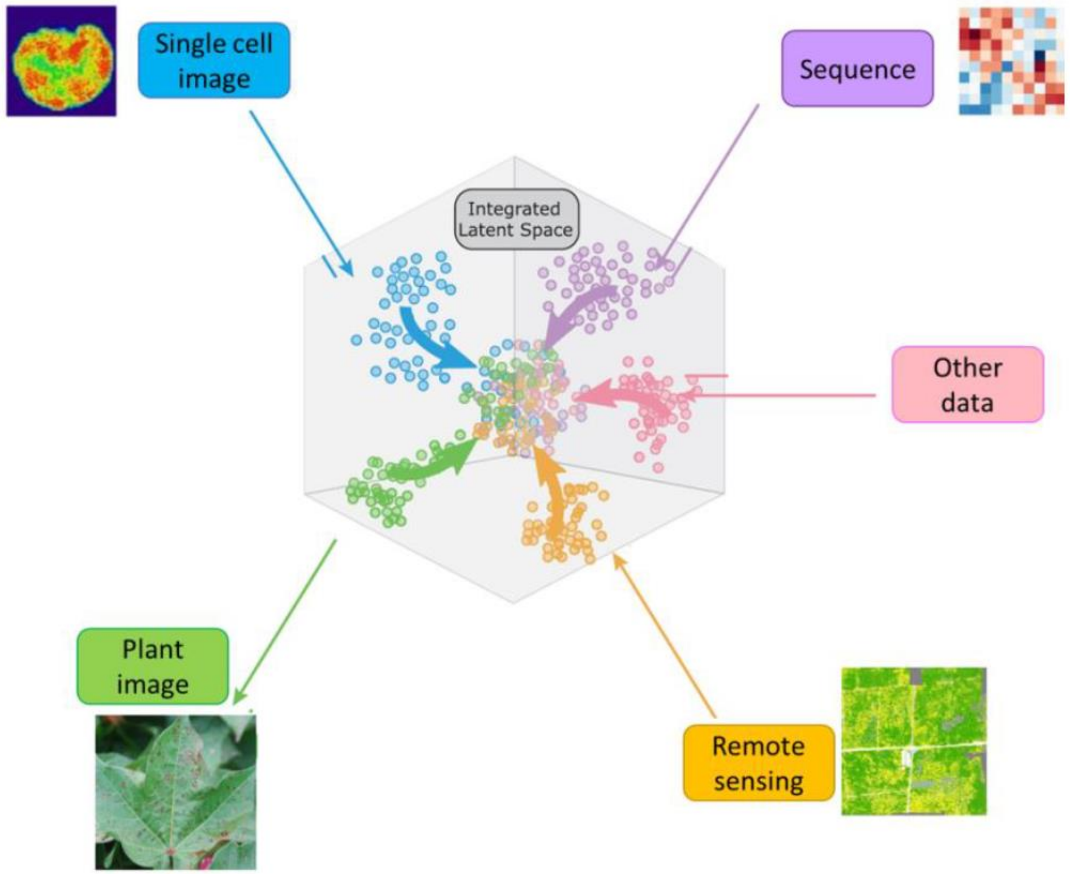
Brookhaven National Laboratory

Multimodal Foundation Models for Plant Disease Diagnose

Heterogeneity of data:

- Remote sensing images (large scale)
- Plant pictures (medium scale)
- live-cell images from Optical microscopy (small scale)
- Electronic microscopy (micro-scale)
- Genomic sequence

Multi-modal Foundation Models are utilized to align and integrate data from various modalities into a common latent space.



Multiple LLM Agents for Scientific Discovery

LLM-based agent

An LLM-based agent is a comprehensive framework consisting of planning, tools, and memory. The LLM acts as a central controller that provides planning and calls external tools to solve complex tasks.

Limitations for existing LLM-based

- Most existing tools uses a single LLM agent.
- Existing tools are not for scientific applications.
- Existing tools usually encounter reliability concerns.

