Follow us
@icicleai

Creating Intelligent Cyberinfrastructure for Democratizing AI: Overview of the Activities at the NSF-AI Institute ICICLE

Talk at UCSB

by

http://icicle.ai

Dhabaleswar K. (DK) Panda
The Ohio State University
E-mail: panda@cse.ohio-state.edu
http://www.cse.ohio-state.edu/~panda

Computing has been evolving over the last three decades with multiple **phases:**

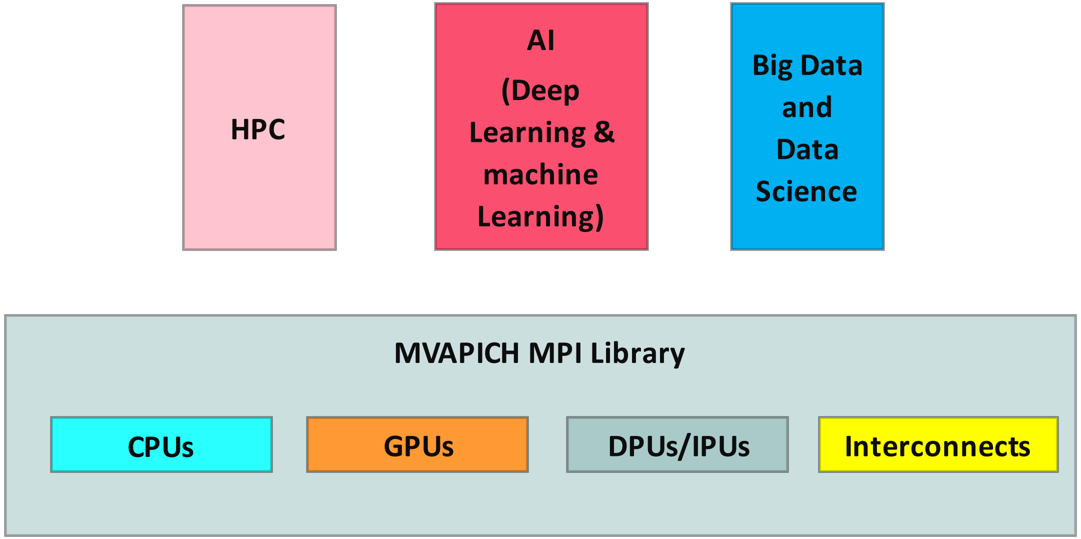
- Phase 1 (1975-): Scientific Computing/HPC
- Phase 2 (2000-): HPC + Big Data Analytics
- Phase 3: (2010-): HPC + AI (Machine Learning/Deep Learning)

Overview of the MVAPICH Project

- High Performance open-source MPI Library
- Support for multiple interconnects
  - InfiniBand, Omni-Path, Ethernet/iWARP, RDMA over Converged Ethernet (RoCE), AWS EFA, OPX, Broadcom RoCE, Intel Ethernet, Rockport Networks, Slingshot 10/11
- Support for multiple platforms
  - x86, OpenPOWER, ARM, Xeon-Phi, GPGPUs (NVIDIA and AMD)
- Started in 2001, first open-source version demonstrated at SC '02
- Supports the latest MPI-3.1 standard
- http://mvapich.cse.ohio-state.edu
- Additional optimized versions for different systems/environments:
  - MVAPICH2-X (Advanced MPI + PGAS), since 2011
  - MVAPICH2-GDR with support for NVIDIA (since 2014) and AMD (since 2020) GPUs
  - MVAPICH2-MIC with support for Intel Xeon-Phi, since 2014
  - MVAPICH2-Virt with virtualization support, since 2015
  - MVAPICH2-EA with support for Energy-Awareness, since 2015
  - MVAPICH2-Azure for Azure HPC IB instances, since 2019
  - MVAPICH2-X-AWS for AWS HPC+EFA instances, since 2019
- Tools:
  - OSU MPI Micro-Benchmarks (OMB), since 2003
  - OSU InfiniBand Network Analysis and Monitoring (INAM), since 2015

- **Used by more than 3,375 organizations in 91 countries**
- **More than 1.77 Million downloads from the OSU site directly**
- Empowering many TOP500 clusters (Nov '23 ranking)
  - 11^{th} , 10,649,600-core (Sunway TaihuLight) at NSC, Wuxi, China
  - 29^{th} , 448, 448 cores (Frontera) at TACC
  - 46^{th}, 288,288 cores (Lassen) at LLNL
  - 61^{st}, 570,020 cores (Nurion) in South Korea and many others
- Available with software stacks of many vendors and Linux Distros (RedHat, SuSE, OpenHPC, and Spack)
- Partner in the 29^{th} ranked TACC Frontera system
- **Empowering Top500 systems for more than 20 years**

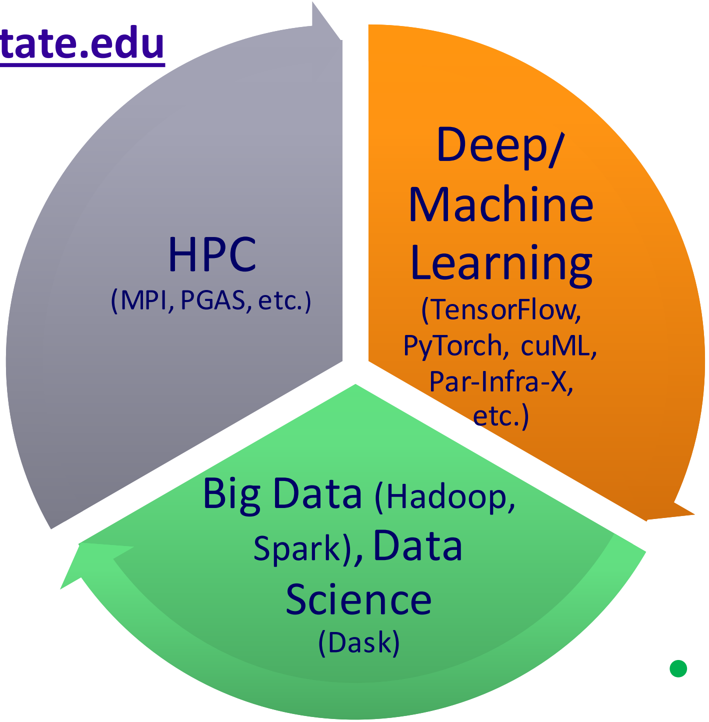
MPI (MVAPICH)-driven Converged Software Stack for HPC, AI, Big Data, and Data Science



