

Knowledge-guided Machine Learning:

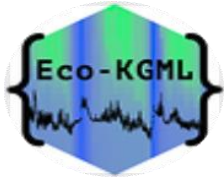
A Paradigm Shift in Advancing AI
by Incorporating Scientific Knowledge

Anuj Karpatne

Professor, Electrical and Computer Engineering
Assistant Director, AI & Informatics Research Institute

University of Florida
akarpatne@ufl.edu

<https://anujkarpatne.github.io/>



UF UNIVERSITY of FLORIDA

UF Artificial Intelligence
& Informatics Research Institute
UNIVERSITY of FLORIDA



BATTELLE UCI



Rensselaer

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL



Cornell



Seattle Children's
HOSPITAL · RESEARCH · FOUNDATION



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

neon
National Ecological Observatory Network

USGS
science for a changing world



OAK
RIDGE
National Laboratory



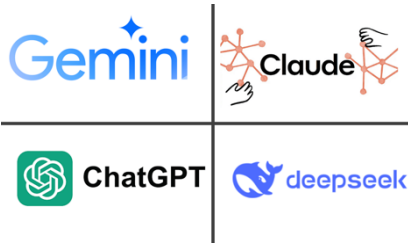
NAIRR Pilot

\$10M SAS Grant from USDA
\$18M NSF PIPP Grant # 2200045
\$15M NSF HDR Grant # 2118240
\$1.1M NSF MSB Grant # 2213550
\$1M NSF Medium Grant # 2107332
\$600K NSF CAREER Award # 2239328
NSF NAIRR Pilot # 240161

Growing Priority in AI for Science

Innovations
in AI

Opportunities in
Science & Engineering



Representation
Learning

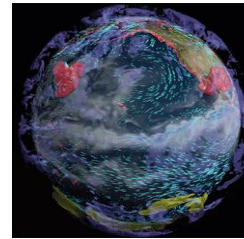
Foundation
Models

Generative AI

Agentic AI

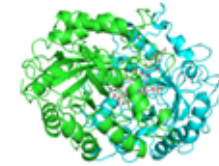
- Trained on Internet-scale data
- Driven by problems in **vision** and **language**
- **“Data-only”** (rely mostly on patterns in data)

Modeling Climate
& Weather



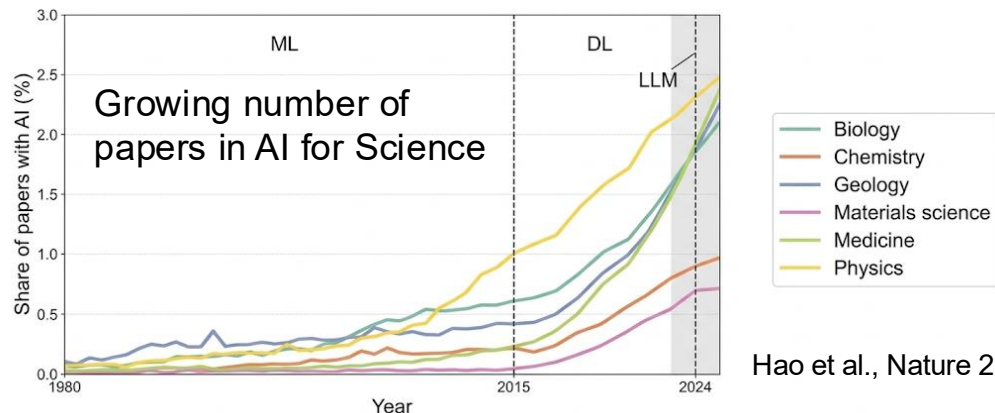
Securing Energy

Drug
Discovery



Autonomous Laboratories, ...

Modeling Astrophysics
& Cosmology



Genesis Mission

Winning the Race
AMERICA'S
AI ACTION PLAN

SCIENCE
A New Golden Age

Growing Priority in AI for Science

Innovations
in AI

Opportunities in
Science & Engineering



Representation
Learning

Foundation
Models

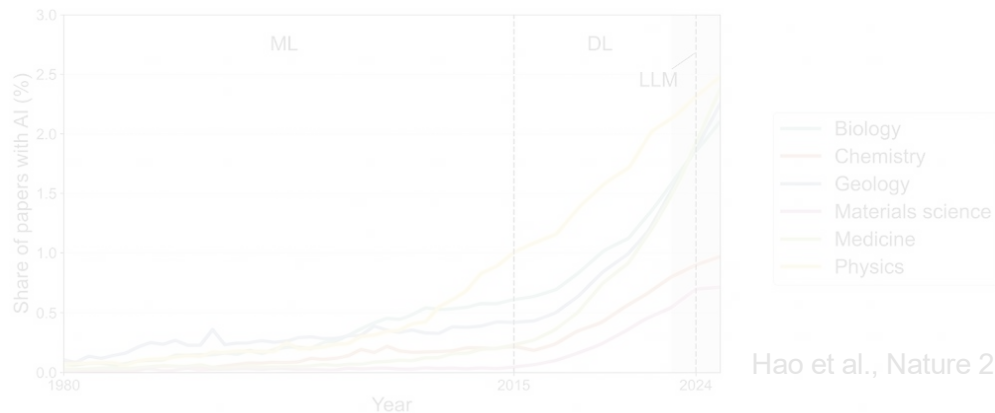
Modeling Climate
& Weather

Designing
Materials

Drug Discovery

Can the current paradigm of **scaling** “data-only”
AI models lead us to a *scientific revolution*?

- Trained on Internet
- Driven by problems in **vision** and **language**
- “**Data-only**” (rely mostly on patterns in data)



Genesis Mission

A National Mission to Accelerate Science Through
Artificial Intelligence

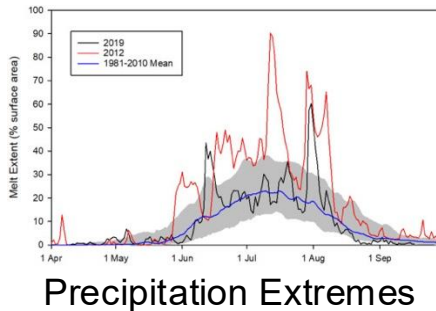
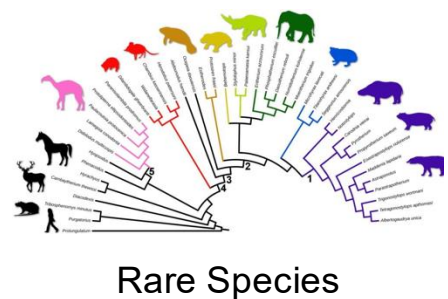
\$320M awarded (Dec 2025)

\$293M announced (March 2026)

Gaps in “Data-only” AI for Science

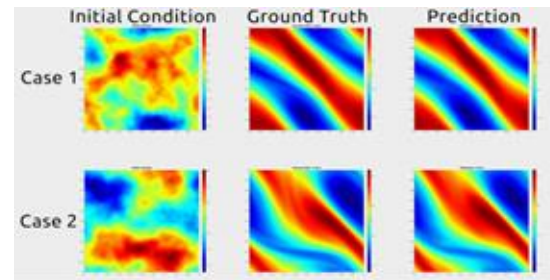
- Poor out-of-distribution (OOD) generalizability w/limited data
 - Data-driven models are designed to *interpolate* between samples
 - Struggle with **extrapolating** to new scenarios (central focus in science)

Long-tailed distributions

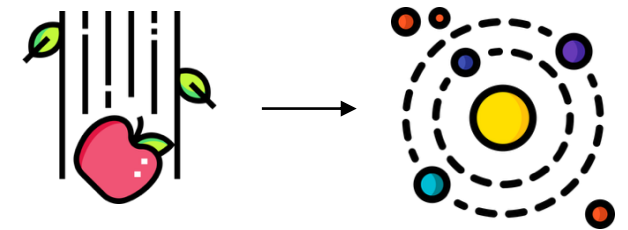


Extrapolation in physical parameters

Varying Reynolds number, Young's modulus, ...



Applying theories in new contexts



Welling, “Do we still need models or just more data and compute?” 2019

“Accelerating scientific discovery with the common task framework,” Kutz et al., 2025

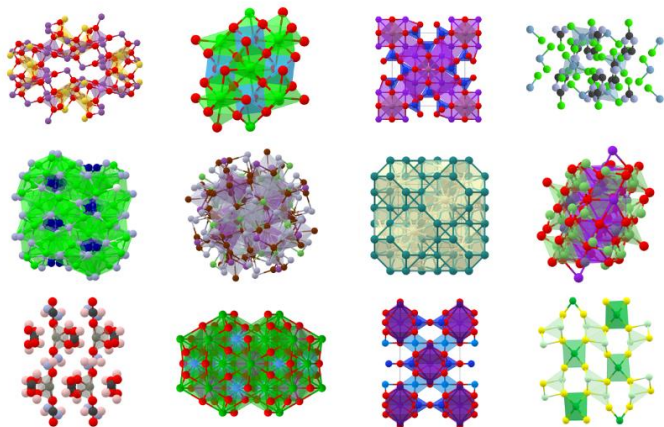
Gaps in “Data-only” AI for Science

- Poor out-of-distribution (OOD) generalizability
 - Data-driven models are designed to *interpolate* between samples
 - Struggle with **extrapolating** to new scenarios (central focus in science)
- Tendency to hallucinate (of the undesirable kind)
 - Generated samples may not be novel, useful, or consistent with scientific theories



GNoME 2023

Millions of new
materials discovered
with deep learning



Cheetham et al. 2024

Galactica: A Large Language Model for Science



**Why Meta’s latest large language
model survived only three days online**

MIT Technology Review 2022

Gaps in “Data-only” AI for Science

- Poor out-of-distribution (OOD) generalizability
 - Data-driven models are designed to *interpolate* between samples
 - Struggle with **extrapolating** to new scenarios (central focus in science)
- Tendency to hallucinate (of the undesirable kind)
 - Generated samples may not be novel, useful, or consistent with scientific theories
- Lack of mechanistic interpretability
 - Critical for reliability and discovery of new knowledge
- Limited control with scientific queries/needs
 - Hypothesis testing, scenario analysis
- High Compute and Data Needs, ...



Where is the *dorsal fin* of this fish and how does it change with environment?

Gaps in “Data-only” AI for Science

- Poor out-of-distribution (OOD) generalizability
 - Data-driven models are designed to *interpolate* between samples
 - Struggle with **extrapolating** to new scenarios (central focus in science)
- Tendency to hallucinate (of the undesirable kind)
 - Generated hypotheses are not grounded in scientific theories
- Lack of integration with scientific theories
 - Can we overcome these challenges in “data-only” AI by taking *guidance* from **scientific theories and principles**?
 - Critical for reliability and discovery of new knowledge
- Limited control with scientific queries/needs
 - Hypothesis testing, scenario analysis
- High Compute and Data Needs, ...

Opportunities in Science to Transform AI

- Rich body of *explainable* concepts and theories for grounding AI mechanisms
 - Governing equations, mechanistic models, symmetries, taxonomies, ontologies, ...
- New opportunities to evaluate AI performance
 - OOD generalizability, robustness, hallucinations, ...
- Broader class of problem formulations
 - Forward and inverse modeling, multi-scale modeling, hypothesis generation, ...

Series of NSF-sponsored workshops on
“AI-enabled Scientific Revolution” (2023 to 2025)

npj | artificial intelligence

Perspective | [Open access](#) | Published: 11 August 2025

AI-enabled scientific revolution in the age of generative AI: second NSF workshop report

[Anuj Karpatne](#), [Aryan Deshwal](#), [Xiaowei Jia](#), [Wei Ding](#), [Michael Steinbach](#), [Aidong Zhang](#) & [Vipin Kumar](#)

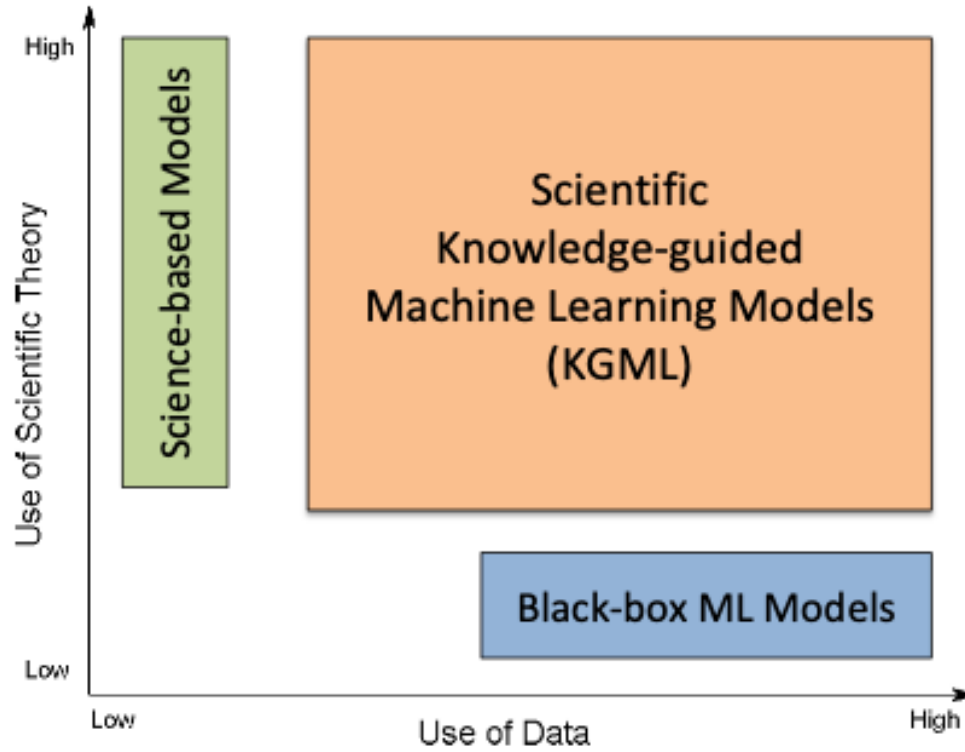
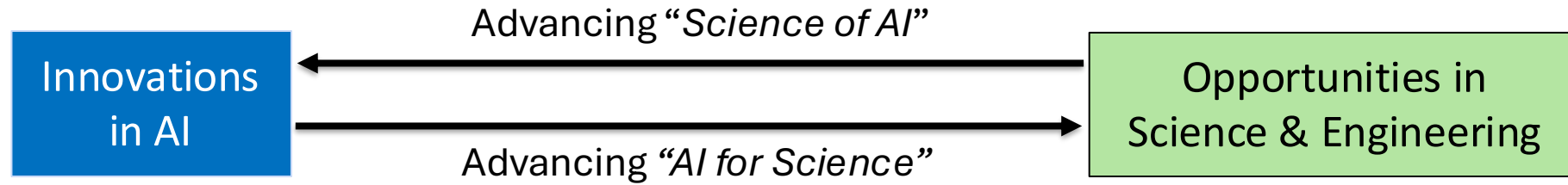
“Advancing the
Science of AI”

AMERICA'S
AI ACTION PLAN

“AI systems need to understand the
physical world, not just language”



Scientific Knowledge-guided Machine Learning



Work in this area also referred to as:

- **Knowledge/Theory/Science/Physics-**
- **Guided/Informed/Aware/Infused...**
- **ML/AI/NN/Data Science...**

KGML has the potential to

- Improve **OOD generalizability**
- Offer **Parameter and Label Efficiency**
- Produce **Interpretable** and **Scientifically Meaningful** AI solutions, leading to generation of novel **Scientific Hypotheses**

Karpatne et al. "Theory-guided data science: A new paradigm for scientific discovery," TKDE 2017

KGML is a Rapidly Growing Community

- Workshops, symposiums, tutorials
 - KGML Bridge at AAAI (2024-26); KGML AAAI Fall Symposium (2020-22); KGML Workshops at UMN (2020,21,24), ICML 2023, ECML-PKDD 2023; KGML Tutorial at KDD (2021), ORNL (2024)

U-M Knowledge-Guided Machine Learning (KGML) Workshop:
Leading the New Paradigm of AI for Science



- First comprehensive book on KGML
 - <https://sites.google.com/vt.edu/kgml-book/>



PyKGML: Python library for knowledge-guided machine learning

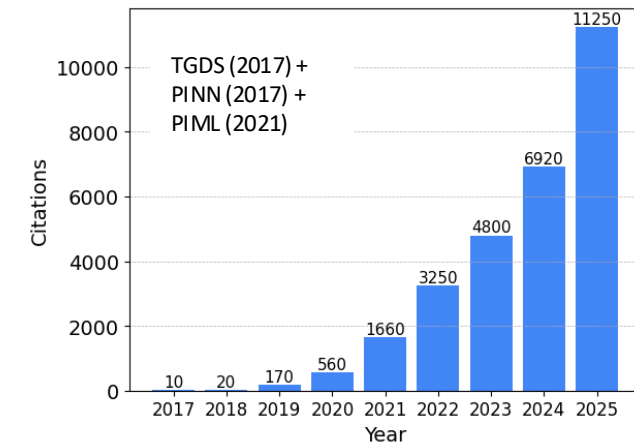
DOE EFRC NOFO, Feb 2026

"AI/ML approaches must be explicitly informed by physics and chemistry. Purely black-box or simulation-only approaches risk overlooking what truly governs material behavior."

- Large Scale Centers/Institutes
 - NSF AI-LEAF Institute (UMN)
 - NSF HDR Imageomics Institute (Ohio State)
 - NSF Eco-KGML Group (UW-Madison)
 - USDA-SAS Project (Ohio State)

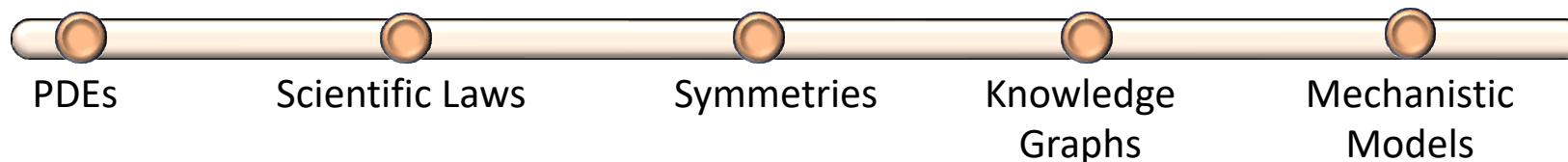
 **Genesis Mission RFA Quotes**
"physics-informed AI for flow modeling",
"physics-constrained AI-enabled models",
"physics-aware AI frameworks",
"physics-informed AI methods",
"hybrid AI-physical models",
"incorporating mechanistic knowledge",

March 17, 2026



Organizing KGML Research: A Multi-Dimensional View

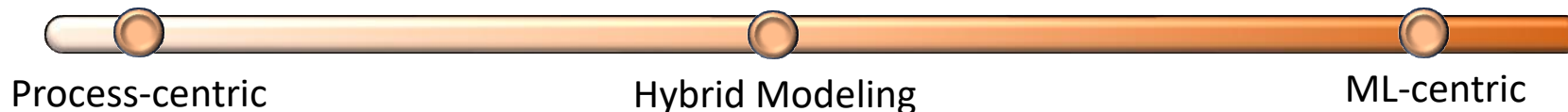
Format Used for Representing Knowledge



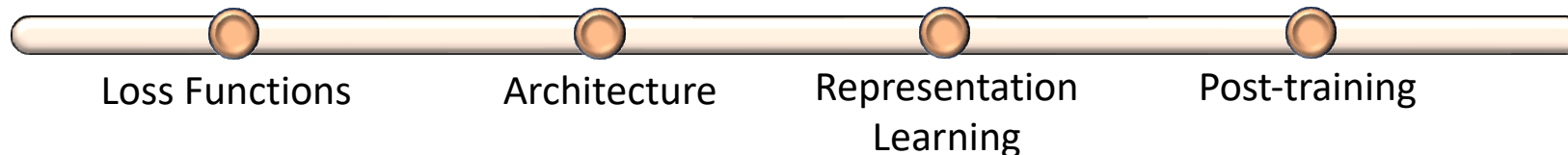
Type of Scientific Knowledge



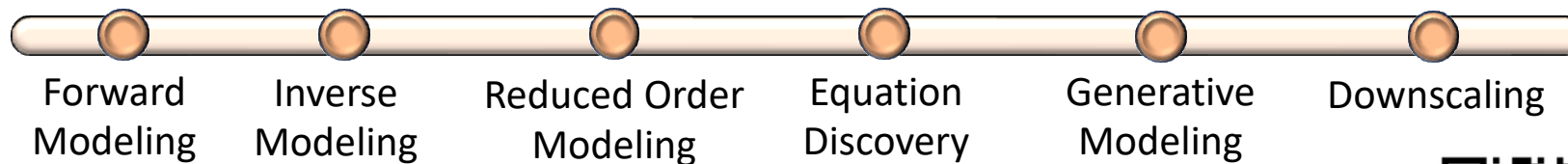
Form of Knowledge-ML Integration



Method for Incorporating Scientific Knowledge in ML

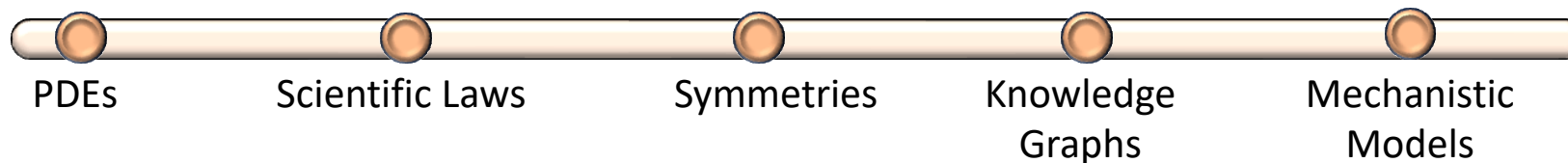


Scientific Use-cases of KGML



Organizing KGML Research: A Multi-Dimensional View

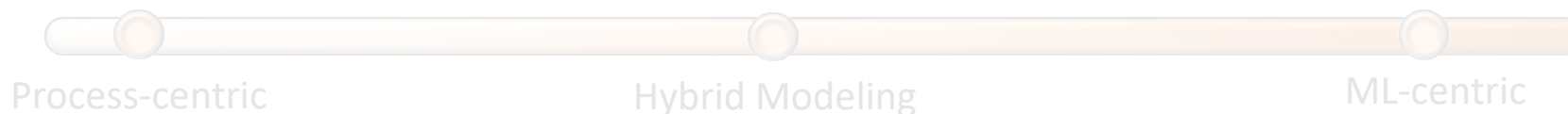
**Format Used for
Representing Knowledge**