Multiple LLM Agents for Scientific Discovery

LLM-based agents for scientific research

- ***Involving the planning agent and evaluation agent.*** The planning agent first makes the decisions. Then, the evaluation agent evaluates the results, provide feedback and guidance on how to refine the decisions.
- ***Integrating advanced scientific tools and evaluation tools.*** Enhancing the suitability of LLM-based agents by integrating advanced scientific tools and evaluation tools.
- ***Incorporating human expertise into the loop.*** we leverage the human verbal description into the prompt and make the agent align with human values and preferences. Human interaction can take place both during the planning and evaluation phases

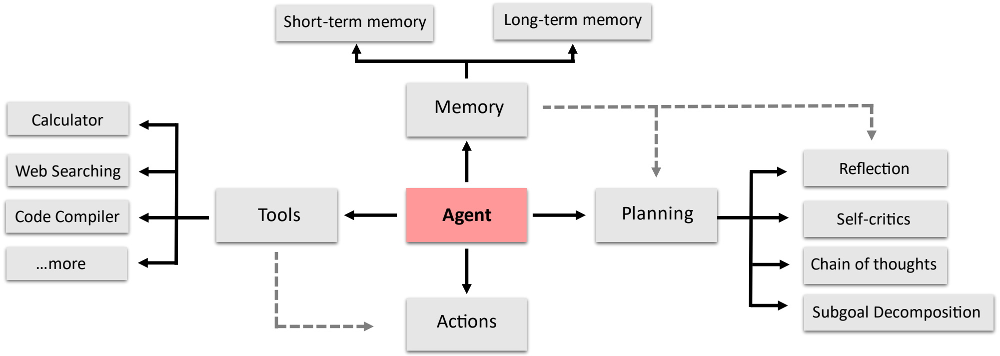


Figure 1 The overall structure of existing LLM-based agents.

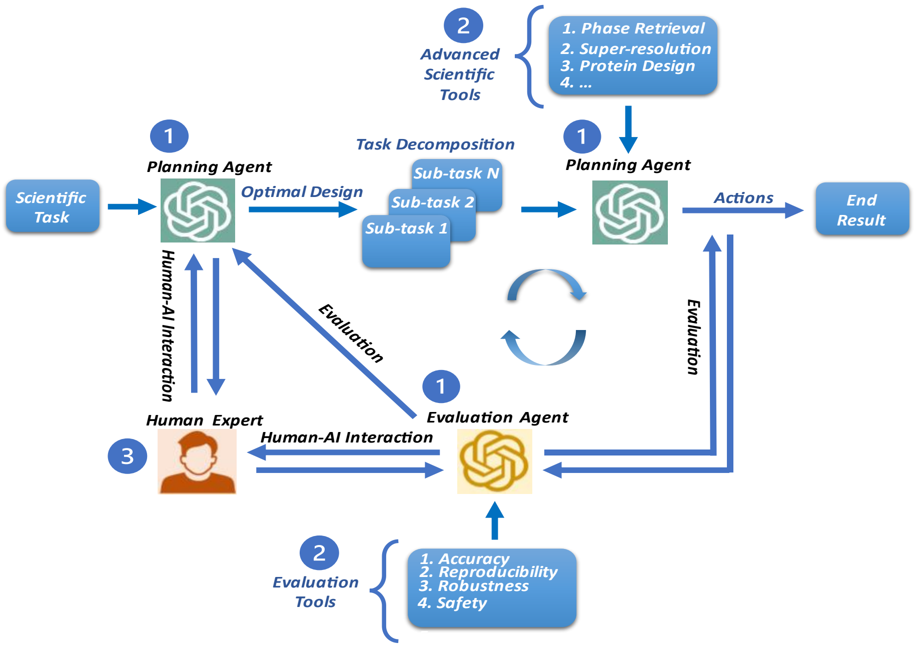


Figure 2. The overall architecture of the proposed scientific LLM-based agents

Brookhaven National Laboratory

Future Works

Other Architectures other than Transformer

- Mamba -- Selective State Spaces

Model compression

- Knowledge distillation
- Quantization

Dataset condensation: condensing large datasets into a compact set of synthetic samples

Other strategy

- Mixture of Experts (MOE)

Thank you!