MVAPICH-Driven Converged Software Stack for AI, Big Data and Data Science

http://mvapich.cse.ohio-state.edu

- MVAPICH



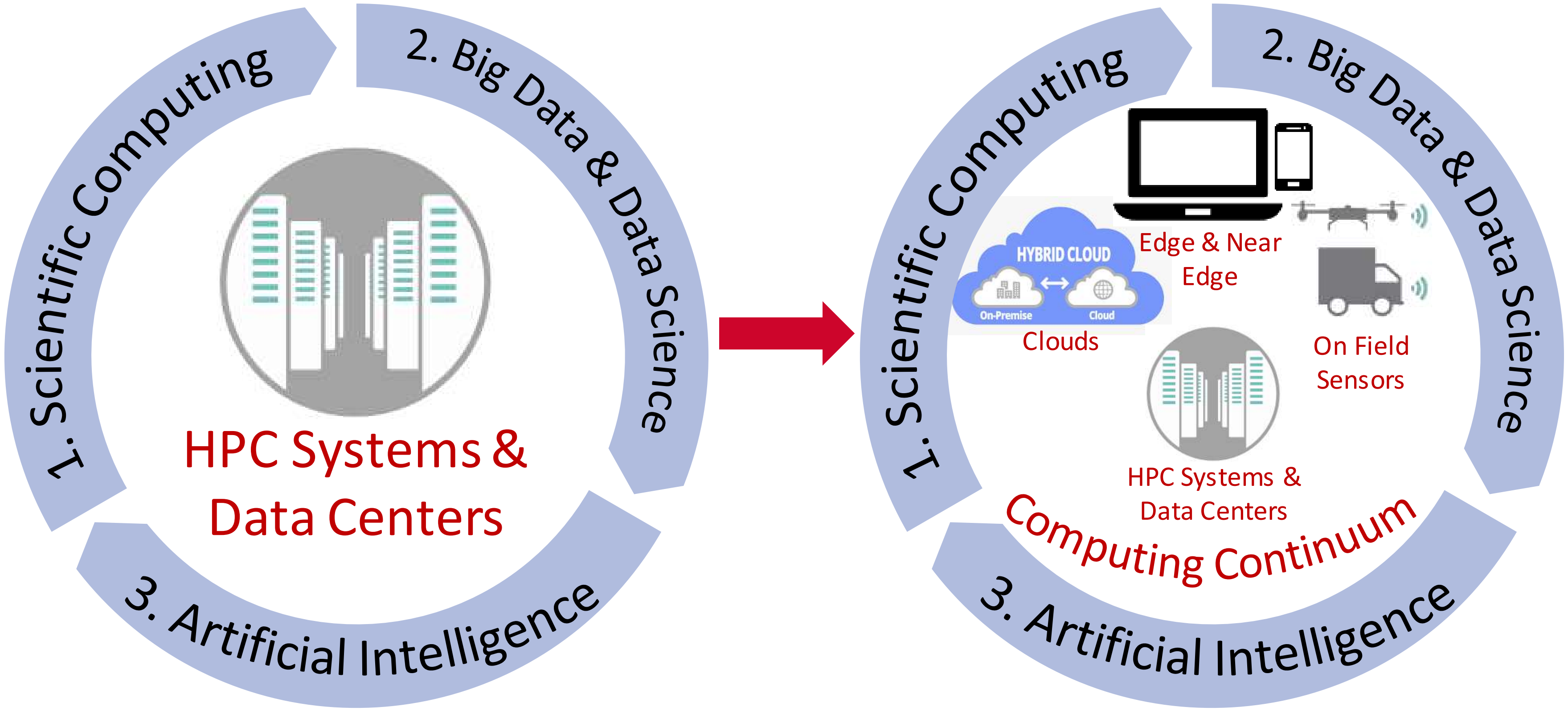
http://hidl.cse.ohio-state.edu

- MPI4DL
- MPI4cuML

http://hibd.cse.ohio-state.edu

- MPI4Spark
- MPI4Dask

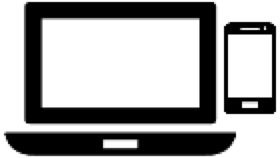
Emergence of the Computing Continuum



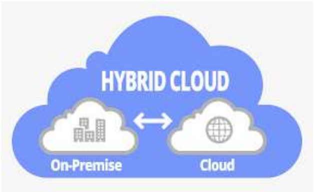
Data Movement and Control in Computing Continuum

Emerging Computing Continuum

On Field Sensors



Edge & Near Edge



Clouds

HPC Systems & Data Centers

Data Movement and Control

How to Create Intelligent Cyberinfrastructure for Computing Continuum in the Modern AI Era
&
Democratizing AI?

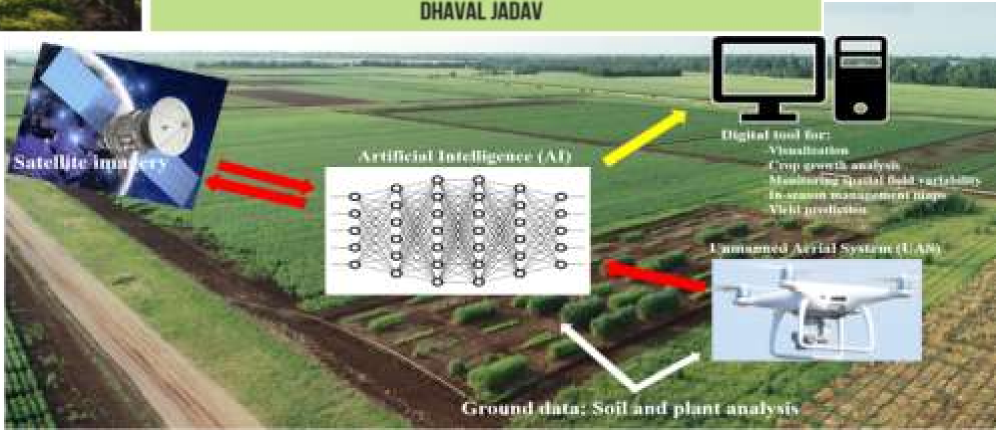
Credits to all ICICLE Team Members!!