**Type of Scientific
Knowledge**



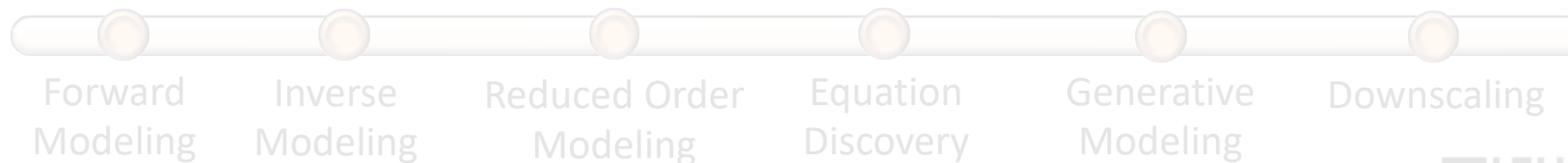
**Form of Knowledge-ML
Integration**



**Method for
Incorporating Scientific
Knowledge in ML**



**Scientific Use-cases of
KGML**



Organizing KGML Research: A Multi-Dimensional View

Format Used for Representing Knowledge



PDEs

Scientific Laws

Symmetries

Knowledge
Graphs

Mechanistic
Models

$$\frac{\partial u(x, t)}{\partial t} + \mathcal{N}(\lambda, u(x, t)) = 0$$

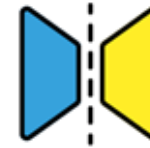
Navier Stokes Equation,
Wave Equation,
Schrodinger Equation, ...

PINNs: Raissi et al. 2019
DeepONets: Lu et al. 2021
Fourier Neural Operators
(FNOs): Li et al. 2021

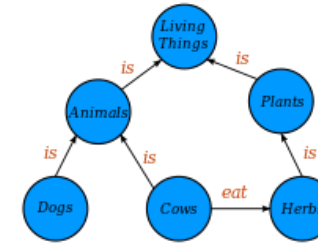
$$\frac{\partial u}{\partial x} \propto \frac{\partial^2 u}{\partial x^2}$$

$$a < \frac{\partial u}{\partial x} < b$$

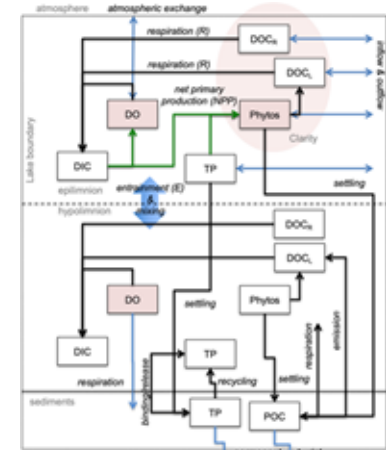
PGNN: Karpatne et al. 2017
PGRNN: Jia et al. 2019
PGA-LSTM: Daw et al. 2020



NequIP: Batzner et al. 2022
Cormorant: Anderson et al. 2019
Equivariant-Net: Wang et al. 2021



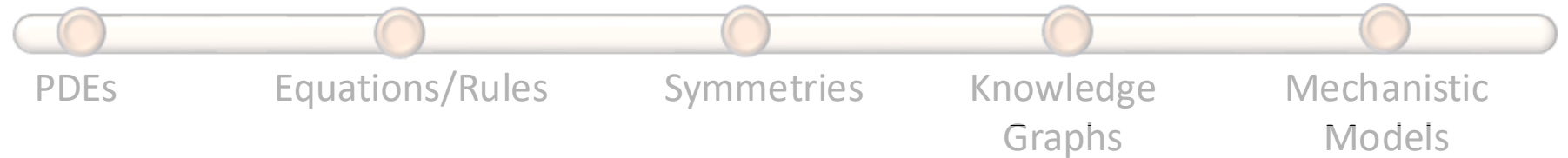
Zareian et al. 2020



MCL: Ladwig et al. 2024
dPL: Shen et al. 2023

Organizing KGML Research: A Multi-Dimensional View

Format Used for
Representing Knowledge



Type of Scientific
Knowledge



Example: Solving *known* PDEs

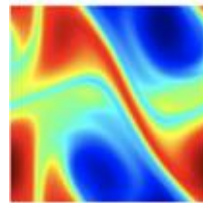
Example: Modeling complex dynamical systems with missing/imperfect physics

Navier Stokes' Equation

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla - \nu \nabla^2 \right) \mathbf{u} = -\frac{1}{\rho} \nabla p + \mathbf{g}.$$

Solve

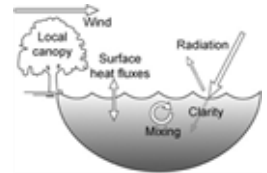
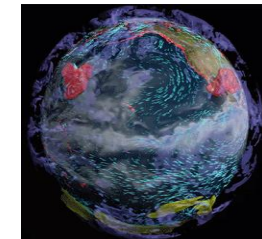
Pressure / Velocity Fields



Primary Objective: Improve Computational Efficiency

PINNs: Raissi et al. 2019, **DeepONets**: Lu et al. 2021, **FNOs**: Li et al. 2021

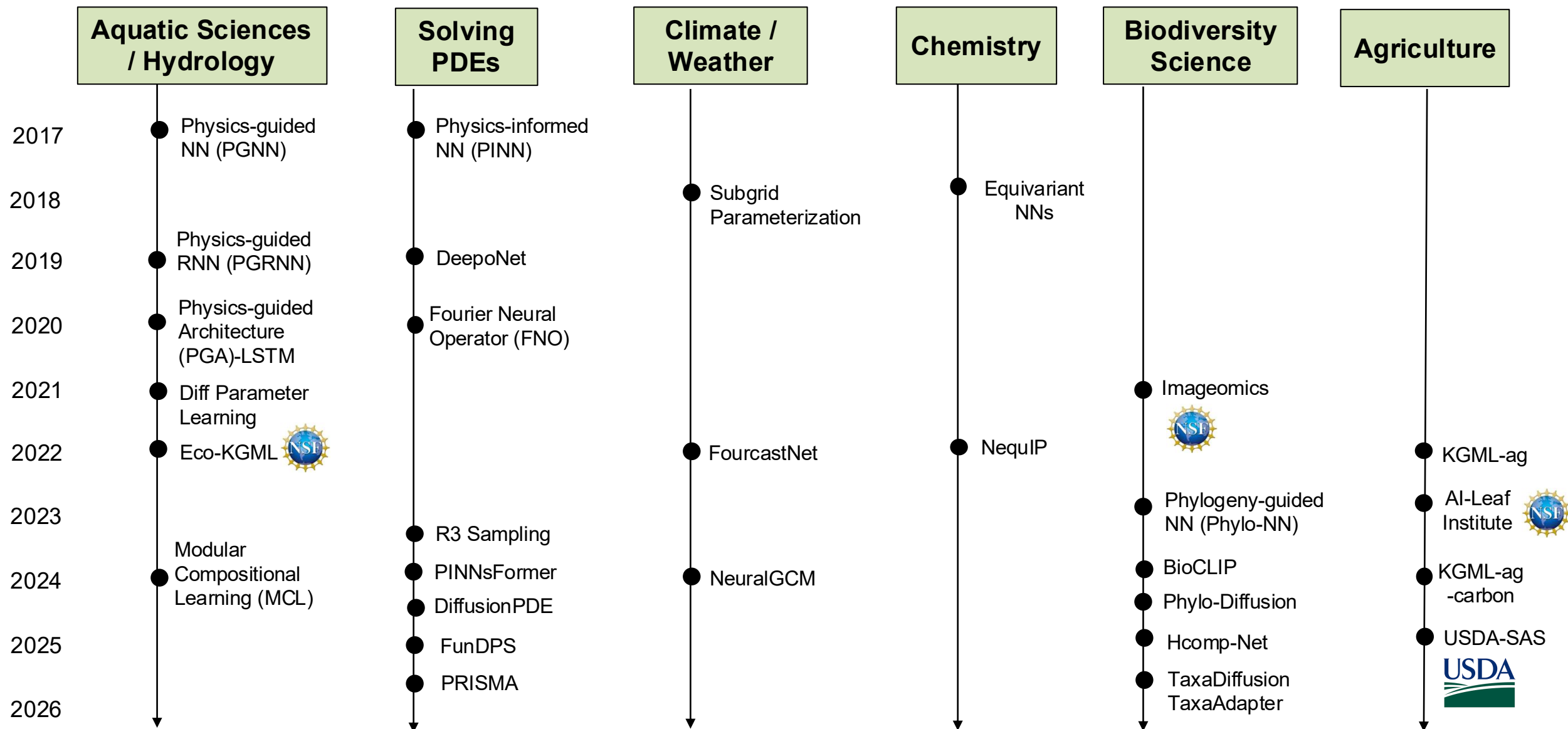
Modeling Turbulence,
Multi-phase Flow,
Cloud Physics, Aerosols, ...



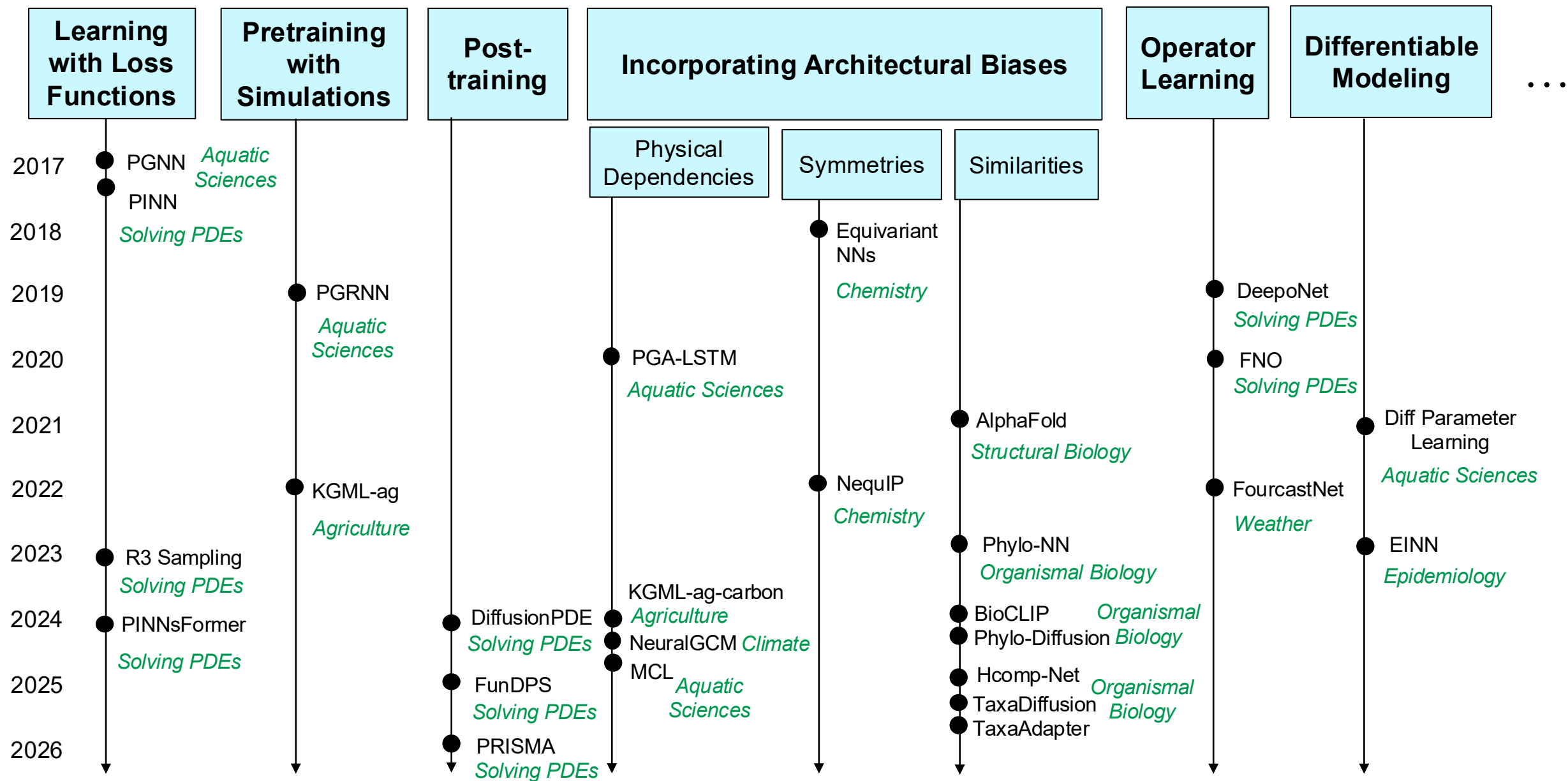
Additional Objective: Improve Modeling Accuracy

PGNN: Karpatne et al. 2017, **PGRNN**: Jia et al. 2019, **PGA-LSTM**: Daw et al. 2020

How is KGML Impacting Sciences?

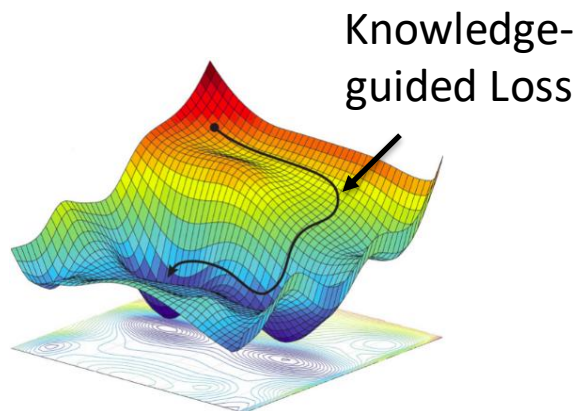
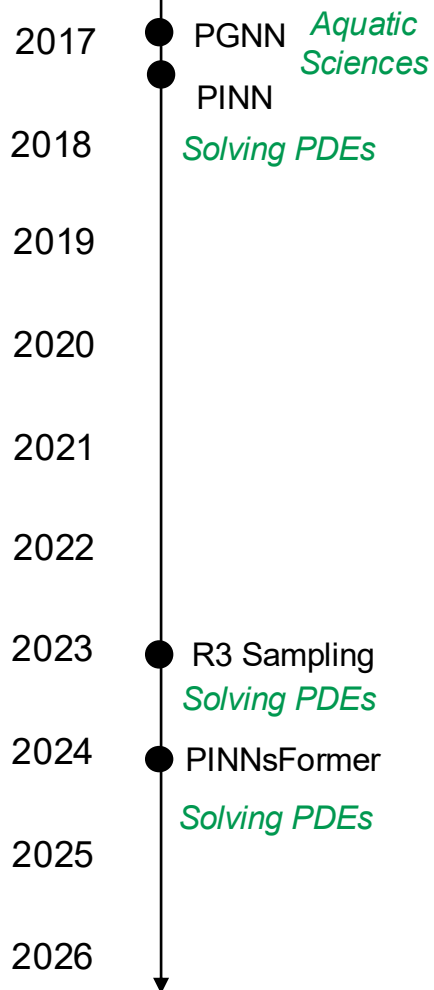


How is KGML Advancing the Science of AI?



How is KGML Advancing the Science of AI?

Learning with Loss Functions



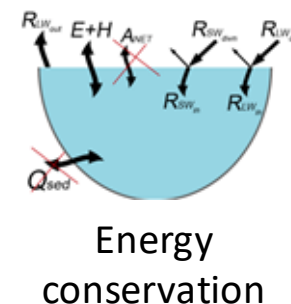
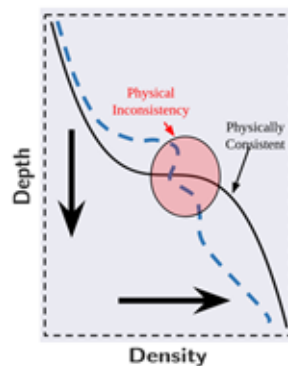
$$\text{Prediction Loss } (Y_{true}, Y_{pred}) + \lambda_{KG} \text{ Knowledge-guided Loss } (Y_{pred})$$

Solving PDEs

$$\frac{\partial u(x, t)}{\partial t} + \mathcal{N}(\lambda, u(x, t)) = 0$$

PINN

Incorporating Imperfect / Partial Knowledge



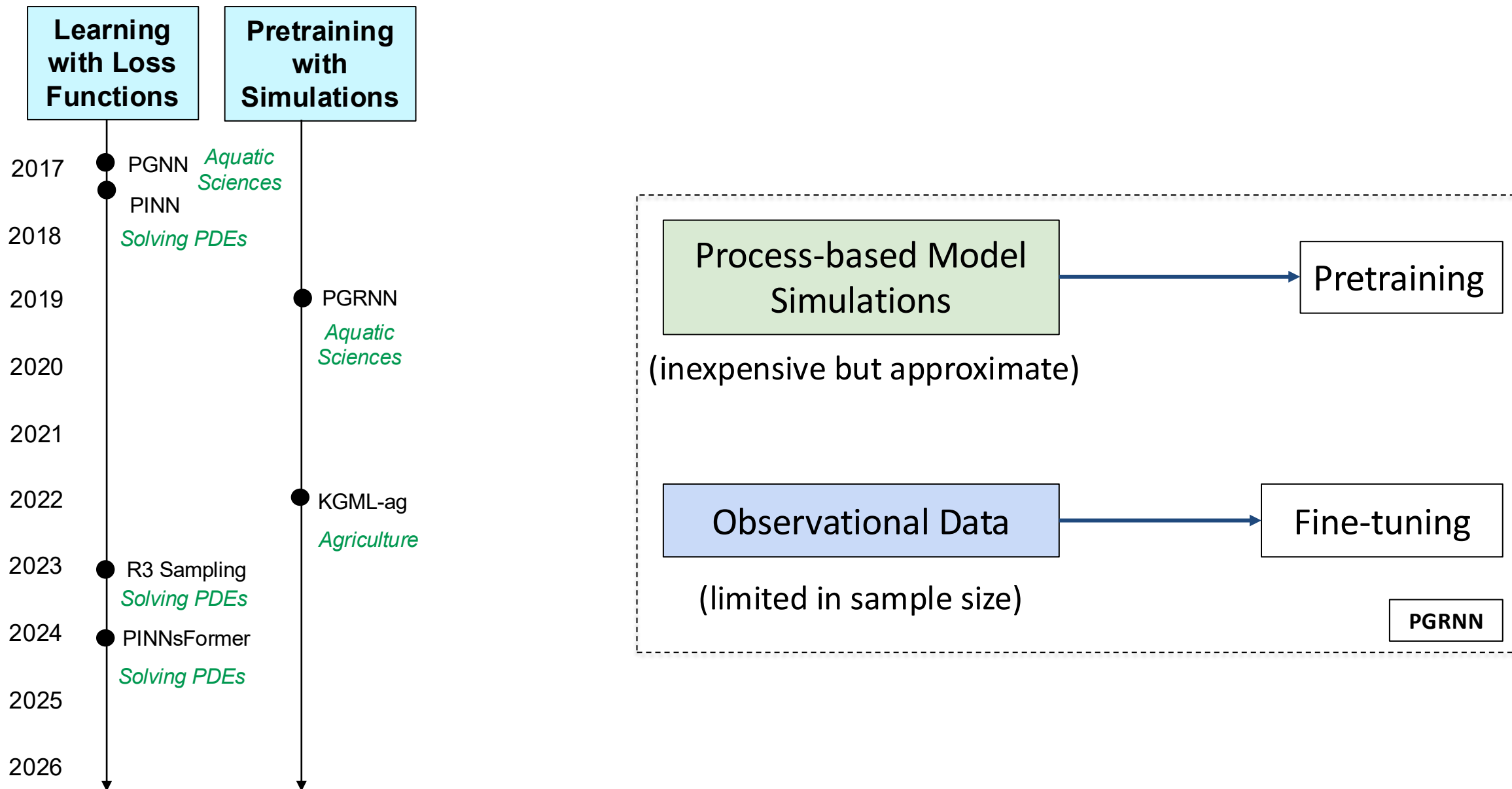
Linear Dependencies,
Inequalities, ...

$$\frac{\partial u}{\partial x} \propto \frac{\partial^2 u}{\partial x^2} \quad a < \frac{\partial u}{\partial x} < b$$

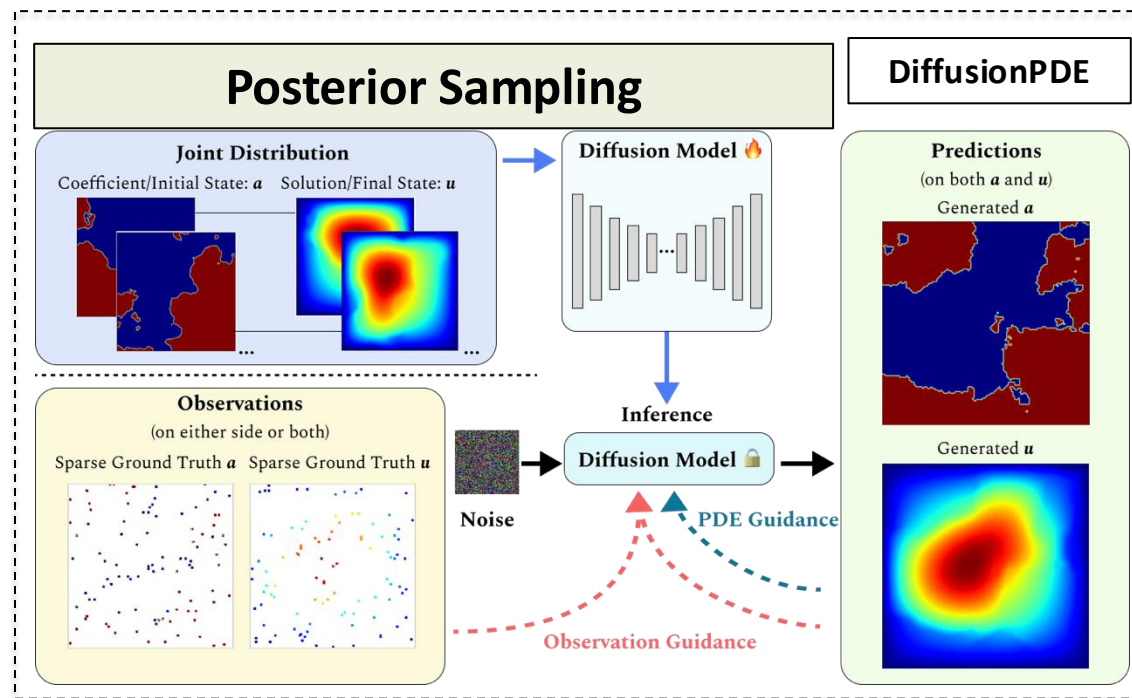
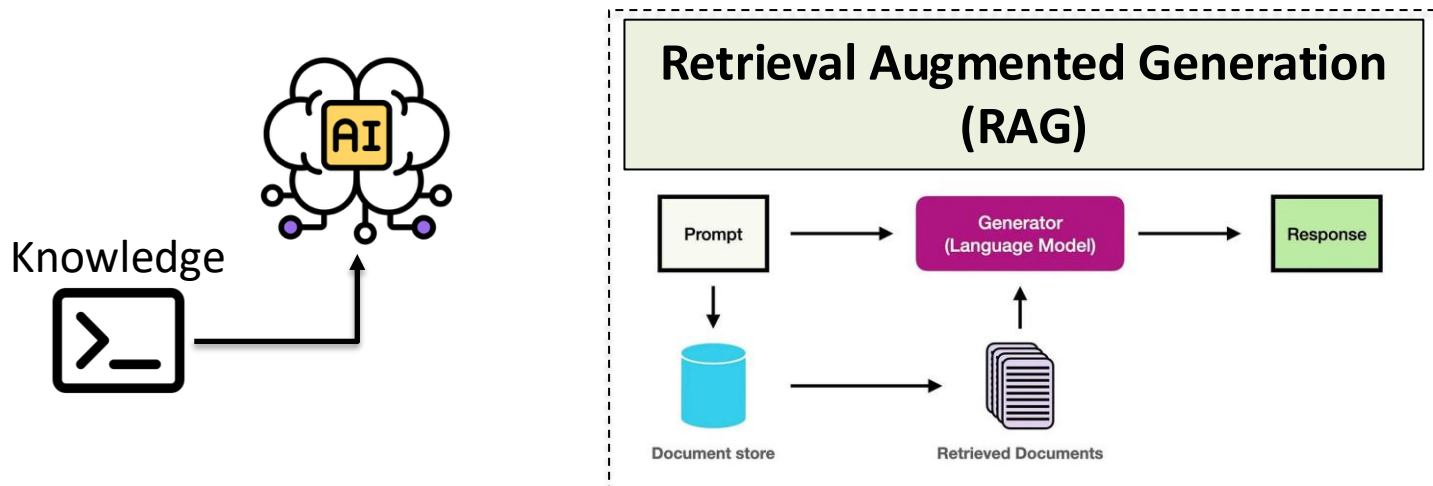
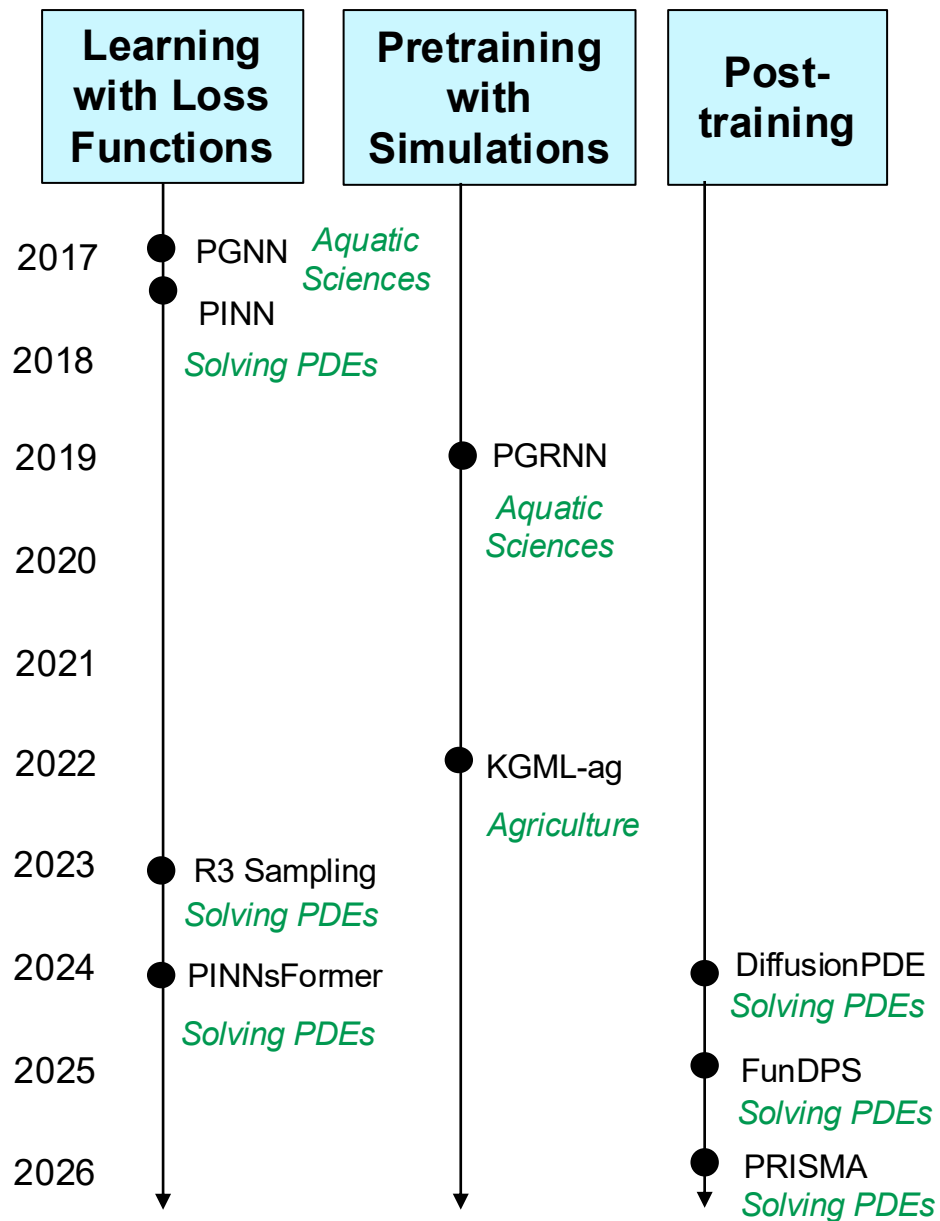
PGNN

PGRNN

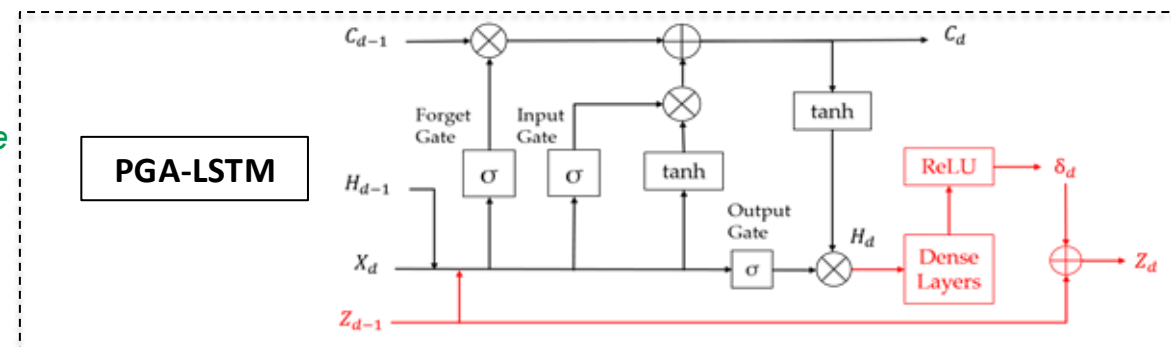
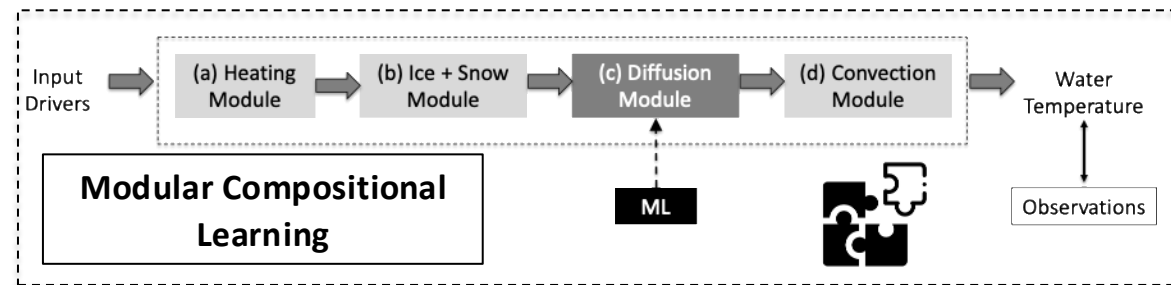
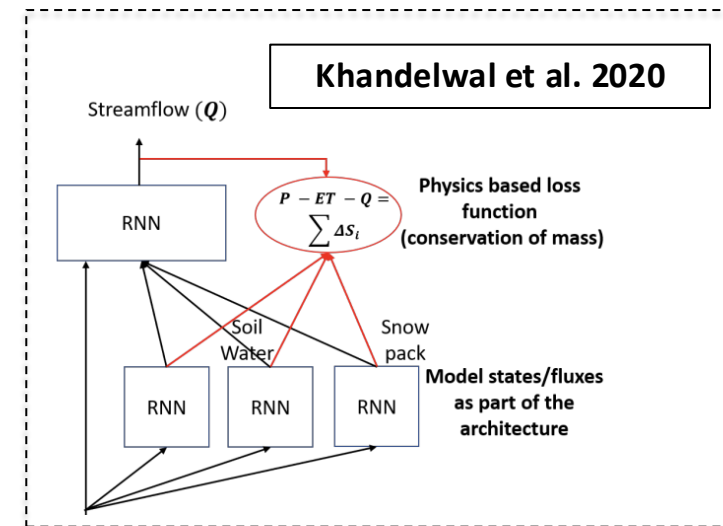
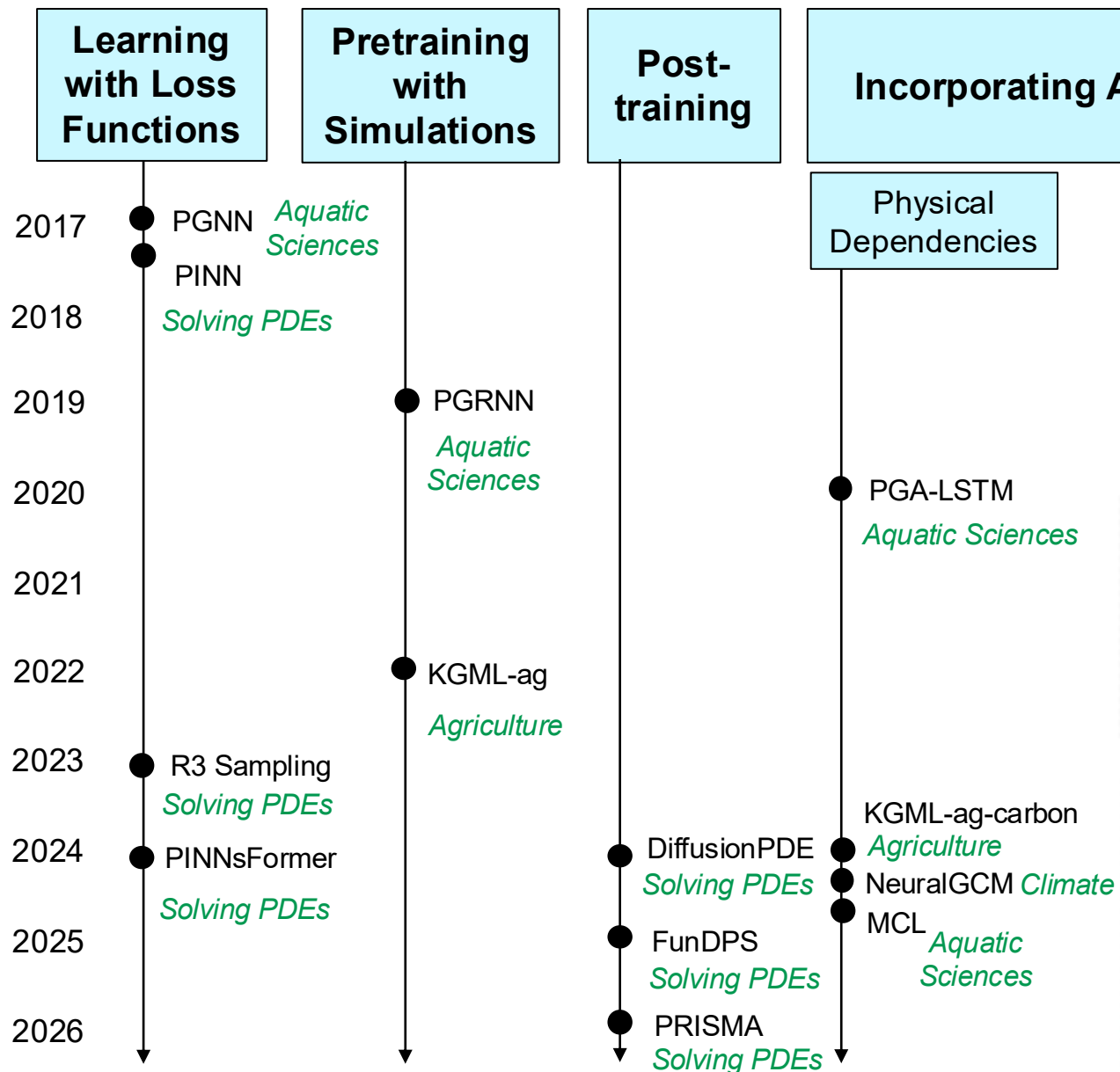
How is KGML Advancing the Science of AI?



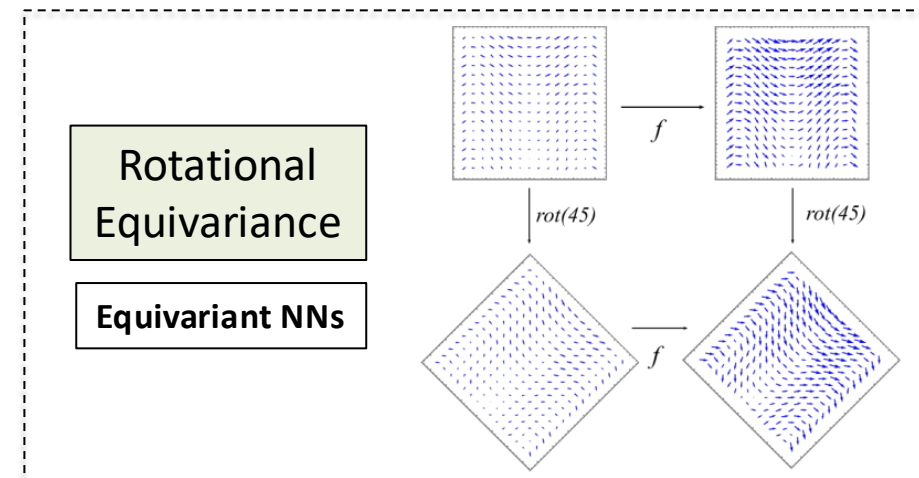
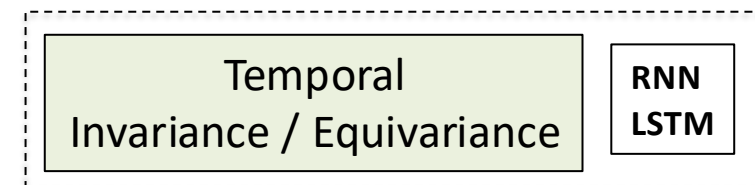
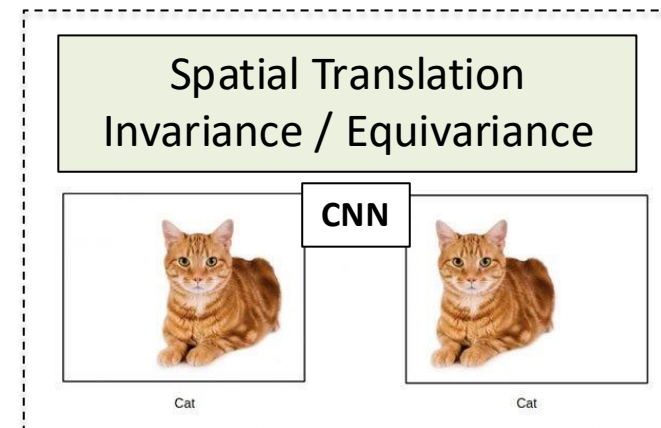
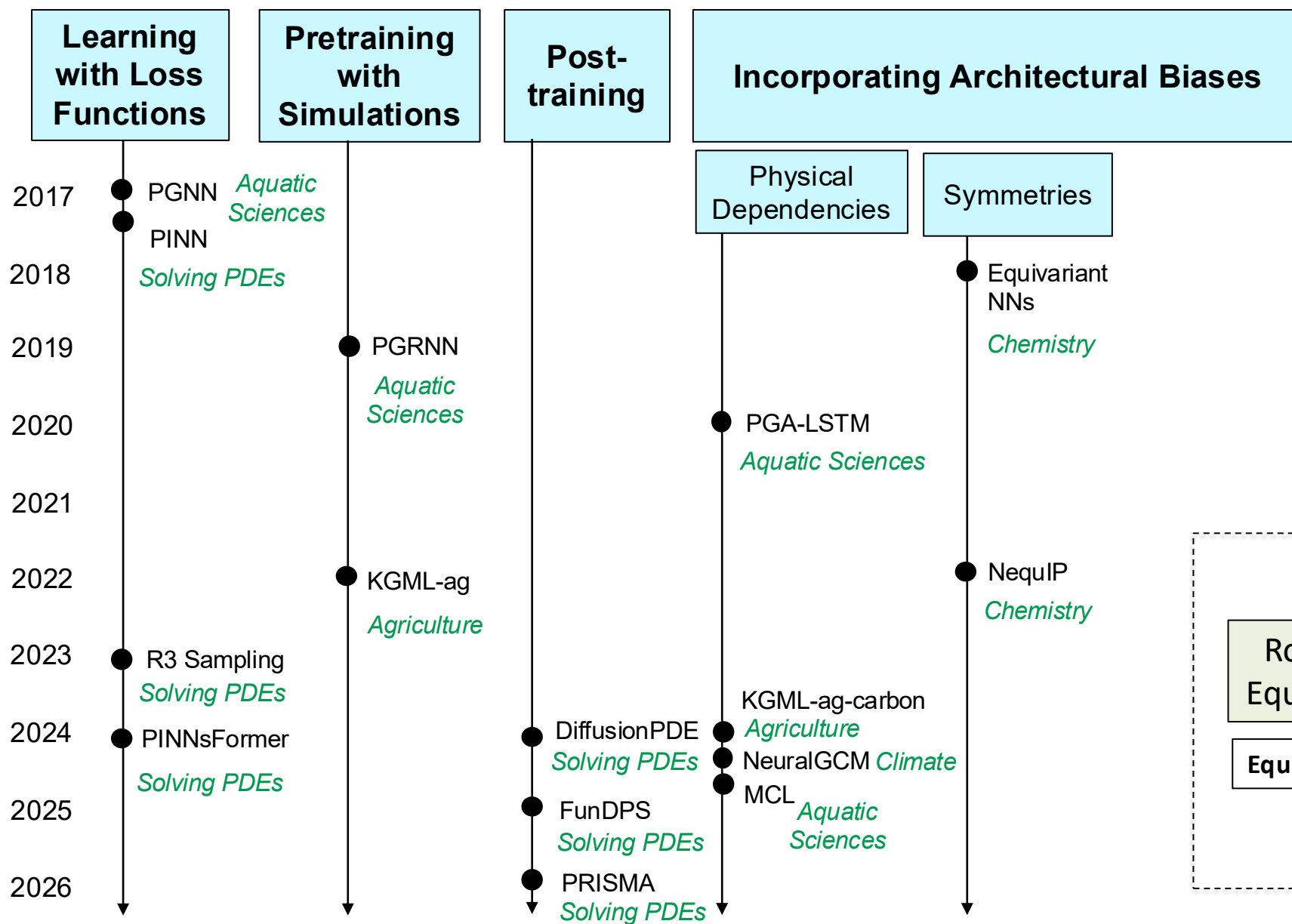
How is KGML Advancing the Science of AI?