ICICLE Members Attending All-Hands-Meeting In-Person (Nov '23)

Outline

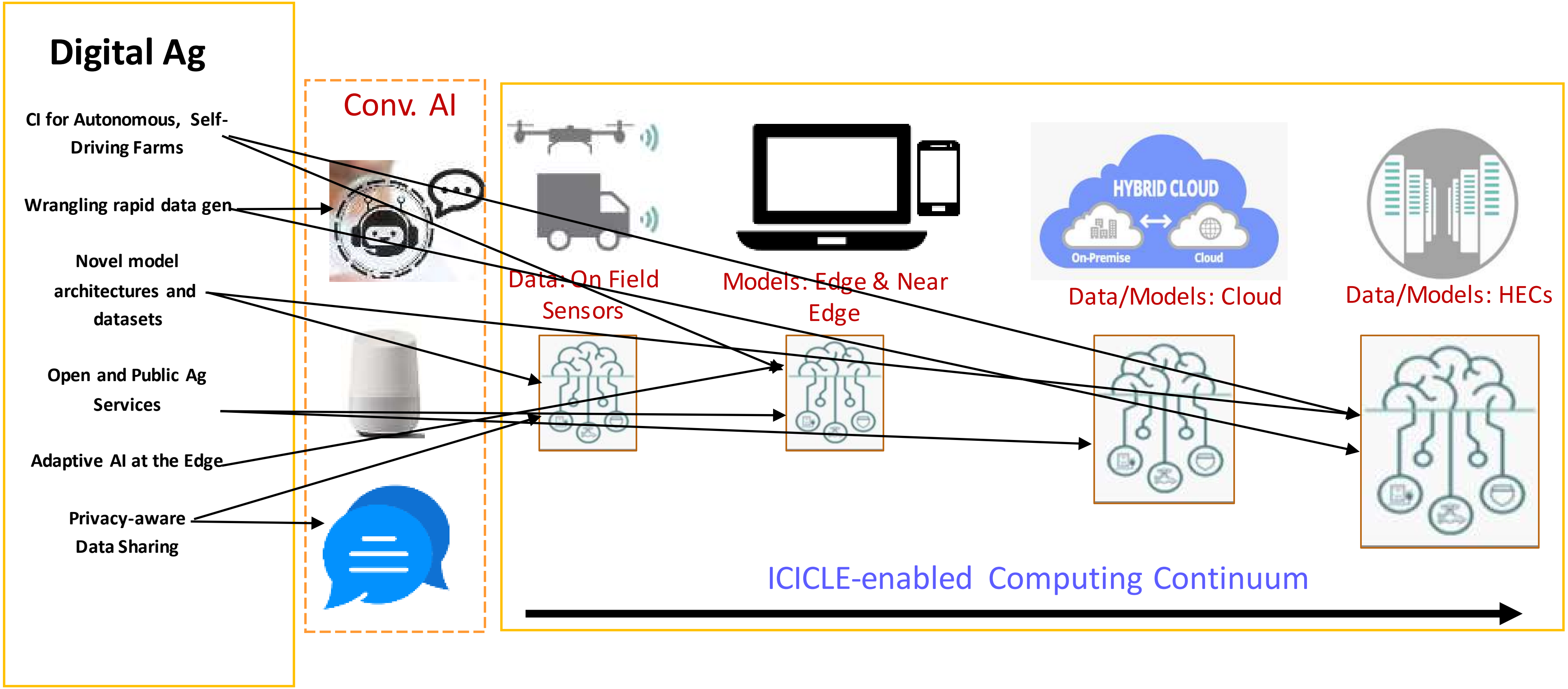
- ICICLE Vision and Goals
- Research Challenges Addressed
- Highlights of Selected Accomplishments
- How to Get Engaged?
- Conclusions

AI-Driven Digital Agriculture



https://ccag.tamu.edu/research-project/digital-agriculture/ https://medium.datadriveninvestor.com/artificial-intelligence-in-agriculture-62f71f8f6ae6

Challenges in Designing AI-Driven CI for Digital Agriculture in Computing Continuum



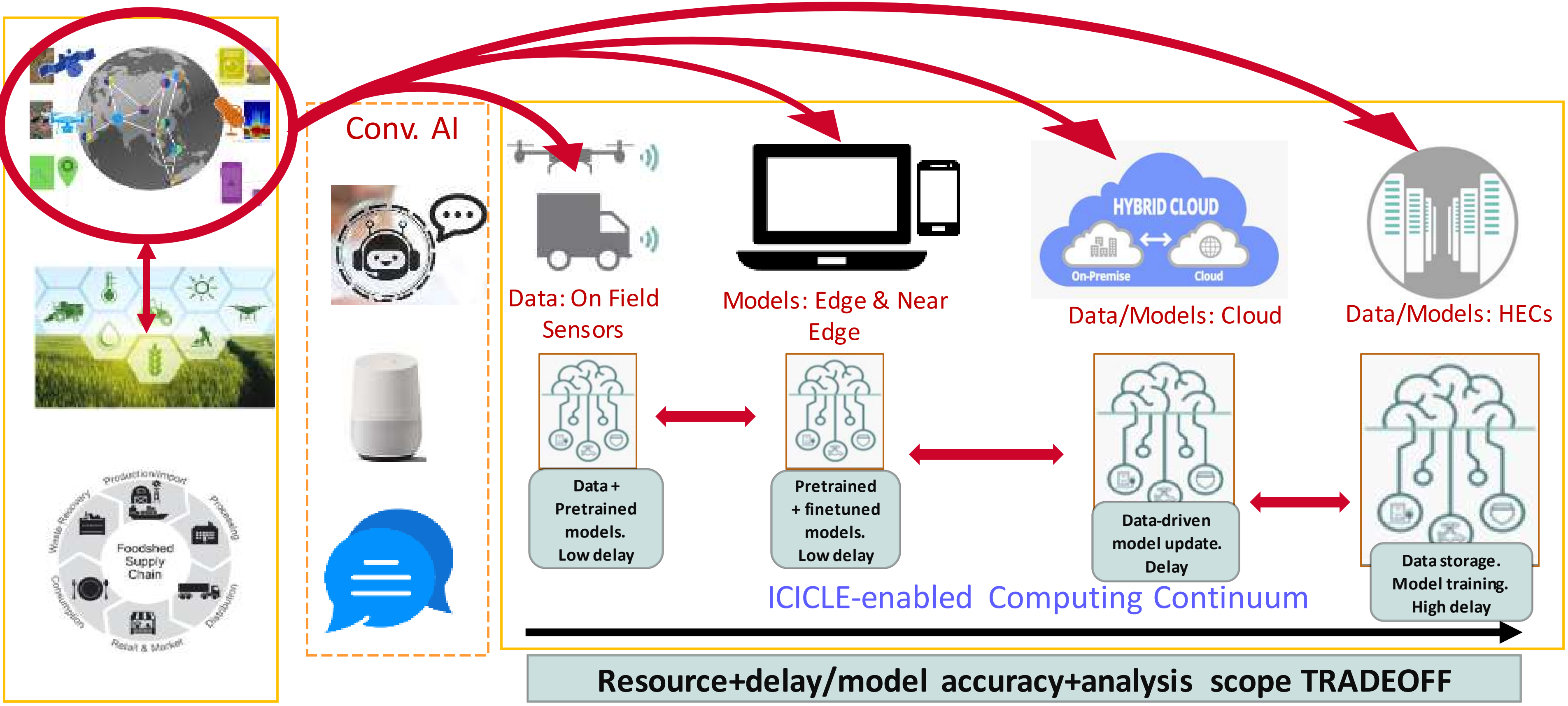
Societal Challenge (Example #2): Animal Ecology

- **Basic science:** The focus of Animal Ecology is understanding the functioning and behavior of animals individually and in groups *in the context of environment* and evolution.