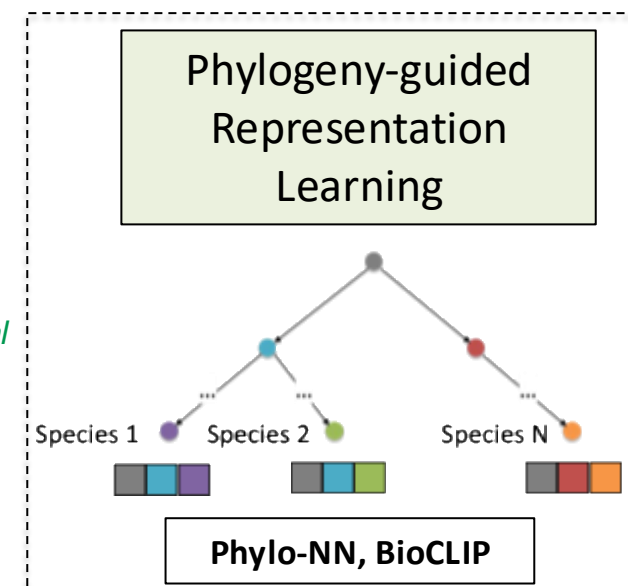
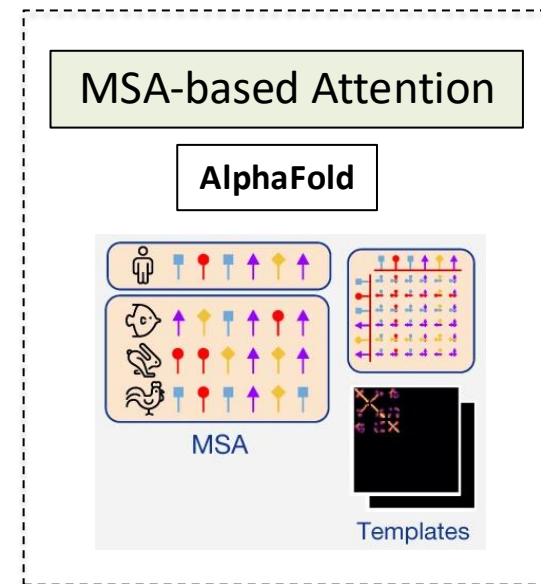
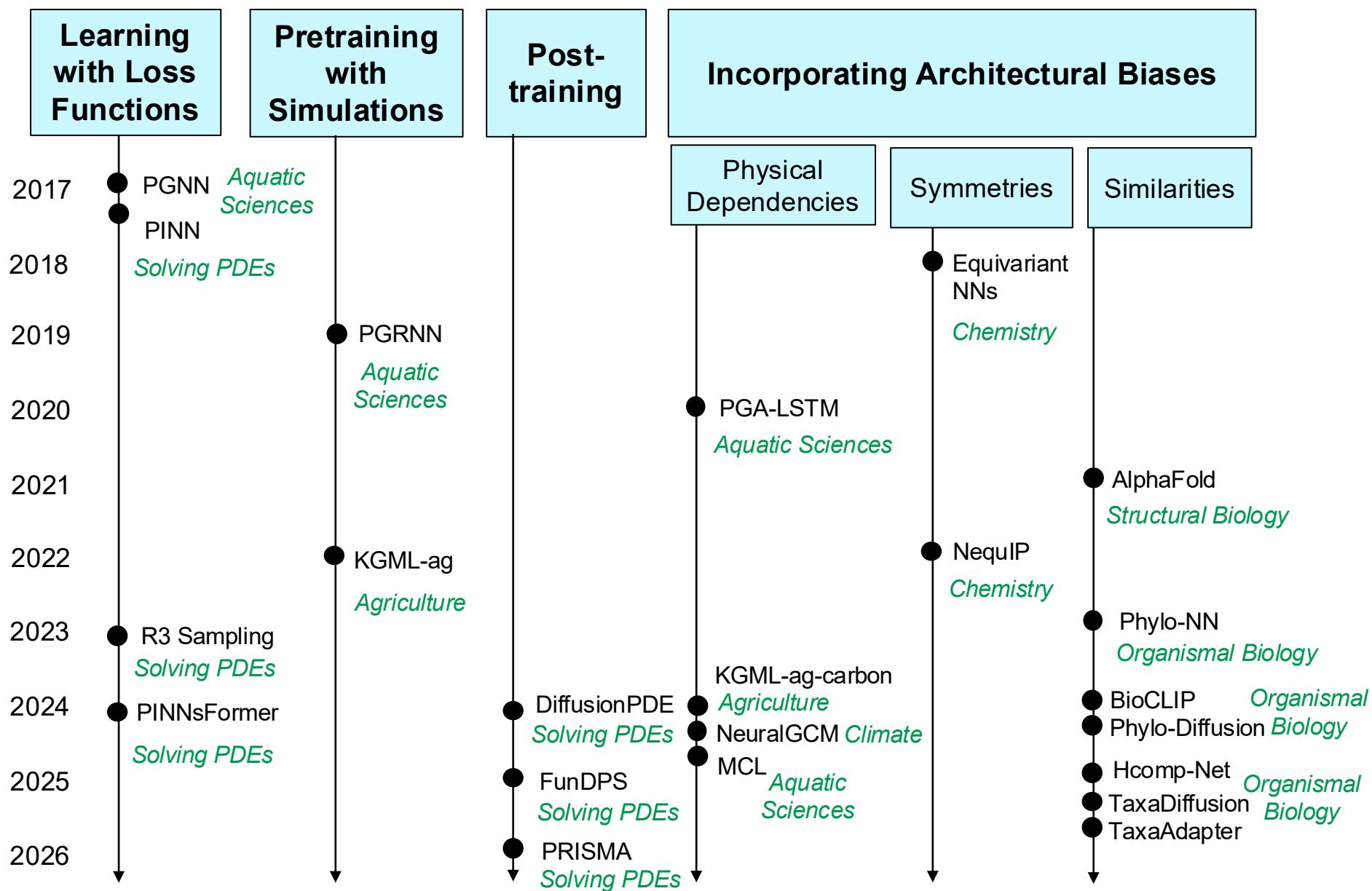
How is KGML Advancing the Science of AI?



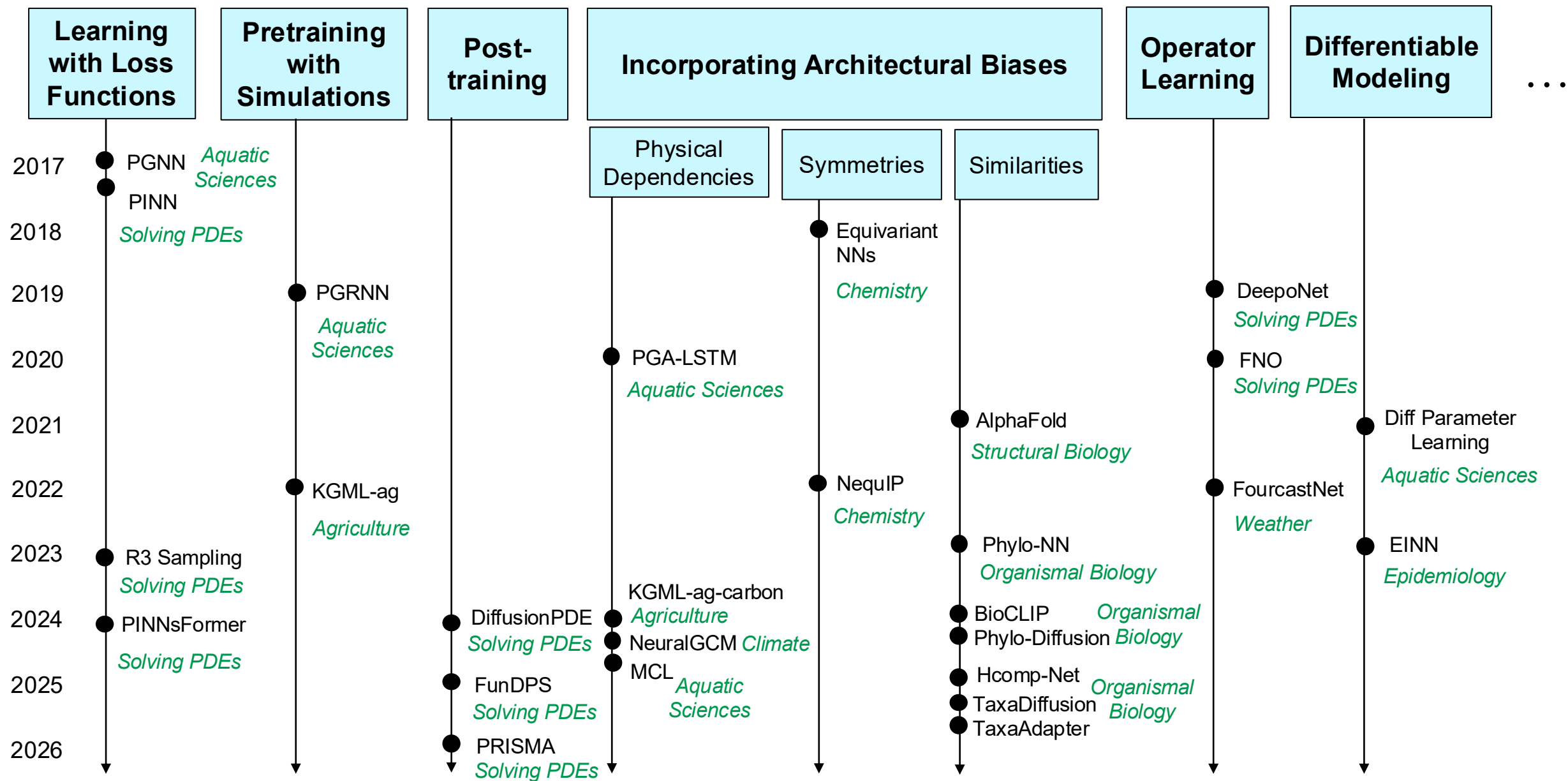
How is KGML Advancing the Science of AI?



How is KGML Advancing the Science of AI?



How is KGML Advancing the Science of AI?



Overview of Research Projects in KGML

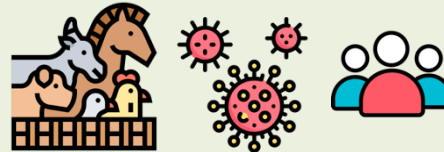
Biodiversity Science: Discovering Biological Traits from Images



\$15M NSF HDR
Institute Grant

Predicting Mutations in **Zoonotic Viruses**

\$18M NSF PIPP
Institute Grant



Aquatic Sciences: Modeling Quality of Water in Lakes



Climate Change



Land Use



Water Demands

\$1.1M NSF MSB Grant



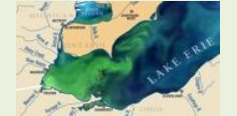
Lakes and Reservoirs

NAIRR Pilot

ReaLSAT: Monitoring Changes in Surface Water Bodies using Satellite Data



Survival of Fish Species



Harmful Algal Blooms



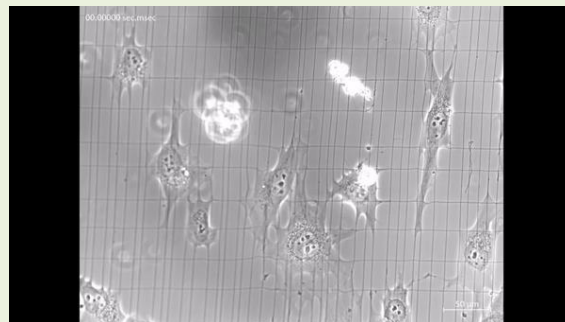
Safe Drinking Water

Farmer Discovery and Soil Organic Carbon Modeling

\$10M
USDA NIFA
SAS Grant



Predicting Mechanics of Cells

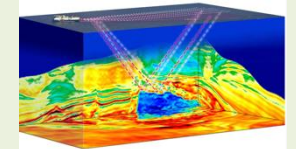
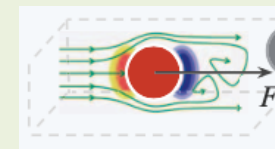


\$1M NSF Medium Grant

Video #1

Solving **PDEs, Inverse Problems** in Geophysics, Quantum Mechanics, Multi-phase Flow, Optics

\$596K NSF CAREER Grant



Overview of Research Projects in KGML

Biodiversity Science: Discovering Biological Traits from Images

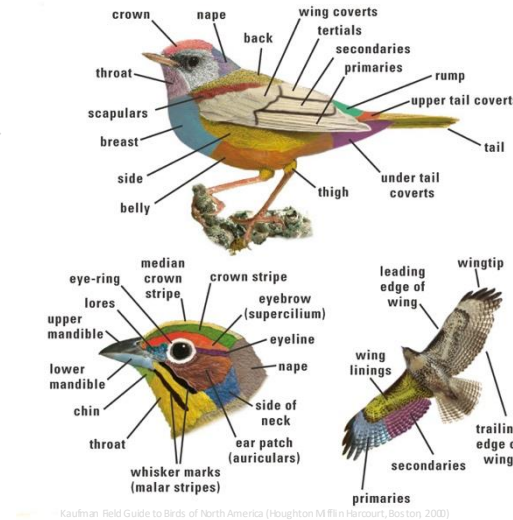


\$15M NSF HDR
Institute Grant

Machine
Learning

Trait-based Bio
Knowledge

Discovery of Biological Traits



Making biological traits computable

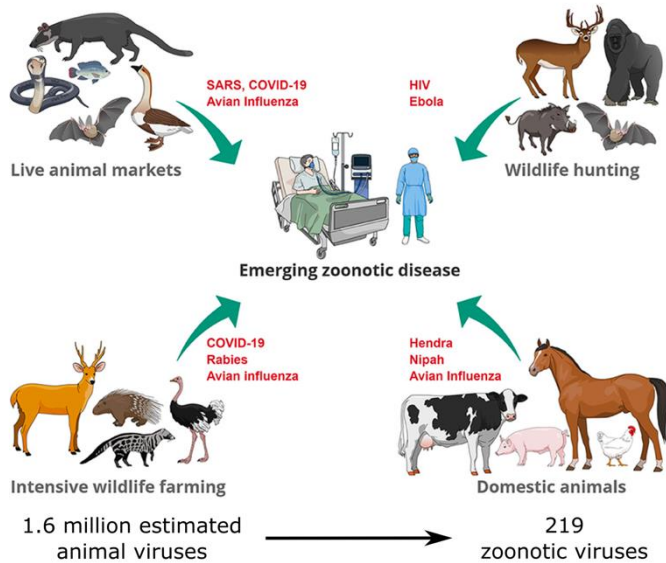
Phylo-NN, KDD 2023
Phylo-Diffusion, ECCV 2024
HCompNet, ICLR 2025
TaxaDiffusion, ICCV 2025
TaxaAdapter, 2026

*Knowledge-guided
Representation Learning,
Architectural Biases,
Generative AI,
Interpretable AI,
Post-training*

Newly established field of “*Imageomics*” -omics
from images

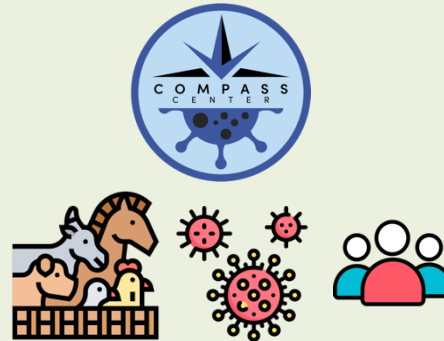


Overview of Research Projects in KGML



Predicting Mutations in Zoonotic Viruses

\$18M NSF PIPP
Institute Grant



Cornell University



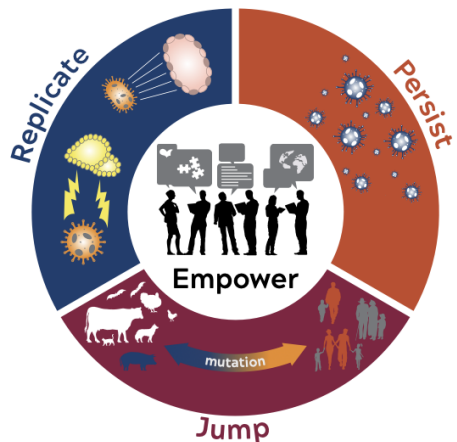
Wake Forest
School of Medicine



COMMunity Empowering Pandemic Prediction and
Prevention from Atoms to Societies (COMPASS) Center

<https://compass-pipp.org/>

*Accurately foresee pandemics and
proactively minimize their impact*



How does a virus

JUMP from one species to another?

REPLICATE inside a new host?

PERSIST in the environment?

How can we **EMPOWER** the public
with pandemic science knowledge?

HAVEN (Hierarchical Attention for Viral protEin-based host iNference)

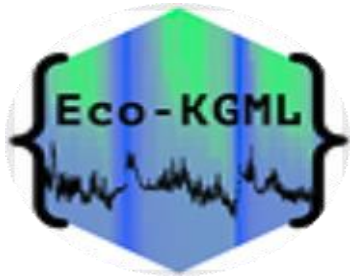
First foundation model for predicting
virus-host interactions

Antony et al., 2026

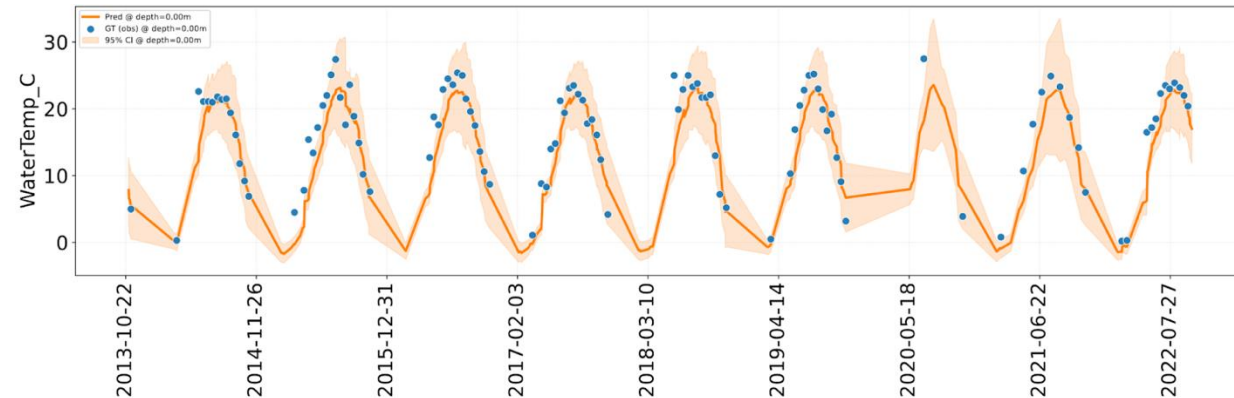
VILLA (Versatile Information Retrieval From Scientific Literature Using Large Language Models)

Antony et al.,
KDD 2026

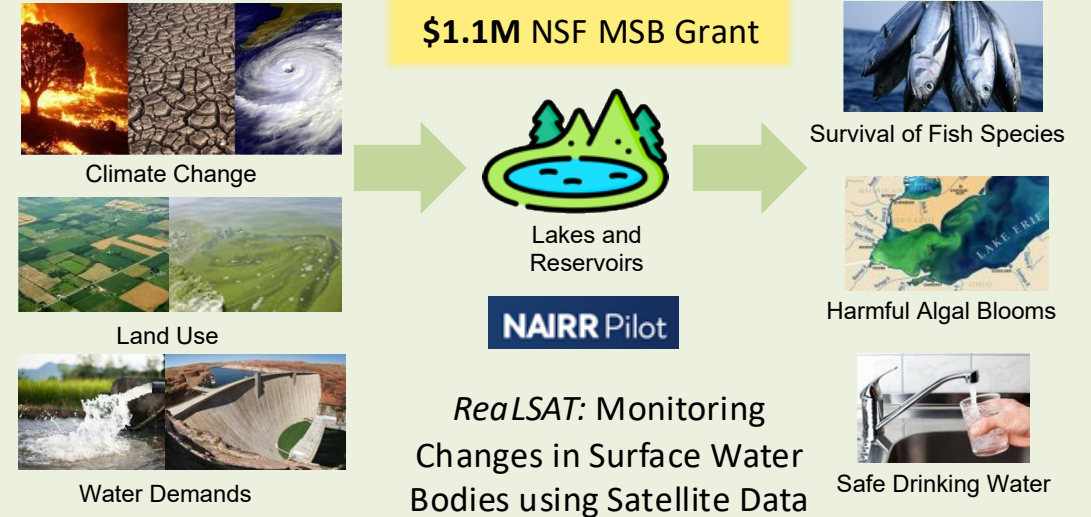
Overview of Research Projects in KGML



Recognized at the **White House** on the launch of NAIRR program in 2024



Aquatic Sciences: Modeling Quality of Water in Lakes



PGNN, 2017
PGRNN, SDM 2019
PGA-LSTM, SDM 2020
RealSAT, Scientific Data 2022
MCL, JAMES 2024
MissTSM, TMLR 2026
LakeFM, 2026

*Knowledge-guided
Loss Functions,
Pretraining,
Architectural Biases*

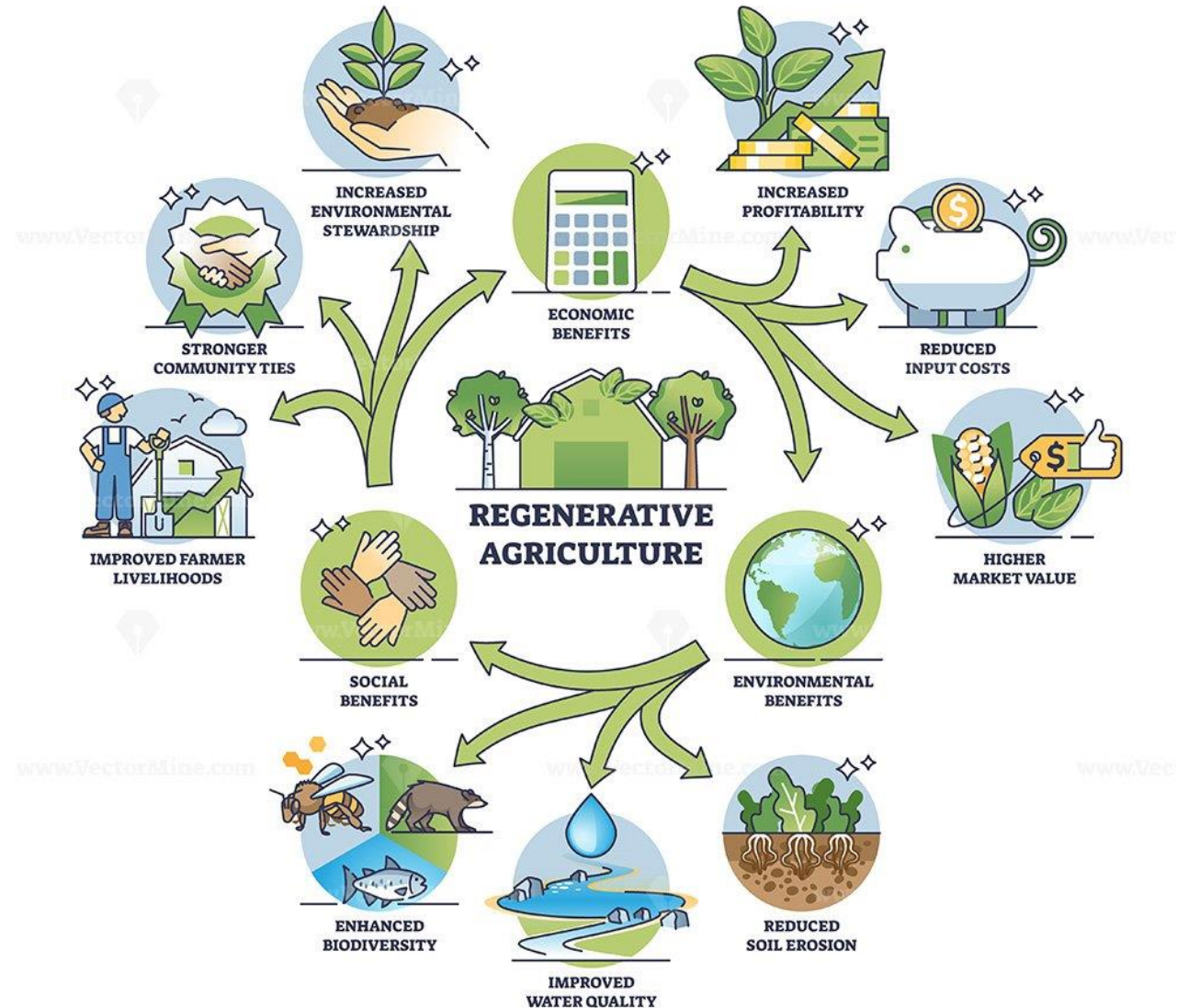
Overview of Research Projects in KGML

Enabling Farmer Discovery and Managing Critical Tradeoffs with the Emergence of National Scale Carbon Markets



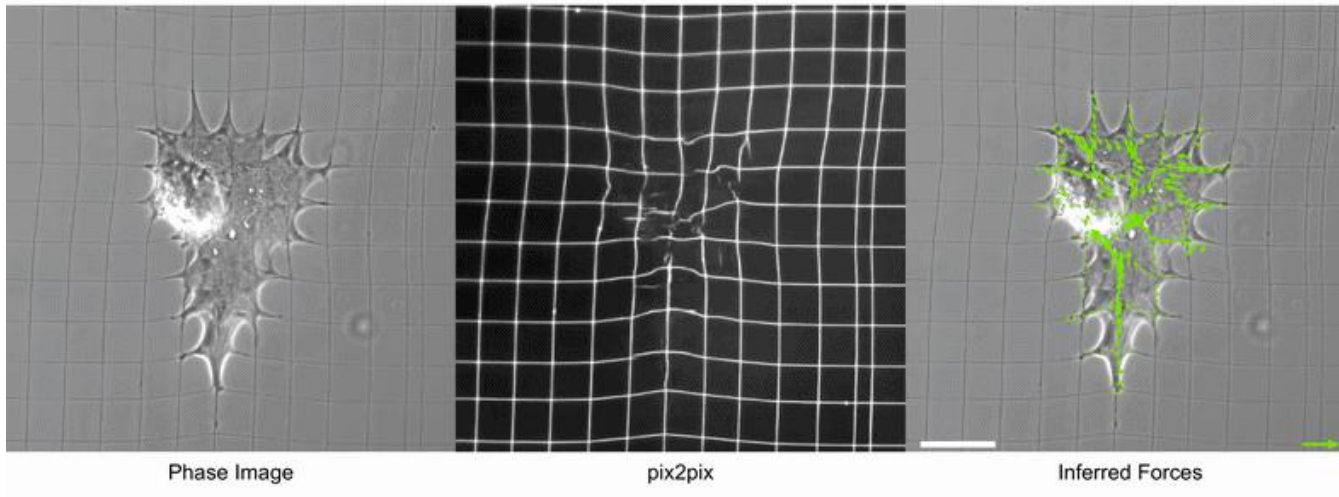
**Farmer Discovery and
Soil Organic Carbon
Modeling**

\$10M
USDA NIFA
SAS Grant



Overview of Research Projects in KGML

0 min



Deep learning reveals how cells pull, buckle, and navigate fibrous environments

Padhi et al., PNAS 2025

Motion Enhanced Multi-Level Tracker (MEMTrack):
A Deep Learning-Based Approach to Microrobot
Tracking in Dense and Low-Contrast Environments

Sawhney et al., Advanced Intelligent Systems 2024

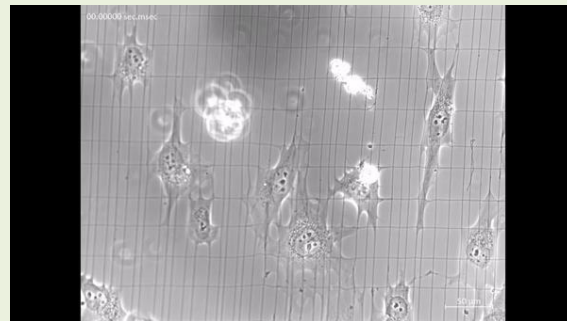


Spinneret-based Tunable
Engineered Parameters Lab

MicroN BASE
Laboratory

Predicting **Mechanics of Cells**

\$1M NSF Medium Grant

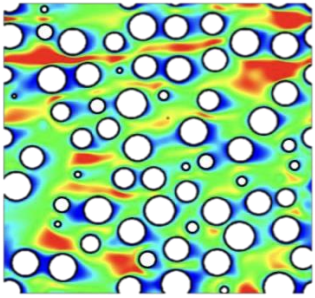


Video #1

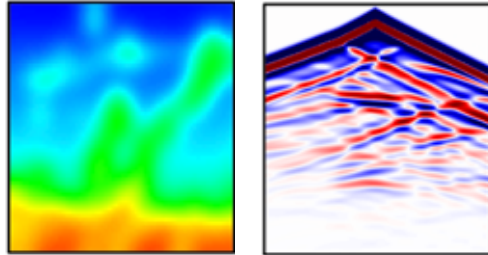
Knowledge-guided
Architectural Biases
Hybrid-physics-AI Models

Overview of Research Projects in KGML

Multi-phase Flow

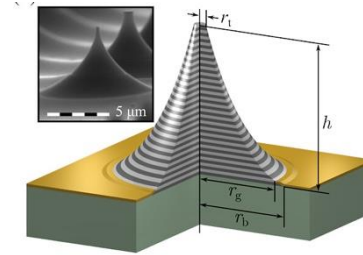


Seismic Imaging

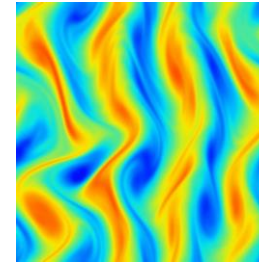


Velocity Maps Seismic Waveform

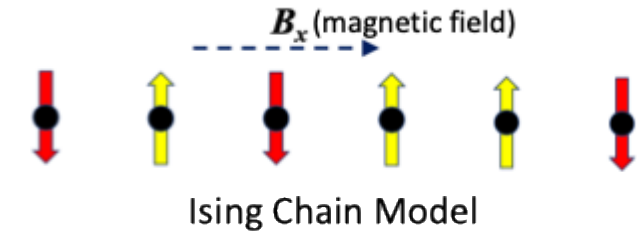
Photonic Funnel



Solving PDEs



Condensed Matter Physics



PhyFlow, ICDM 2021

PhyNet, SDM 2020

GFI, ICLR 2025

Hier-PGNN, ACS Photonics, 2025

QRes, SDM 2021

PID-GAN, KDD 2021

R3 Sampling, ICML 2023

Neuro-Visualizer, ICML 2024

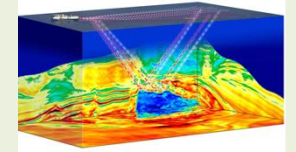
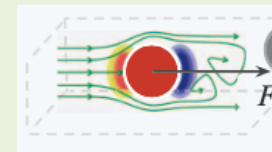
PRISMA, 2026

Cophy-PGNN, ACM TIST 2022

Knowledge-guided
Loss Functions,
Architectural Biases,
Generative AI,
Operator Learning,
Post-training

Solving PDEs, Inverse Problems in
Geophysics, Quantum Mechanics,
Multi-phase Flow, Optics

\$596K NSF CAREER Grant



Overview of Research Projects in KGML

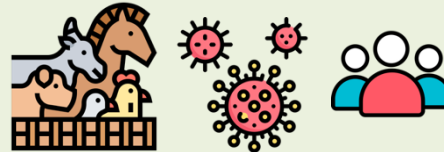
Biodiversity Science: Discovering Biological Traits from Images



\$15M NSF HDR
Institute Grant

Predicting Mutations in **Zoonotic Viruses**

\$18M NSF PIPP
Institute Grant



Aquatic Sciences: Modeling Quality of Water in Lakes



Climate Change



Land Use



Water Demands

\$1.1M NSF MSB Grant



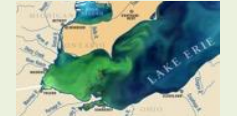
Lakes and Reservoirs

NAIRR Pilot

ReaLSAT: Monitoring Changes in Surface Water Bodies using Satellite Data



Survival of Fish Species



Harmful Algal Blooms



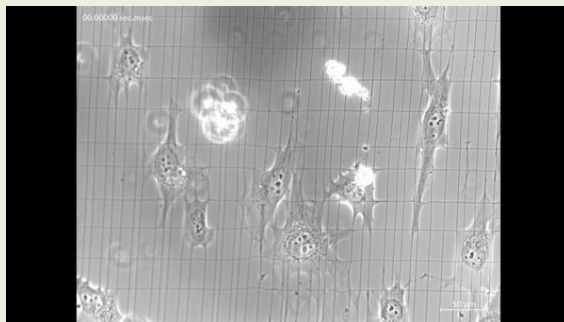
Safe Drinking Water

Farmer Discovery and Soil Organic Carbon Modeling

\$10M
USDA NIFA
SAS Grant



Predicting Mechanics of Cells

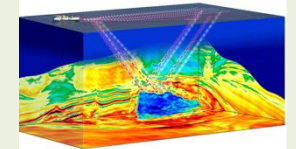
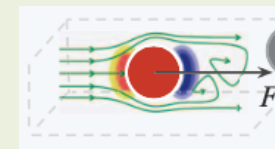


\$1M NSF Medium Grant

Video #1

Solving **PDEs, Inverse Problems** in Geophysics, Quantum Mechanics, Multi-phase Flow, Optics

\$596K NSF CAREER Grant



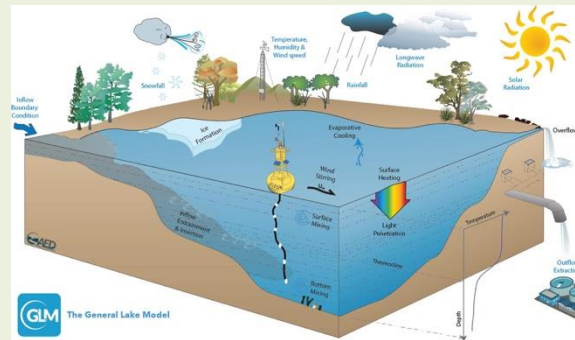
Illustrative Case Studies for This Talk

Biodiversity Science: Discovering Biological Traits from Images



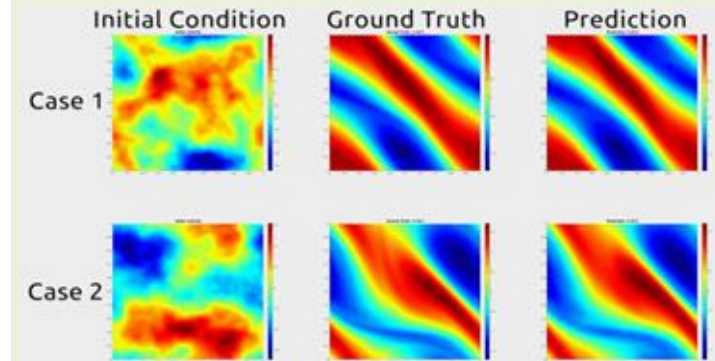
Knowledge:
Phylogenetic Trees

Aquatic Sciences: Modeling Quality of Water in Lakes



Knowledge:
Process-based Models

Solving PDEs (Partial Differential Equations)

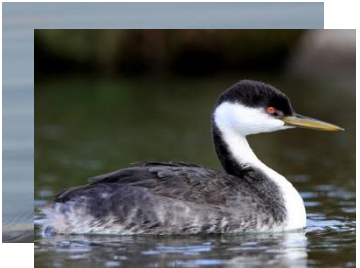
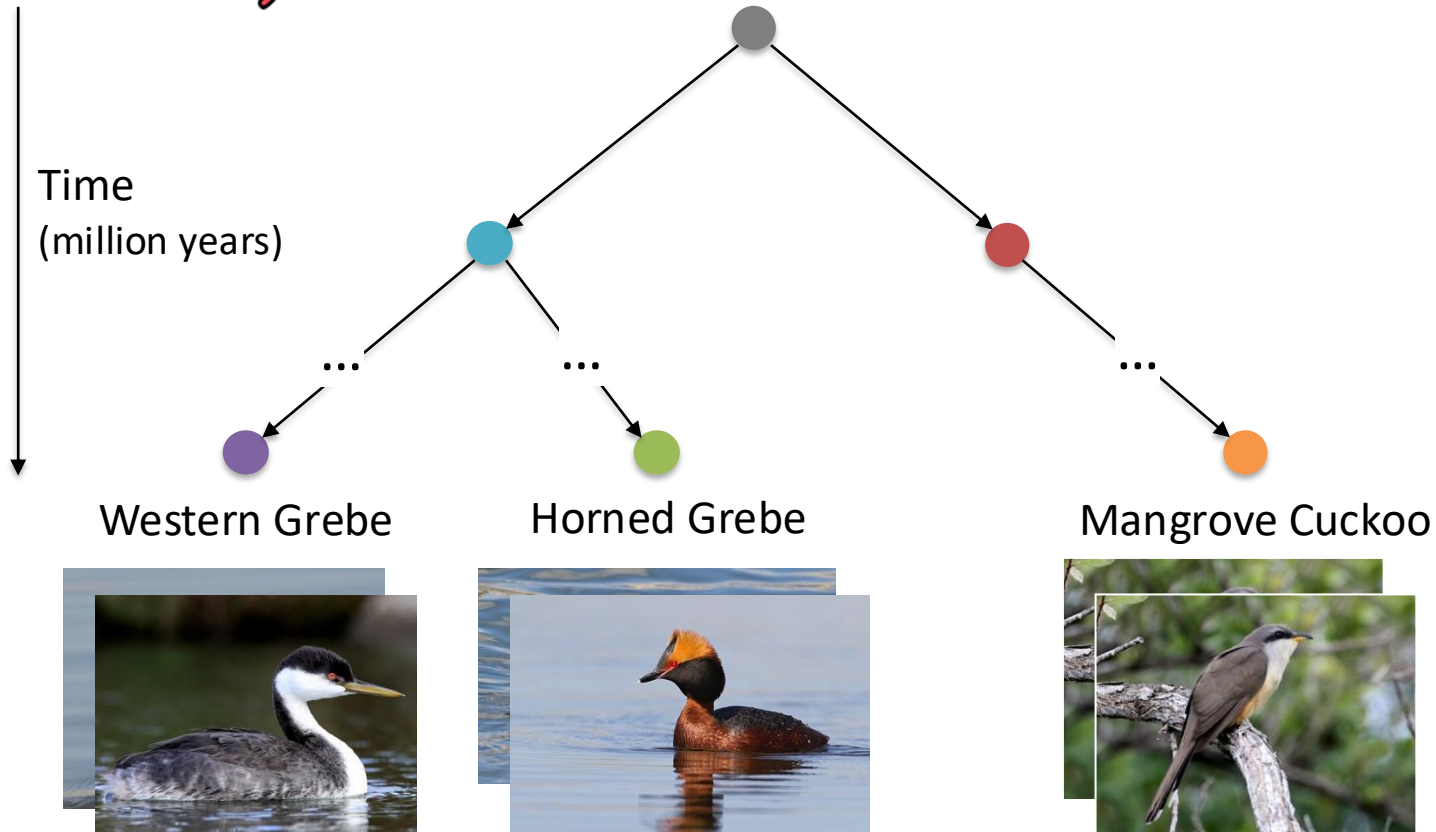