- **Science + translational:**
  - Monitoring, understanding, and protecting biodiversity of the planet
  - Monitoring and understanding the impact of changing habitats on animals that live in them

- **Translational:** biodiversity conservation and mitigating the impact of climate change

Challenges in Designing AI-Driven CI for Animal Ecology in Computing Continuum



Societal Challenge (Example #3): Smart Foodsheds

- **Food Supply Chain Vulnerabilities**
  - Concentration contributes to bottlenecks, lack of resilience to disruptions
- **High Food Insecurity**
  - Supply chain decisions fail to account for impacts on food access, cost, availability
- **Food Waste**
  - Inefficiencies in food supply chains and food systems lead to 30-40% waste
- **Negative Environmental Footprint**
  - Farming and food system has major impacts on environment
- **Holistic Food Systems Planning is Difficult**
  - Data is difficult to access, not coordinated across sectors or food supply chain actors

Challenges in Designing AI-Driven CI for Smart Foodsheds in Computing Continuum

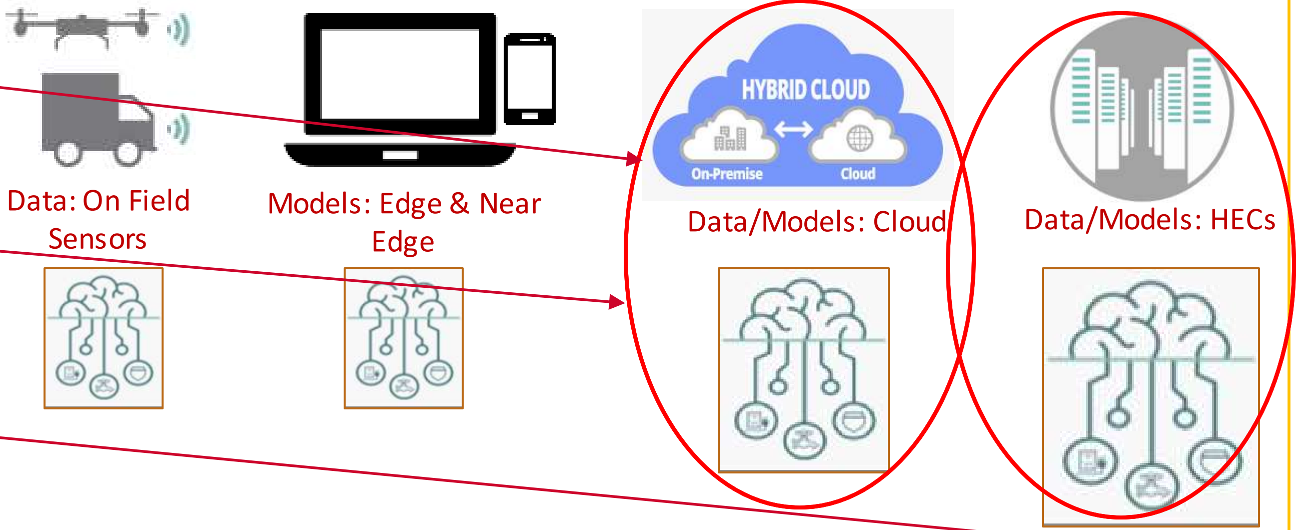
Smart Foodsheds

Diverse data sources **database federation**

Heterogenous food systems and actors
Conversation agents and knowledge graph

Food system planning
Easy-to-use intelligent visual analytics

Conv. AI



ICICLE-enabled Computing Continuum

Many more examples

- Smart Cities
- Smart Manufacturing
- Smart Transportation
- Real-time Surveillance
- Computational Medicine (Pathology, Radiology, ..)

Broad Challenge

Designing the next-generation intelligent cyberinfrastructure for a computing continuum with heterogenous resources that is usable in a plug-and-play manner by stakeholders to solve societal challenges?

Video

https://www.youtube.com/watch?v=gNFk5tDTtoU

Objectives: Intelligent CyberInfrastructure for Computing Continuum

Use Inspired Science Domains

Digital Agriculture

Smart Foodsheds

Animal Ecology

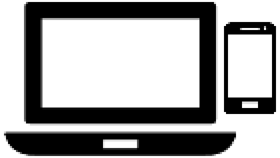
ICICLE: Intelligent CyberInfrastructure with Computational Learning in the Environment

Systems AI Foundational Research for CI

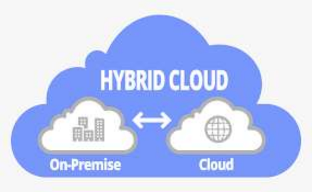
Intelligent Cyber Infrastructure

CI for AI | **AI for "CI for AI"**

On Field Sensors



Edge & Near Edge



Clouds

HPC Systems & Data Centers

Emerging Computing Continuum

Participation:
14 Organizations, 33 faculty, 41 staff, (58 PhD, 16 MS, 16 undergrad, 6 K-12) students & many Collaborators

Govt. Agencies & National Labs

Lawrence Livermore National Laboratory

BROOKHAVEN NATIONAL LABORATORY

Fermilab

Research Institutes

MOLSSI

SGCI Science Gateways Community Institute

TRUSTED CI

ECEP

CAHSI

SC/EC

neon

Jetstream

International

University of Stuttgart Germany

Industry

CNH INDUSTRIAL

Microsoft

FARM 2 FACTS

THE FINDINGS GROUP, LLC

NSF AI Institutes

AIFS

Hospitals & Universities

Nationwide Children's

Collaboration: ICICLE and the Technology Innovation Hub (TIH) at the Indian Institute of Technology Bombay (IIT-B), India

Digital Agriculture

This research collaboration will contribute novel design paradigms for context-adaptive CI and aims to develop next-generation CI for *Digital Agriculture* including AI and machine learning methods targeting 3 core areas.