Knowledge:
Equations

Case Study: Discovering **Evolutionary** Traits from Images



Phylogenetic Tree

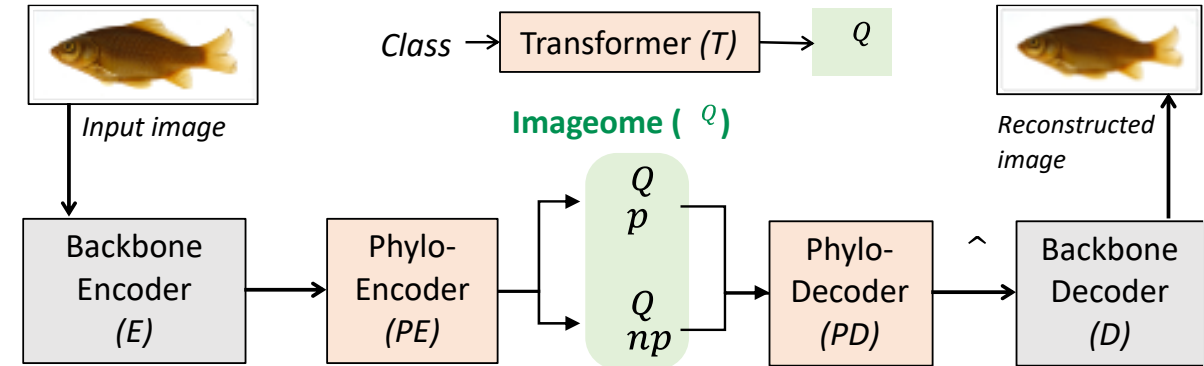
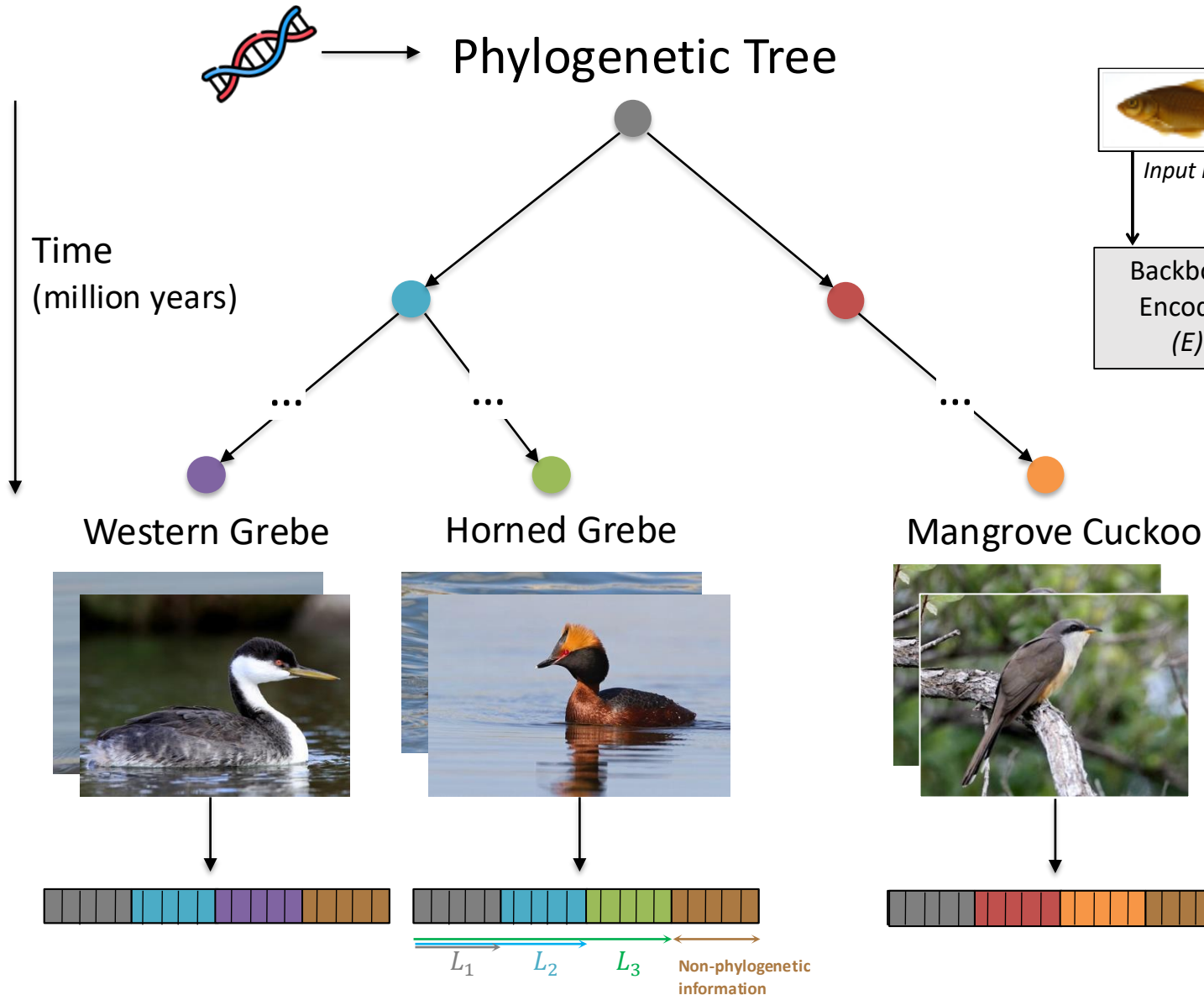


- Can we discover image features (traits) linked to evolution for answering questions such as:
 - What makes a species different from its siblings?
 - When did a species evolve a specific trait?

Phylogenetic tree captures knowledge of species evolution

Phylogeny-guided Neural Networks (Phylo-NNs)

Elhamod et al., KDD 2023



- Learn image features (or **imageomes**, *akin to genomes*) structured by phylogenetic knowledge
- Disentangles information of images into feature segments (*traits*) acquired at every ancestry level

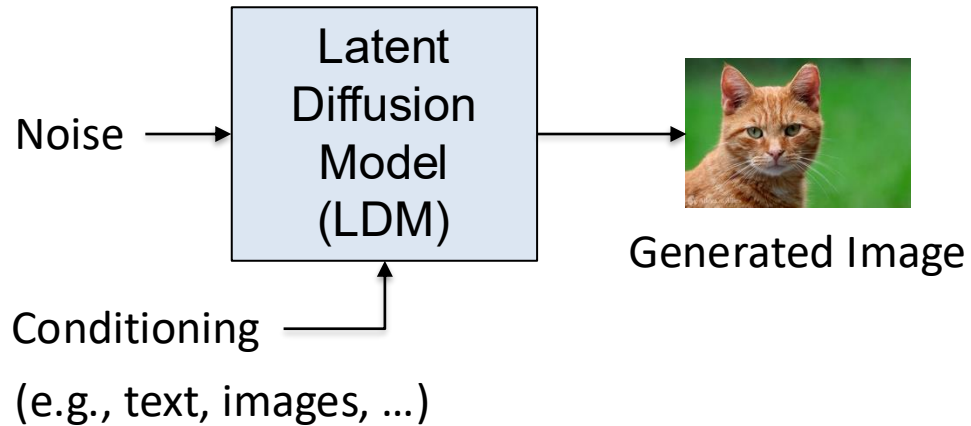
Knowledge-guided Representation Learning

Knowledge-guided Architectural Biases

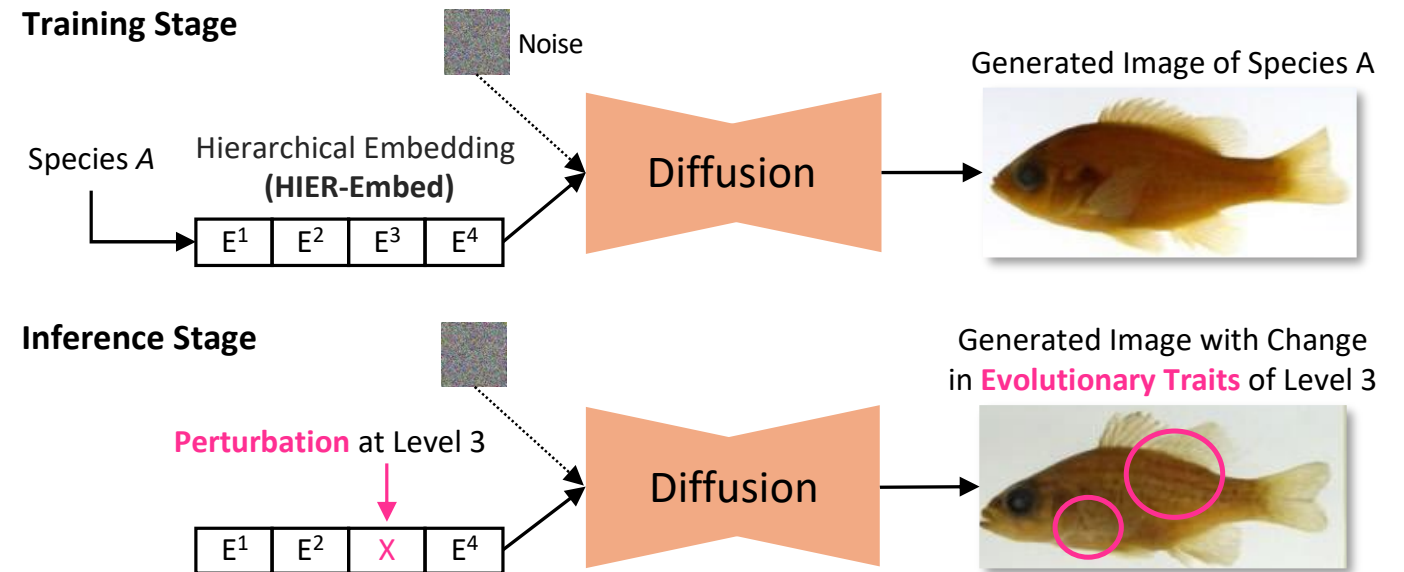
Phylogeny-guided Diffusion Models

Khurana et al., ECCV 2024

<https://imageomics.github.io/phylo-diffusion/>

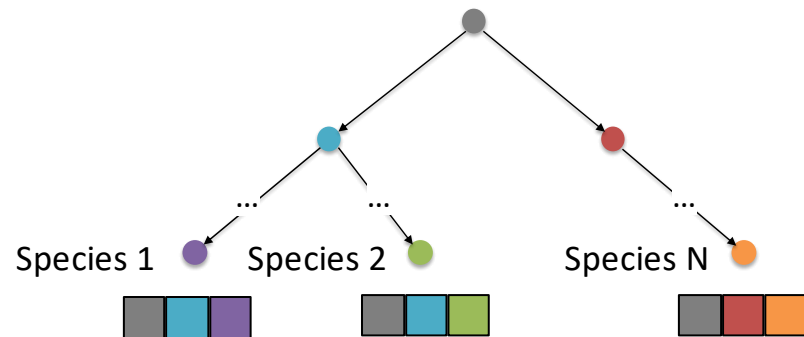


Key Idea: Query LDMs using tree-structured knowledge of phylogeny



Manipulating embeddings can reveal trait changes in images, producing novel insights of biological timing of trait acquisition

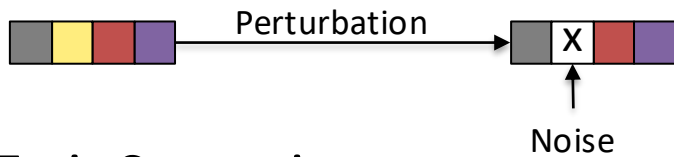
Knowledge-guided Generative AI



Manipulating Embedding Space of Phylo-Diffusion

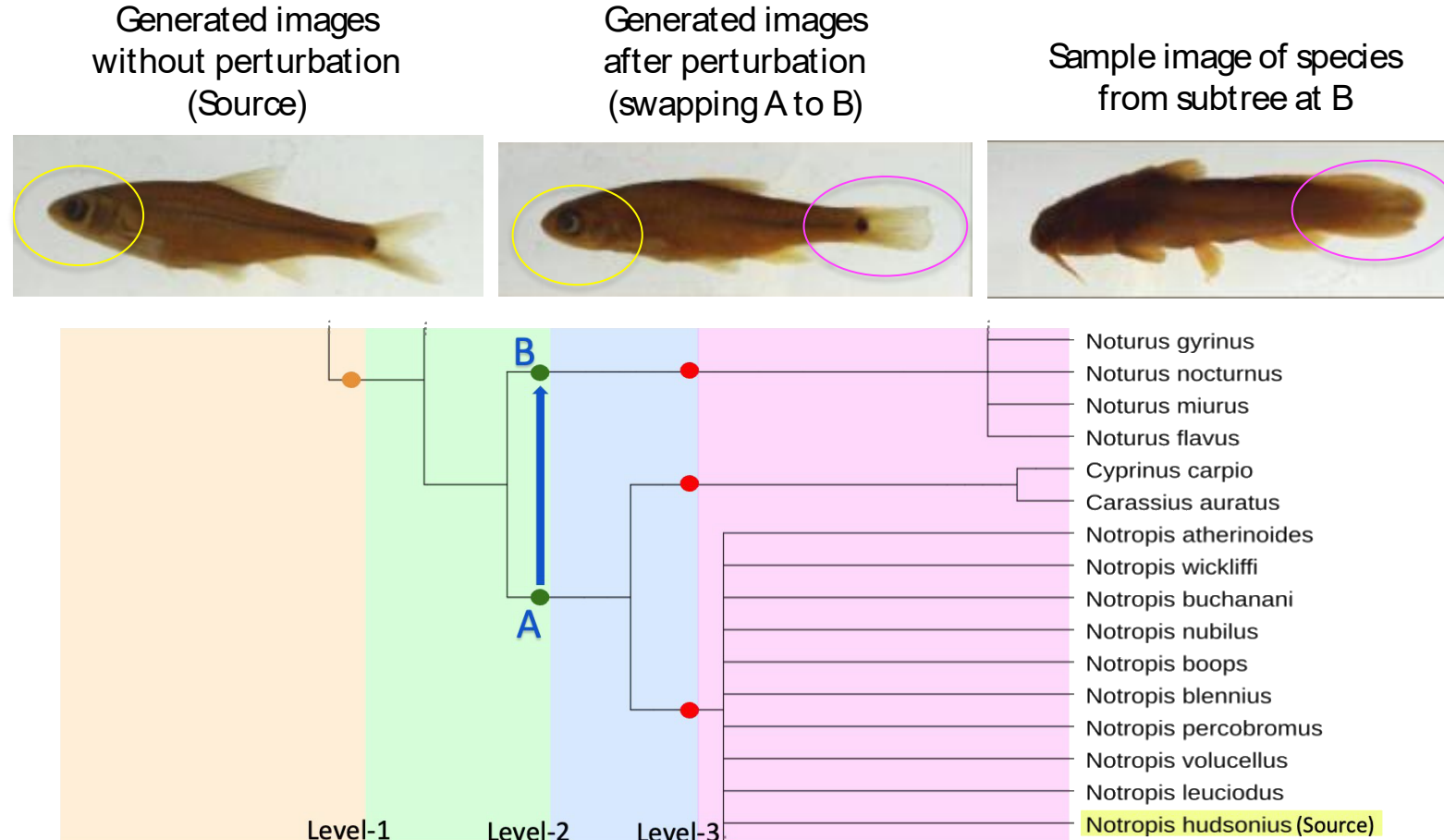
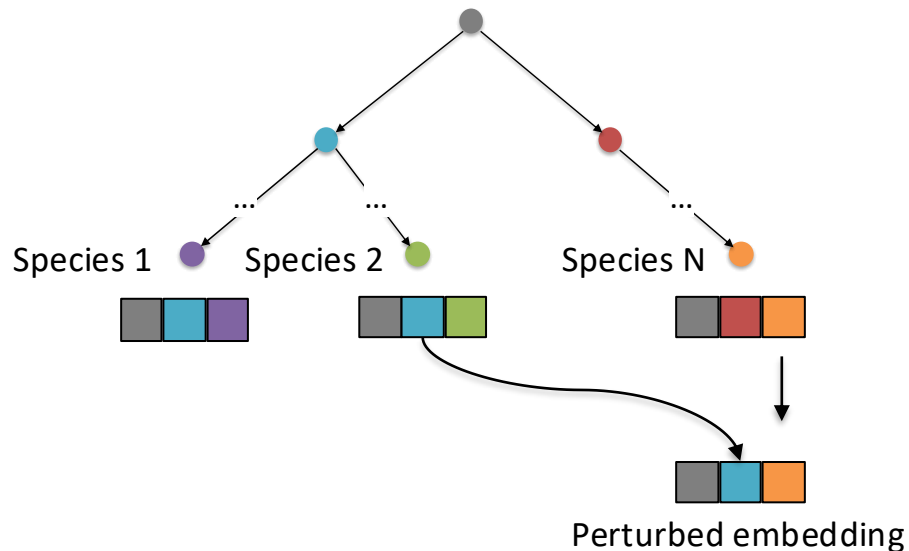
1. Trait Masking

- Replace components with noise
- Analogous to “Gene Knockout”



2. Trait Swapping

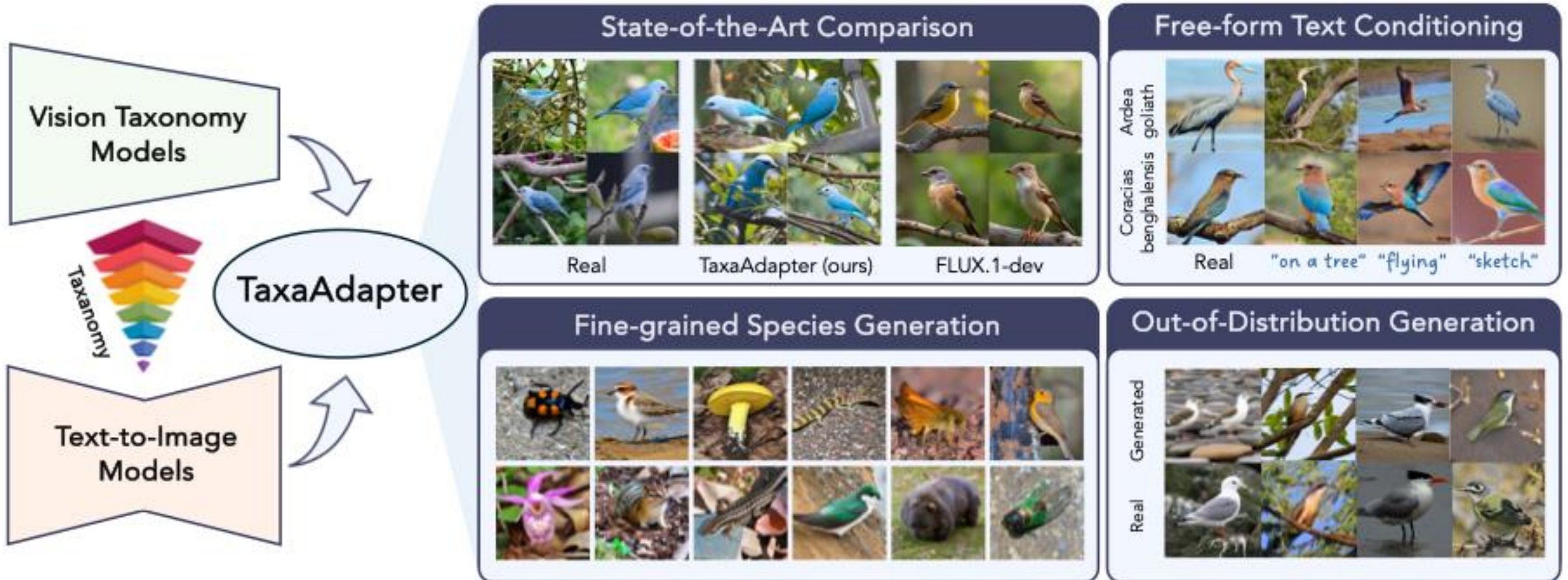
- Replace components with siblings
- Analogous to “CRISPR experiments”



Phylo-Diffusion generates **novel scientific hypotheses** of trait evolution directly from images

TaxaAdapter: Coupling Vision Taxonomy Models with Diffusion Models for Fine-Grained Species Generation

Khurana et al. 2026



Knowledge-guided Post-training

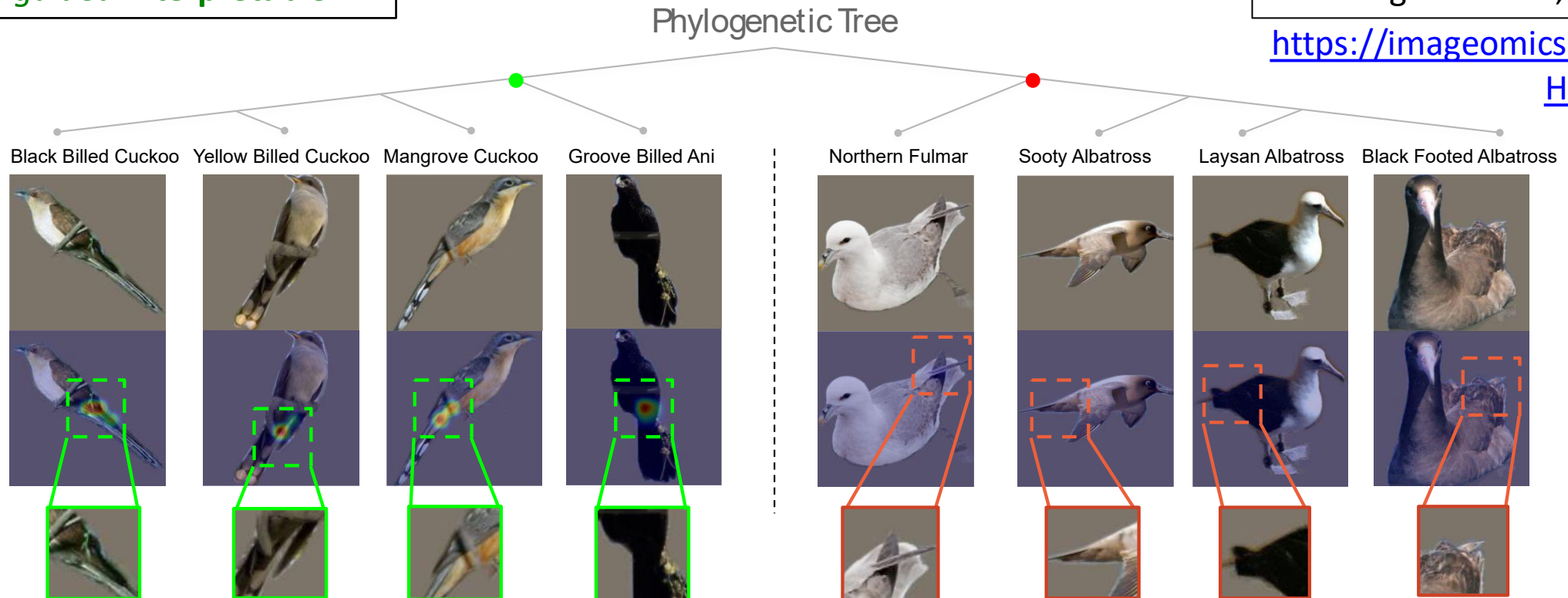
Discovering Evolutionary Traits as Prototypes:

HComP-Net: Hierarchy-aligned **Commonality** through **ProtoP-Nets**

Knowledge-guided Interpretable AI

Manogaran et al., ICLR 2025

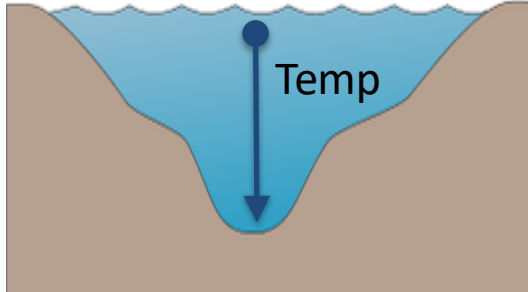
<https://imageomics.github.io/HComPNet/>



What is common among all descendant species of a subtree (green node) that is absent from other species?

- HComP-Net avoids learning over-specific prototypes at internal nodes
- HComP-Net can abstain from finding prototypes at higher levels if there is nothing in common

Case Study 2: Modeling Quality of Water in Lakes

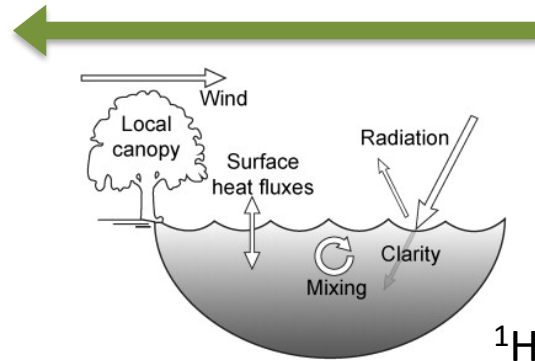


Growth and survival of fisheries



Harmful Algal Blooms

General Lake Model¹ (GLM)



¹Hipsey et al., 2014

Input Drivers (observed):

Short-wave Radiation,
Long-wave Radiation,
Air Temperature, ...

How can we utilize:

- ***imperfect*** knowledge of process-based model runs?
- ***partial*** knowledge of relationships among physical variables?

Limitations of Process-based Models (e.g., GLM):

- Parameter needs to be calibrated for every system (lake)
 - Default settings of parameters often inaccurate
- Modeling biases
 - Approximations for comp. gains
 - Missing or partial knowledge of some processes

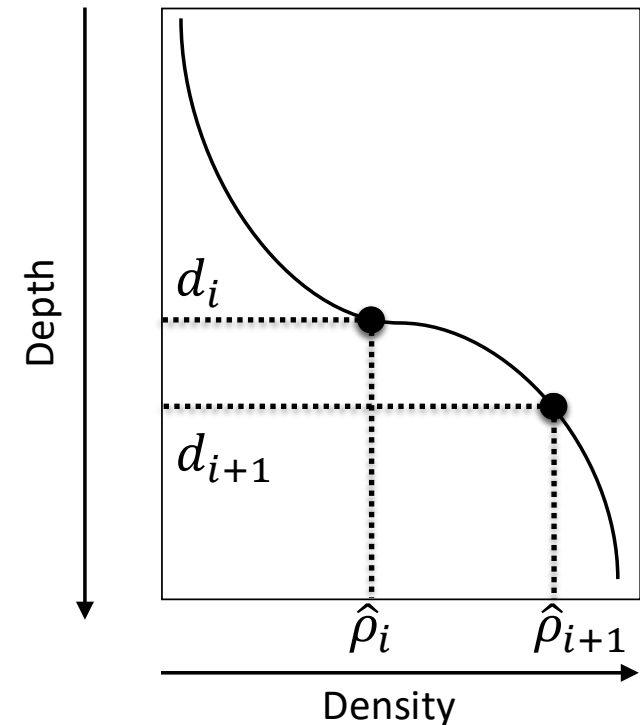
Physics-guided Neural Networks (PGNN):

Incorporating imperfect/partial knowledge

Karpatne et al. 2017

Innovation 1: Use Knowledge-guided Loss Functions

- Physical Constraint: density should increase with depth



For any consecutive depth pair, $d_i < d_{i+1}$

$$\Delta_i = \hat{\rho}_i - \hat{\rho}_{i+1} \leq 0$$

Physical Violation = $\text{ReLU}(\Delta_i)$

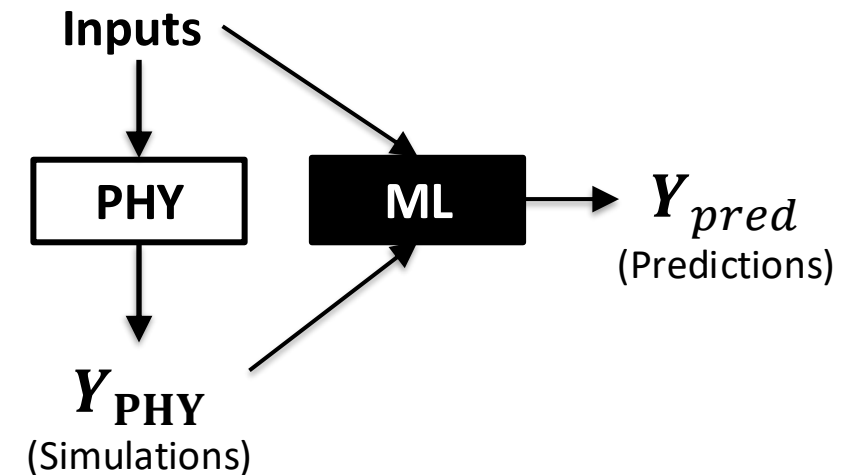
$$\text{Physics-guided Loss}(Y_{pred}) = \sum_i \text{ReLU}(\Delta_i)$$

Advantages:

- Does not require labels (Y_{true}) !
- Can be evaluated on unlabeled data

Innovation 2: Hybrid-physics-AI Modeling

- Use process outputs as additional inputs in ML



Prediction Loss (, ^) $\oplus$ $\sum_i \text{ReLU}(\Delta_i)$ **Physics-guided Loss**

Modular Compositional Learning (MCL): A New Paradigm for Creating Hybrid-Process-ML Models

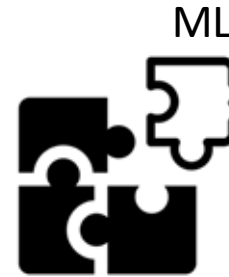
Ladwig et al. JAMES 2024

- Modular approach to process-based modeling
 - Every module can be represented at varying levels of fidelity to meet speed-accuracy trade-offs

Knowledge-guided Interpretable AI

Modular AI Agents for Science

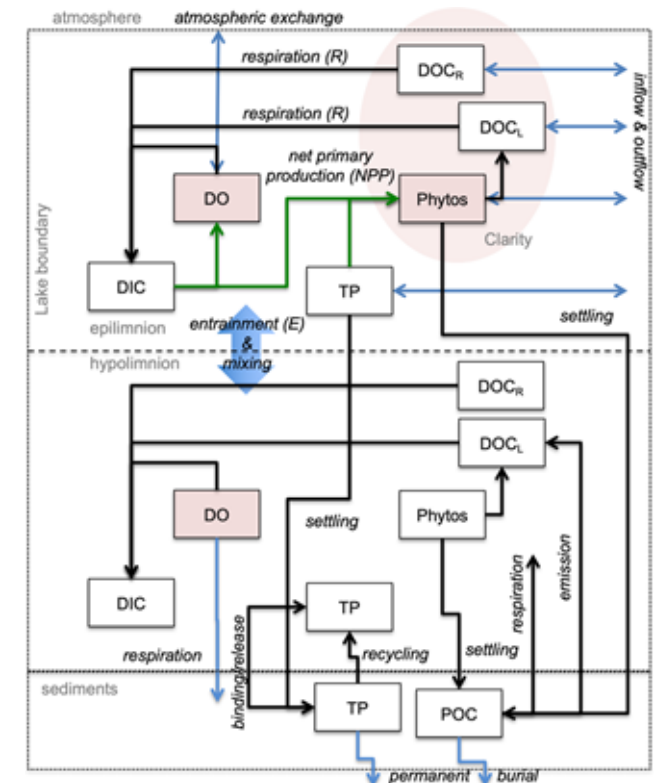
Process-based



Key Idea of Modular Compositional Learning (MCL):

- Use ML to replace/augment a few modules that are known to be deficient, while retaining other modules
- New way of thinking about modular AI agents for science

Box and Arrow Diagrams in Aquatic Sciences

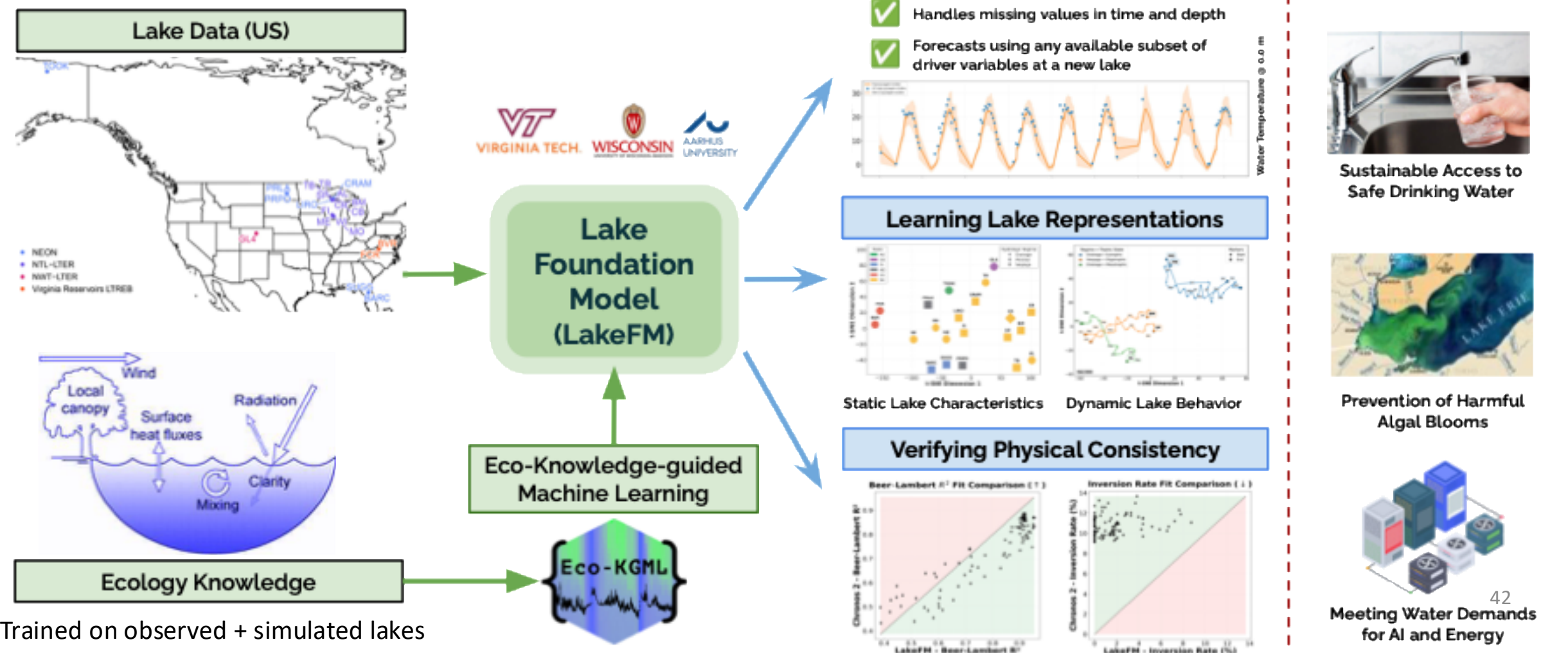


LakeFM: Toward a Foundation Model for Aquatic Sciences

Neog et al. KDD 2026

Recognized at the **White House Event**
Announcing the Launch of NAIRR Program

Knowledge-guided Representation Learning



Trained on observed + simulated lakes

Case Study 3: Solving PDEs

$$\frac{\partial u(x, t)}{\partial t} + \mathbf{N}(\lambda, u(x, t)) = 0$$

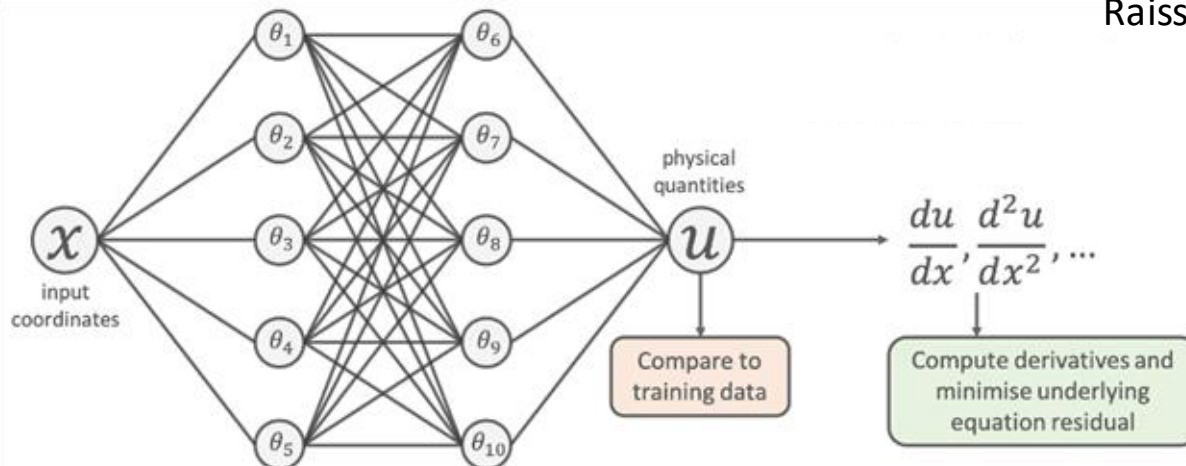
Non-linear differential operator with params λ

Initial/Boundary
Conditions: $u(x, 0) = h(x)$
 $u(x, t) = g(x, t)$

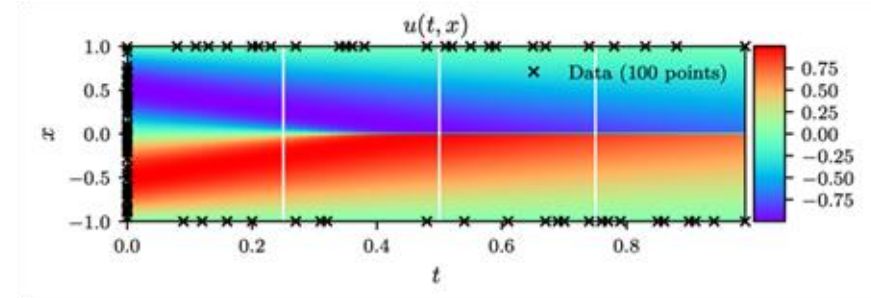
PDE Residuals: $\mathcal{R}_\theta(x, t) = \frac{df_\theta(x, t)}{dt} + \mathcal{N}_x[f_\theta(x, t)]$

Framework of Physics-informed Neural Networks (PINNs)

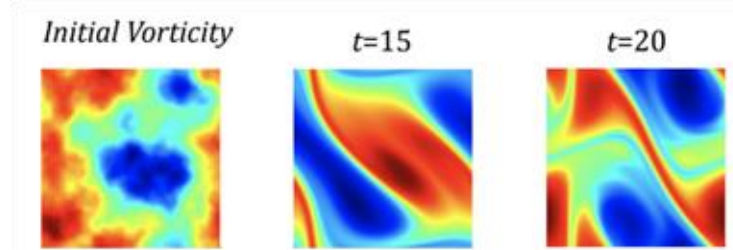
Raissi et al. 2019



Burgers' Equation:



Navier Stokes' Equation:



Training Loss:

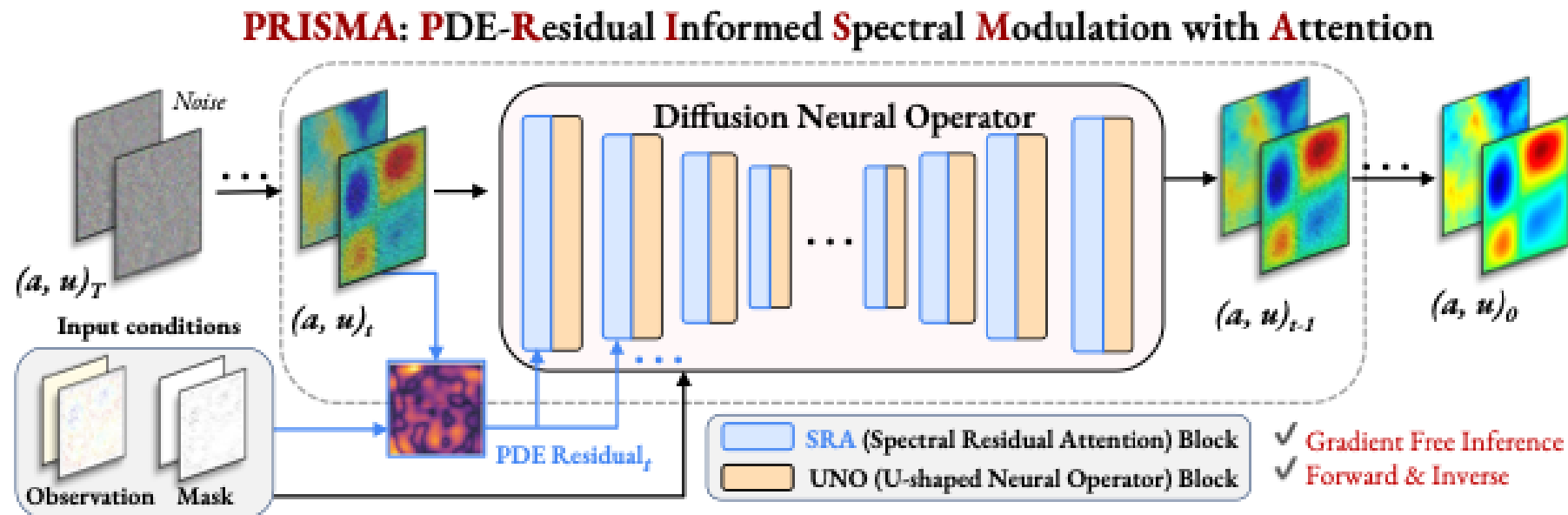
$$\mathcal{L}(\theta) = \lambda_{ic}\mathcal{L}_{ic}(\theta) + \lambda_{bc}\mathcal{L}_{bc}(\theta) + \lambda_r\mathbb{E}_{x \sim p(x)}[\mathcal{R}_\theta(x, t)]^2$$