Crop Health Modeling

- Sense crop health and level context to predict crop yield

- Detect stressors and diseases for geographically diverse crops

- Apply remedies with little human intervention via Internet of Things (IoT) and sensor systems

Aerial Crop Scouting

- CI for fully autonomous aerial systems

- Simplify deployment of UAV in real fields to capture common crop health conditions

- Provide accurate maps that yield valuable insights for crop management

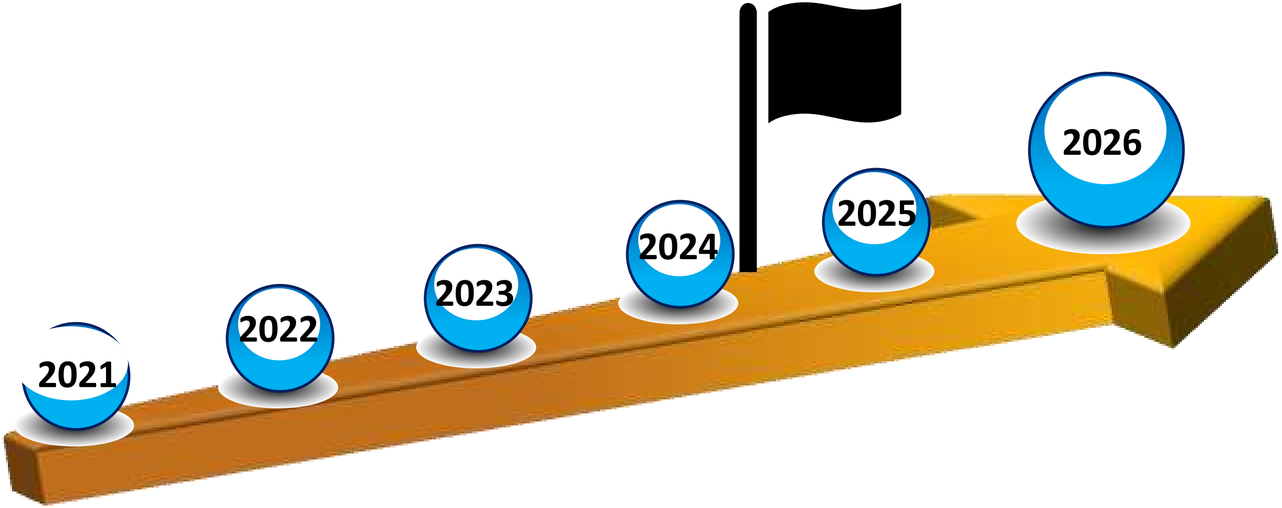
Privacy-Preserving Data Exchange

Create secure, trustworthy, and privacy-preserving platforms that connect farmers and allow them to share information and resources safely.

Building upon the existing ICICLE infrastructure, CI and AI capabilities, researchers will leverage contextual conditions in India for *Digital Agriculture* that differ from the United States to (1) expose brittle CI components, (2) make AI4CI more robust and expansive in the long-term, (3) devise principles that yield context-aware CI

Timeline

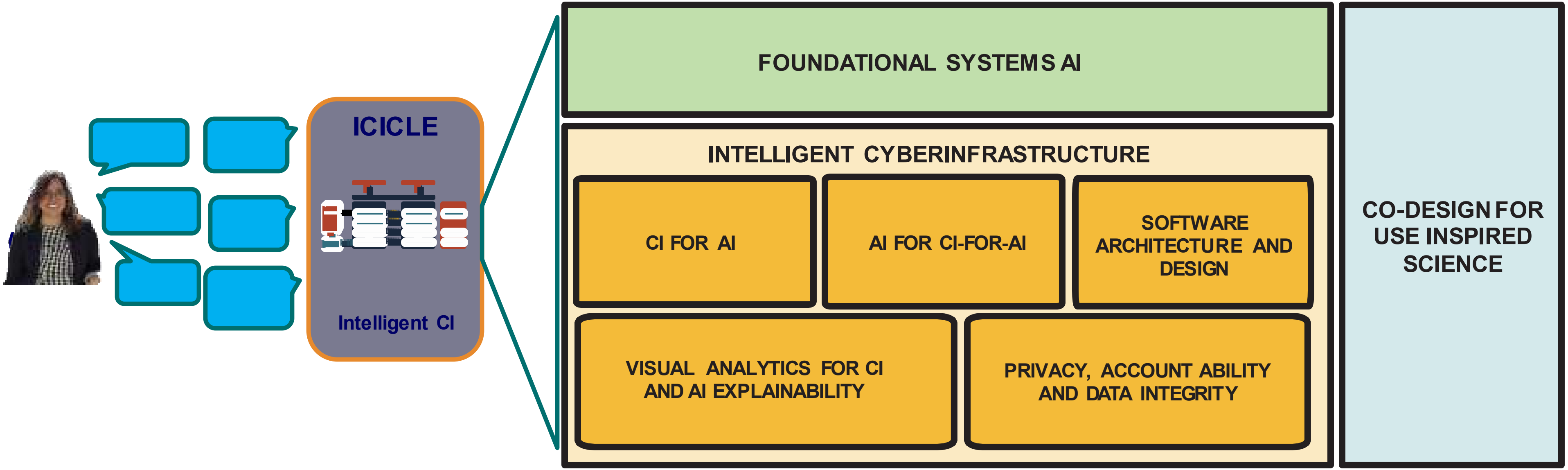
- Started on Nov 1, 2021
- Finished 29 months of the project



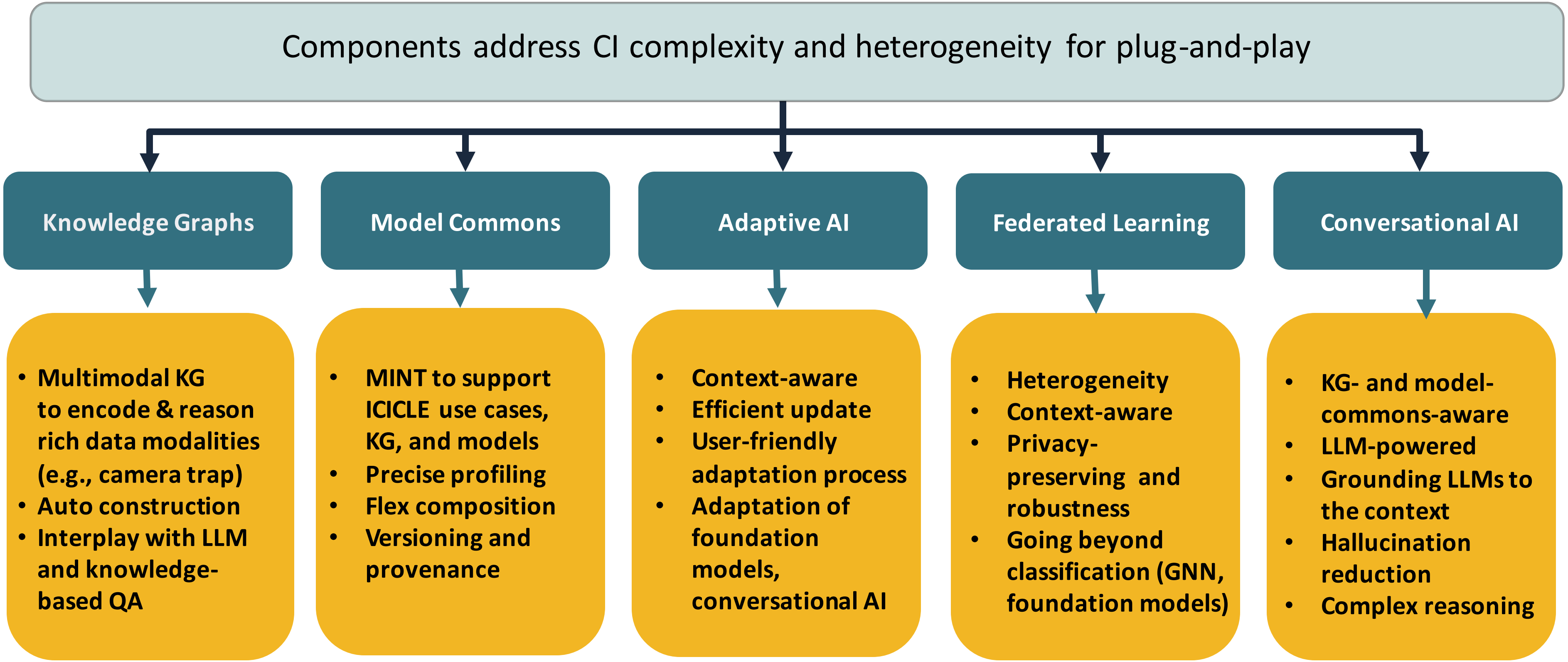
Outline

- ICICLE Vision and Goals
- **Research Challenges being Addressed**
- Selected Accomplishment Highlights
- How to Get Engaged?
- Conclusions

Research Plan: Overall Vision



Thrust: Foundational Systems AI



Thrust: CI4AI

Provides necessary CI to deploy AI throughout computing continuum and make it plug-and-play!

High Perf. Training

- High-performance communication libraries
- Gradient sparsification
- Exploiting data-, model-, pipeline-, and hybrid-parallel paradigms

High Perf. Data Management

- Unified storage of data, model and hyperparameters
- Data location transparency with migration
- Leveraging new hardware

Edge Intelligence

- Performance characterization of edge
- Optimize ML/DL inference on edge devices
- Profiling edge devices to improve quality of service

AI-Adaptive Edge Wireless

- High-throughput, reliable communications
- Predictable Wireless Comm. via Rateless-Coding & Multi-Modal/Path
- AI-adaptive edge wireless prototypes.

Control and Coordination

- Functional/Performance Interface Design
- Intelligent Resource Management with Tapis
- Hardening and Optimizing for Production-ready Service

Thrust: AI4CI

Enhances CI with AI for adaptive and field-optimized machine learning!

KGs & Model Commons for CI

- Investigate and survey existing datasets for CI optimization
- Create new CI components to serve CI data and models for other ICICLE CI and AI4CI components
- Edge-specific CI dataset distributed as KGs

Intelligent Modeling and Optimization

- Collection of baseline performance
- Exploration of analytical metrics
- Use and refinement of hybrid models in a design-space explorer for code optimization

Applications

- Application Selection and Performance Profiling
- Building Performance Models
- Designing Features for Applications, Frameworks, and Hardware

Middleware

- Develop a set of intelligent linear algebra kernels for sparse-matrix operations
- Leverage data sparsity in all computational kernels.
- Cross-layer Optimizations

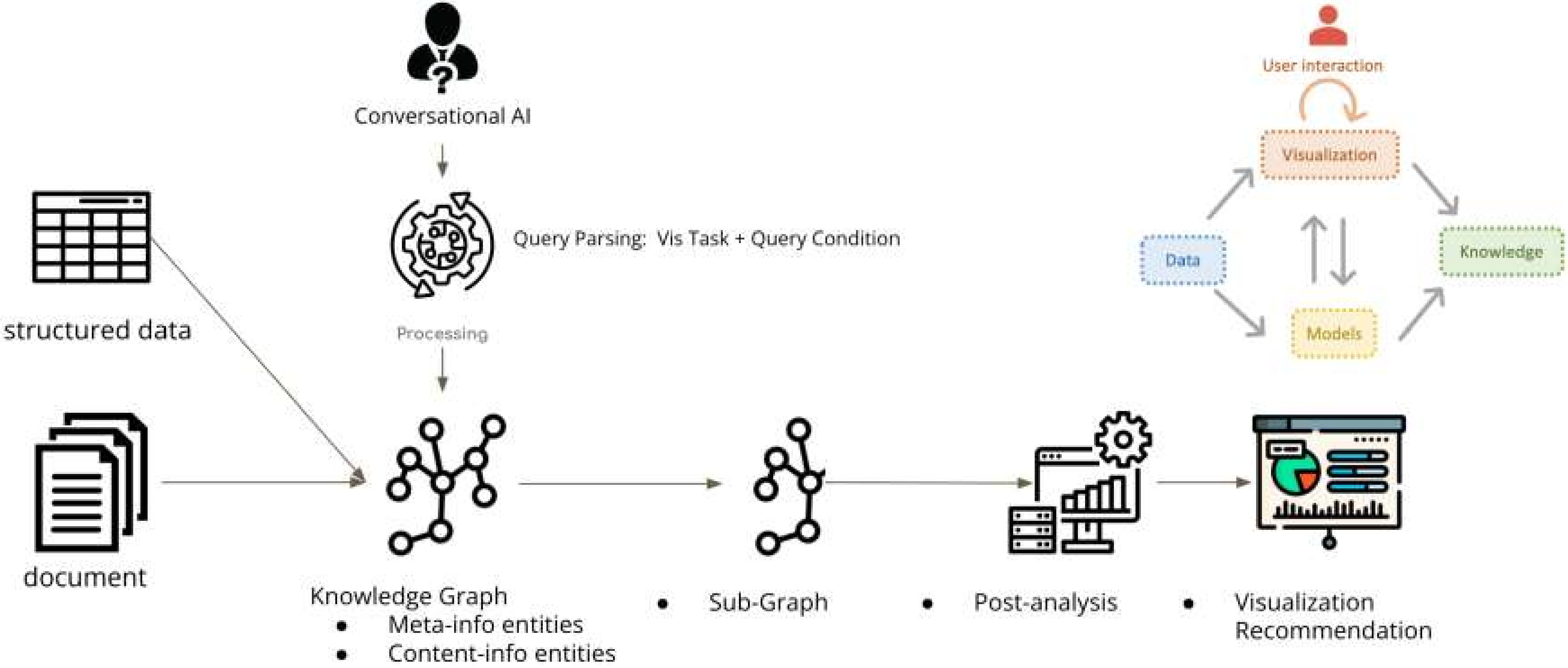
Systems

- Resource allocation optimizer for ML training
- Develop an optimizing middleware for ML inference placement based on our use cases.
- Intelligent Wireless Communications