Initial Condition Loss Boundary Condition Loss **PDE Residual Loss**

- Mesh-free, unsupervised!
- Evaluated on a random sample of **collocation points**

Diffusion Models for Solving PDEs

- Prior works: Posterior sampling methods (DiffusionPDE, FunDPS)
 - Perform gradient gradient of physics residuals during inference
 - Challenge: *slow inference speed* due to expensive gradient calculations
- **PRISMA**
 - Gradient descent-free inference using **physics-informed architecture** (spectral attention)
 - **15x to 250x faster** inference (time) compared to posterior sampling methods

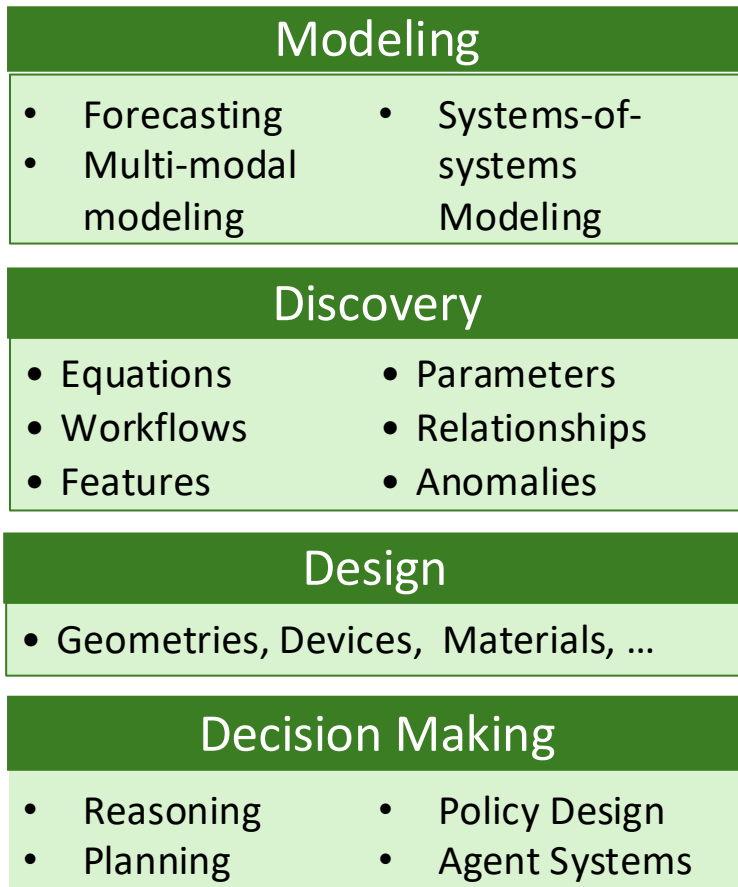


Sawhney et al. 2026

<https://arxiv.org/abs/2512.01370>

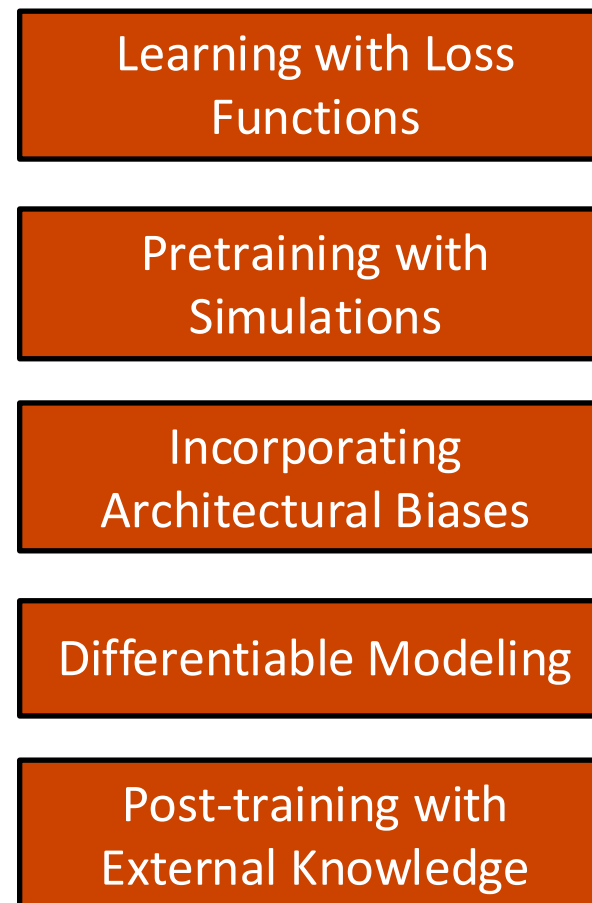
Concluding Remarks

Advancing
AI for Science
transcending disciplinary boundaries



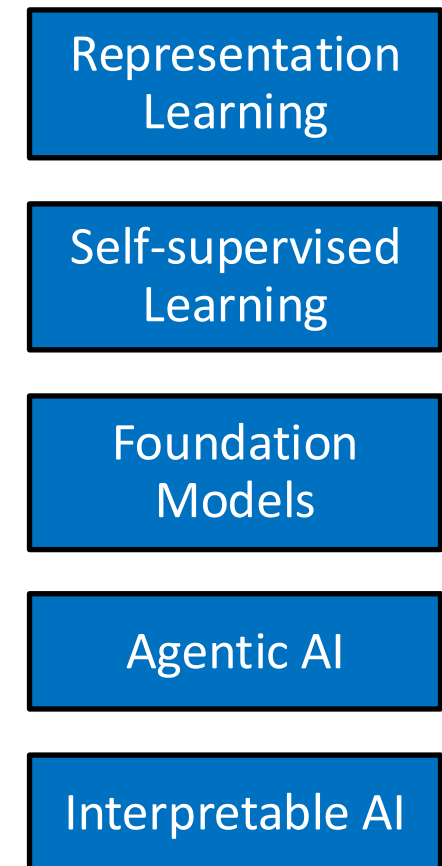
...

← **Knowledge-guided
Machine Learning
(KGML)** →



...

Building
Science of AI
inspired by scientific principles



...

Thank you!

<https://kgml-lab.github.io/>
<https://anujkarpatne.github.io/>
akarpadne@ufl.edu

- Funding Agencies:
 - NSF: CAREER Award # 2239328, HDR Grant # 2118240, HDR Grant # 1940247, Medium Grant # 2107332, MSB Grant # 2213550, EAGER Grant # 2026710, PIPP Grant # 2200045; PIPP Grant # 2412389, NAIRR Pilot Award, NEEC Grant
- Collaborators:
 - **Aquatic Sciences:** **VT:** Cayelan Carey, Quinn Thomas, Mary Thompson; **U Wisconsin:** Paul Hanson, Bennett McAfee; **Aarhus U:** Robert Ladwig
 - **Organismal Biology:** **VT:** Josef Uyeda, Leanna House; **OSU:** Tanya Berger-Wolf, Wei-Lun Chao; **Battelle:** Paula Mabee; **Tulane:** Hank Bart; **UCI:** Wasila Dahdul, **Princeton:** Dan Rubenstein, **RPI:** Chuck Stewart; **Duke:** Hilmar Lapp
 - **Mechanobiology:** **VT:** Amrinder Nain, Sohan Kale, Bahareh Behkam
 - **Solving PDEs:** **U Penn:** Paris Perdikaris
 - **Fluid Dynamics:** **VT:** Danesh Tafti, Naren Ramakrishnan
 - **Virology:** **VT:** T.M. Murali; **Cornell:** Colin Parrish; **U Michigan:** Adam Luring
 - **Quantum Mechanics and Electromagnetism:** **SUNY Binghamton:** Wei-Cheng Lee; **U Mass Lowell:** Viktor Podolskiy; **OSU:** Anish Arora
 - **Aerosol Modeling:** **VT:** Elena Lind
 - **Seismology:** **UNC and LANL:** Youzuo Lin

Thank you!

<https://kgml-lab.github.io/>
<https://anujkarpatne.github.io/>
akarpadne@ufl.edu

- Students:

- PhD Students:



Medha Sawhney



Kazi Mehrab



Mridul Khurana



Abhilash Neog



Sepideh Fatemi



Amartya Dutta



Snehal More

- MS Student:
 - Postdoc:



Sneha Aggarwal



Blessy Antony

- Alumni (PhD):



Nikhil Dhiman



Suhrid Das



Nikhil Muralidhar
(Stevens Institute)



Mohannad Elhamod
(Boston University)



Arka Daw
(Amazon AGI)



Jie Bu
(Apple)



M Maruf
(Amazon AGI)



Blessy Antony
(Virginia Tech)

- Alumni (MS):

- Aanish Pradhan, Harish Manogaran, Naveen Gupta, Sandhya Bhaskar, Ioannis Papakis, Prathamesh Mandke, Arya Shahadi, Reza Sepasdar, Zheng Li

Thank you!

References

- [1] Karpatne, A., et al. (2017). [Theory-guided data science: A new paradigm for scientific discovery from data.](#) IEEE Transactions on Knowledge and Data Engineering, 29(10), 2318–2331.

- [2] Karpatne, A., Jia, X., & Kumar, V. (2024). [Knowledge-guided Machine Learning: Current Trends and Future Prospects.](#) arXiv:2403.15989.

- [3] Karpatne, A., Deshwal, A., Jia, X., Ding, W., Steinbach, M., Zhang, A., & Kumar, V. (2025). [AI-enabled scientific revolution in the age of generative AI: second NSF workshop report.](#) npj Artificial Intelligence, 1, 18.

- [4] Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2017). [Physics-guided Neural Networks \(PGNN\): An Application in Lake Temperature Modeling.](#) arXiv:1710.11431.

- [5] Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., & Kumar, V. (2019). [Physics Guided RNNs for Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles.](#) SIAM International Conference on Data Mining (SDM), 558–566.

References

- [6] Daw, A., Thomas, R. Q., Carey, C. C., Read, J. S., Appling, A. P., & Karpatne, A. (2020). [Physics-Guided Architecture \(PGA\) of Neural Networks for Quantifying Uncertainty in Lake Temperature Modeling.](#) SDM, 532–540.
- [7] Khandelwal, A., Karpatne, A., Ravirathinam, P., Ghosh, R., Wei, Z., Dugan, H. A., Hanson, P. C., & Kumar, V. (2022). [ReaLSAT, a global dataset of reservoir and lake surface area variations.](#) Scientific Data, 9, 356.
- [8] Ladwig, R., Daw, A., Albright, E. A., Buelo, C., Karpatne, A., Meyer, M. F., Neog, A., Hanson, P. C., & Dugan, H. A. (2024). [Modular Compositional Learning Improves 1D Hydrodynamic Lake Model Performance by Merging Process-Based Modeling With Deep Learning.](#) Journal of Advances in Modeling Earth Systems, 16.
- [9] Neog, A., et al. (2026). [Investigating a Model-Agnostic and Imputation-Free Approach for Irregularly-Sampled Multivariate Time-Series Modeling.](#) Transactions on Machine Learning Research.
- [10] Neog, A., et al. (2026). [LakeFM: Toward a Foundation Model for Aquatic Ecosystems Using Irregular Multivariate Multi-depth Time Series Data.](#) KDD 2026.

References

- [11] Elhamod, M., et al. (2023). [Discovering Novel Biological Traits From Images Using Phylogeny-Guided Neural Networks.](#) ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).
- [12] Khurana, M., et al. (2024). [Hierarchical Conditioning of Diffusion Models Using Tree-of-Life for Studying Species Evolution.](#) European Conference on Computer Vision (ECCV), 137–153.
- [13] Manogaran, H. B., et al. (2025). [What Do You See in Common? Learning Hierarchical Prototypes over Tree-of-Life to Discover Evolutionary Traits.](#) International Conference on Learning Representations (ICLR).
- [14] Monsefi, A. K., Khurana, M., Ramnath, R., Karpatne, A., Chao, W.-L., & Zhang, C. (2025). [TaxaDiffusion: Progressively Trained Diffusion Model for Fine-Grained Species Generation.](#) IEEE/CVF International Conference on Computer Vision (ICCV), 8579–8589.
- [15] Khurana, M., et al. (2026). [TaxaAdapter: Vision Taxonomy Models are Key to Fine-grained Image Generation over the Tree of Life.](#) arXiv:2603.26128.

References

- [16] Antony, B., Haghani, M., Luring, A. S., Karpatne, A., & Murali, T. M. (2025). [HAVEN: Hierarchical Attention for Viral protEin-based host iNference.](#) bioRxiv.
- [17] Antony, B., et al. (2026). [VILLA: Versatile Information Retrieval From Scientific Literature Using Large LAnguage Models.](#) KDD 2026, AI Sciences Track.
- [18] Padhi, A., et al. (2025). [Deep learning reveals how cells pull, buckle, and navigate fibrous environments.](#) Proceedings of the National Academy of Sciences, 122(47), e2424047122.
- [19] Sawhney, M., et al. (2024). [Motion Enhanced Multi-Level Tracker \(MEMTrack\): A Deep Learning-Based Approach to Microrobot Tracking in Dense and Low-Contrast Environments.](#) Advanced Intelligent Systems.
- [20] Muralidhar, N., et al. (2020). [PhyNet: Physics Guided Neural Networks for Particle Drag Force Prediction in Assembly.](#) SDM, 559–567.

References

- [21] Muralidhar, N., Bu, J., Cao, Z., Raj, N., Ramakrishnan, N., Tafti, D., & Karpatne, A. (2021). [PhyFlow: Physics-Guided Deep Learning for Generating Interpretable 3D Flow Fields.](#) IEEE International Conference on Data Mining (ICDM).
- [22] Bu, J., & Karpatne, A. (2021). [Quadratic Residual Networks: A New Class of Neural Networks for Solving Forward and Inverse Problems in Physics Involving PDEs.](#) SDM, 675–683.
- [23] Daw, A., Maruf, M., & Karpatne, A. (2021). [PID-GAN: A GAN Framework based on a Physics-informed Discriminator for Uncertainty Quantification with Physics.](#) KDD.
- [24] Elhamod, M., et al. (2022). [CoPhy-PGNN: Learning Physics-guided Neural Networks with Competing Loss Functions for Solving Eigenvalue Problems.](#) ACM Transactions on Intelligent Systems and Technology.
- [25] Daw, A., Bu, J., Wang, S., Perdikaris, P., & Karpatne, A. (2023). [Mitigating Propagation Failures in Physics-informed Neural Networks using Retain-Resample-Release \(R3\) Sampling.](#) International Conference on Machine Learning (ICML).

References

- [26] Elhamod, M., & Karpatne, A. (2024). [Neuro-Visualizer: An Auto-encoder-based Loss Landscape Visualization Method.](#) ICML.
- [27] Gupta, N., Sawhney, M., Daw, A., Lin, Y., & Karpatne, A. (2025). [A Unified Framework for Forward and Inverse Problems in Subsurface Imaging using Latent Space Translations.](#) ICLR.
- [28] Lynch, S., LaMountain, J., Fan, B., Bu, J., Raju, A., Wasserman, D. L., Karpatne, A., & Podolskiy, V. A. (2025). [Physics-Guided Hierarchical Neural Networks for Maxwell's Equations in Plasmonic Metamaterials.](#) ACS Photonics.
- [29] Sawhney, M., Neog, A., Khurana, M., & Karpatne, A. (2026). [PRISMA: Improving the Accuracy-Latency Frontier of Diffusion-based PDE Solvers Using Physics-Informed Spectral Attention.](#) arXiv:2512.01370v2.
- [30] Hao, Q., Xu, F., Li, Y., & Evans, J. (2026). [Artificial intelligence tools expand scientists' impact but contract science's focus.](#) Nature, 649, 1237–1243.

References

- [31] Kutz, J. N., et al. (2025). [Accelerating scientific discovery with the common task framework](#). arXiv:2511.04001.

- [32] Cheetham, A. K., & Seshadri, R. (2024). [Artificial Intelligence Driving Materials Discovery? Perspective on the Article: Scaling Deep Learning for Materials Discovery](#). Chemistry of Materials.

- [33] Merchant, A., et al. (2023). [Scaling deep learning for materials discovery](#). Nature, 624, 80–85.

- [34] Taylor, R., et al. (2022). [Galactica: A Large Language Model for Science](#). arXiv:2211.09085.

- [35] Karniadakis, G. E., et al. (2021). [Physics-informed machine learning](#). Nature Reviews Physics.

References

- [36] Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). [Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations.](#) Journal of Computational Physics, 378, 686–707.
- [37] Lu, L., et al. (2021). [Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators.](#) Nature Machine Intelligence.
- [38] Li, Z., et al. (2021). [Fourier Neural Operator for Parametric Partial Differential Equations.](#) ICLR.
- [39] Batzner, S., et al. (2022). [E\(3\)-equivariant graph neural networks for data-efficient and accurate interatomic potentials.](#) Nature Communications.
- [40] Anderson, B., Hy, T. S., & Kondor, R. (2019). [Cormorant: Covariant Molecular Neural Networks.](#) NeurIPS.

References

- [41] Wang, R., Walters, R., & Yu, R. (2021). [Incorporating Symmetry into Deep Dynamics Models for Improved Generalization.](#) ICLR.
- [42] Shen, C., et al. (2023). [Differentiable modelling to unify machine learning and physical models for geosciences.](#) Nature Reviews Earth & Environment, 4, 552–567.
- [43] Zhao, Z., Ding, X., & Prakash, B. A. (2024). [PINNsFormer: A Transformer-Based Framework For Physics-Informed Neural Networks.](#) ICLR.
- [44] Jumper, J., et al. (2021). [Highly accurate protein structure prediction with AlphaFold.](#) Nature.
- [45] Pathak, J., et al. (2022). [FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators.](#) arXiv:2202.11214.

References

- [46] Kochkov, D., et al. (2024). [Neural general circulation models for weather and climate](#). Nature, 632, 1060–1066.
- [47] Rodríguez, A., Cui, J., Ramakrishnan, N., Adhikari, B., & Prakash, B. A. (2023). [EINNs: Epidemiologically-Informed Neural Networks](#). AAAI Conference on Artificial Intelligence.
- [48] Liu, L., et al. (2022). [KGML-ag: a modeling framework of knowledge-guided machine learning to simulate agroecosystems: a case study of estimating N₂O emission using data from mesocosm experiments](#). Geoscientific Model Development, 15, 2839–2858.
- [49] Liu, L., et al. (2024). [Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems](#). Nature Communications.
- [50] Stevens, S., et al. (2024). [BioCLIP: A Vision Foundation Model for the Tree of Life](#). CVPR, 19412–19424.

References

- [51] Huang, J., Yang, G., Wang, Z., & Park, J. J. (2024). [DiffusionPDE: Generative PDE-Solving Under Partial Observation](#). NeurIPS.
- [52] Yao, J., Mammadov, A., Berner, J., Kerrigan, G., Ye, J. C., Azizzadenesheli, K., & Anandkumar, A. (2025). [Guided Diffusion Sampling on Function Spaces with Applications to PDEs](#). NeurIPS.
- [53] Khandelwal, A., et al. (2020). [Physics Guided Machine Learning Methods for Hydrology](#). arXiv:2012.02854.
- [54] Zareian, A., Karaman, S., & Chang, S.-F. (2020). [Bridging Knowledge Graphs to Generate Scene Graphs](#). ECCV.
- [55] Rasp, S., Pritchard, M. S., & Gentine, P. (2018). [Deep learning to represent subgrid processes in climate models](#). PNAS, 115(39), 9684–9689.