Thrust: Privacy, Accountability and Data Integrity (PADI)

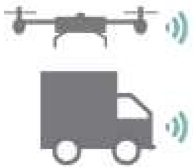
- PADI contributes to
  - ICICLE vision as *transparent and trustworthy* infrastructure for AI-enabled future
  - An ethically aligned infrastructure and workforce through an *AI ethics framework*
- PADI advances both technical and non-technical innovations and best practices that collectively contribute to a trusted environment
  - e.g., where stakeholders (farmers, industry partners, etc.) are comfortable contributing data and AI models for ICICLE AI research (and more broadly for AI research).
- PADI addresses both research questions and issues of practice (project norms and practice)

Thrust: Visual Analytics

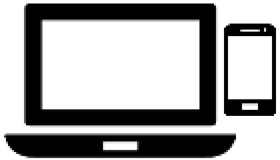


Co-Designing with use-inspired domains

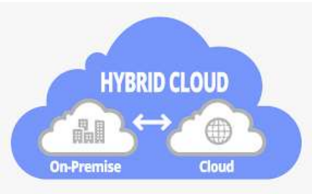
Conv. AI



Data: On Field Sensors

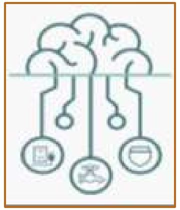


Models: Edge & Near Edge



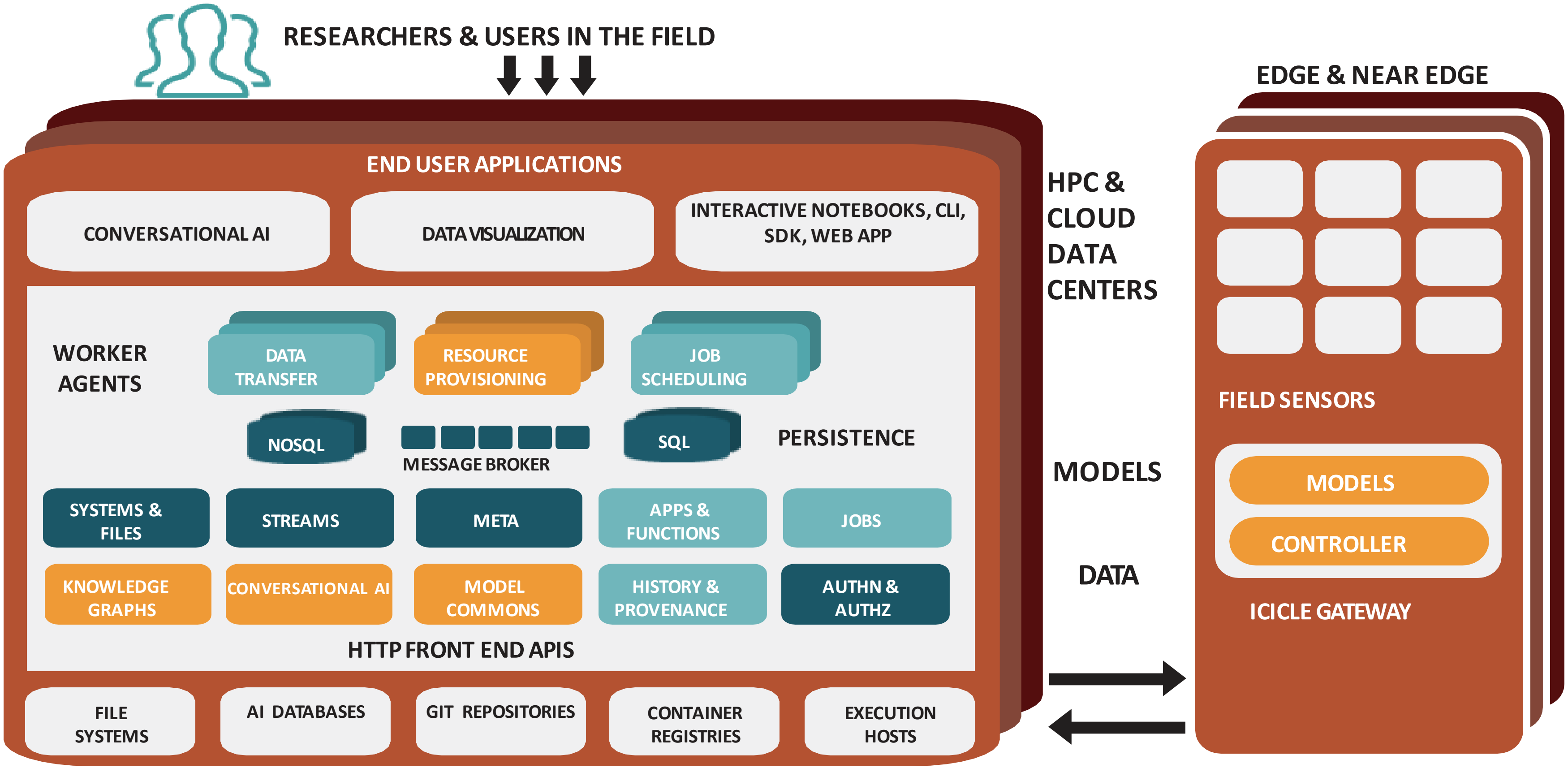
Data/Models: Cloud

Data/Models: HECs



ICICLE-enabled Computing Continuum

The Deliverable: The ICICLE Software Stack



Broader Impacts Backbone Network (BIBN)

BIBN is a consortium with the goal of democratizing AI!

Oversees activities towards broader impacts and engagement:

- Diversity Equity and Inclusion (DEI)
- Broaden Participation in Computing (BPC)
- Workforce Development (WFD)
- Collaboration and Knowledge Transfer (CKT)

Outline

- ICICLE Vision and Goals
- Research Challenges being Addressed
- **Selected Accomplishment Highlights**
    - **CI/Software Released**
    - Digital Agriculture
    - Animal Ecology
    - Smart Foodsheds
    - Intelligent Scaling and Scheduling
- How to Get Engaged?
- Conclusions

CI/Software Components Released (so far)

2023.04 Release (04/30/23)

- **AI4CI**
  - HPC Application Runtime Predictor (HARP) v1.0
  - Intelligent Sparse Library (iSpLib) v1.0
- **Software and Reference Architecture**
  - Base ICICLE Tapis Software v1.3.0
  - Event Engine v0.2.0
  - Hello ICICLE Authentication Clients v0.0.1
  - Tapis Pods Service v1.3.0
  - CI Components Catalog v0.1.0
- **Animal Ecology**
  - Camera-Traps Edge Simulator v0.3.0
- **Digital Agriculture**
  - SoftwarePilot v1.2.5
- **Smart Foodsheds**
  - Persons-Projects-Organizations-Datasets (PPOD) Schema v0.9.1
  - Smart Foodsheds Visual Analytics (VA) Dashboard v0.1

https://icicle.osu.edu/cyberinfrastructure/software

2023.06 Release (06/30/23)

- **AI Foundations**
  - ICICLE Foodshed Parser v0.1
  - Species Classification using Multimodal Heterogeneous Context v0.1.0
  - Region2vec v1.0
- **Software and Reference Architecture**
  - Tapis Federated Authentication Service v1.3.4
  - ICICONSOLE v0.0.10
  - TapisCL-ICICLE v0.1.4
  - Tapis Pods Service v1.3.2
- **Animal Ecology**
  - Camera-Traps Edge Simulator v0.3.1
- **Digital Agriculture**
  - ICICLE Digital Agriculture Hub v1.0
  - Far-Edge Edge Simulator v1.0
  - In-Field Helper for Crop Scouts v1.0
- **Smart Foodsheds**
  - Persons-Projects-Organizations-Datasets_California (PPOD_CA) Knowledge Graph v23.06
  - Kroger Store Closure v0.1
  - Smart Foodsheds Visual Analytics (VA) Dashboard v0.2

CI/Software Components Released (so far)

2023.10 Release (10/06/23)

- AI for CI-for-AI
  - High Performance Computing Applications Dataset v1.0
  - HPC Application Runtime Predictor (HARP) v2.0
- Software and Reference Architecture
  - iciflaskn v1.0
  - TapisCL-ICICLE v1.0.11
  - ICICONSOLE v0.8.0
- Animal Ecology
  - Camera-Traps Edge Simulator v0.3.2
- Smart Foodsheds
  - Smart Foodsheds Visual Analytics (VA) Dashboard v0.3

2024.01 Release (01/26/24)

- AI Foundations
  - Iluvatar Functions as a Service (FaaS) Control Plane v1.0.0
- Software Architecture and Design
  - Tapis Federated Authentication Service v1.5.0
  - Tapis Pods Service v1.5.3

https://icicle.osu.edu/cyberinfrastructure/software

Outline

- ICICLE Vision and Goals
- Research Challenges being Addressed
- **Selected Accomplishment Highlights**
  - CI/Software Released
  - **Digital Agriculture**
  - Animal Ecology
  - Smart Foodsheds
  - Intelligent Scaling and Scheduling
- How to Get Engaged?
- Conclusions

Digital Agriculture

What does CI for digital agriculture look like?

How to build CI that connects a wide range of digital agriculture stakeholders?

Why use-inspired CI will be transformative?

ICICLE Use-Inspired Science: Digital Agriculture

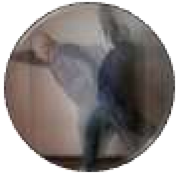
Scott Shearer
Food, Agriculture and Biological Eng.

Christopher Stewart
Computer Science & Eng

Co-Leads

Zichen Zhang

Jenna Kline



John C. Chumley

Kevyn Angueira Irrizary

Ohio State University

Digital Agriculture Hub and Use-Inspired Technologies

P. Sadayappan
University of Utah

Jinghua Yan
University of Utah

Artificial Intelligence for Cyberinfrastructure

Hari Subramoni

Nawras Alnaasan

Cyberinfrastructure for AI-Driven Digital Agriculture

Erman Ayday
Case Western

Beth Plale
Indiana University

Privacy-aware, Explainable AI, & Democratization

Alfonso Morales
University of Wisconsin

Stakeholder Engagement

Multiple Challenges

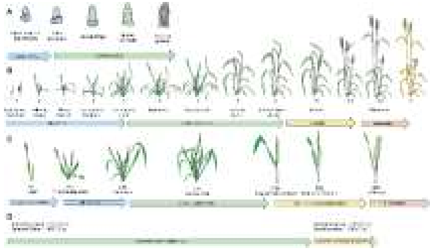
- **Application Domain**
- **Data Labeling**
- **Distributed Training with Semi-Supervised Learning**
- Quantization on Edge Devices
- Aerial Crop Scouting
- End-to-end CI

The Application Domain Challenge (Digital Agriculture)

Computer Vision (CV) based classification scenarios are ubiquitous in use-inspired science domains such as Digital Agriculture

Residue Cover on Soil Surface

Soil Aggregate Size



Wheat Development

Non-Uniform Emergence

Nitrogen Deficiency

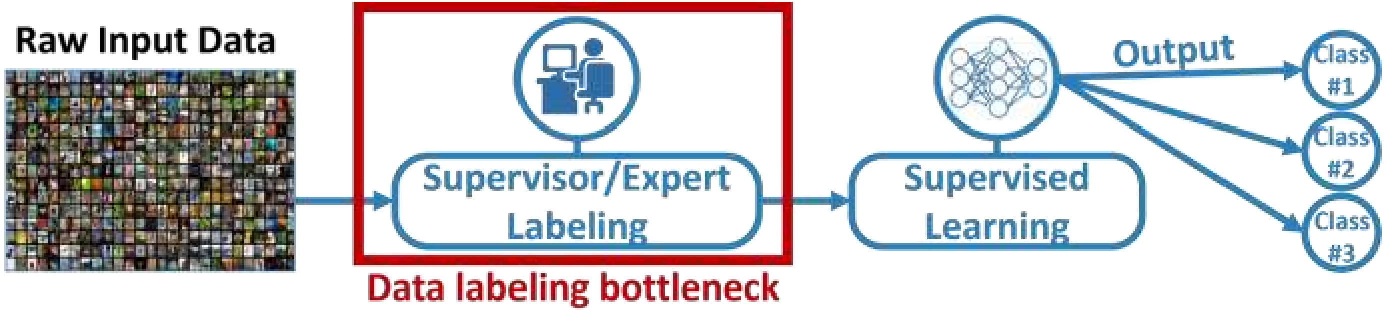
European corn borer

Corn leaf aphid

Mexican bean beetle defoliation

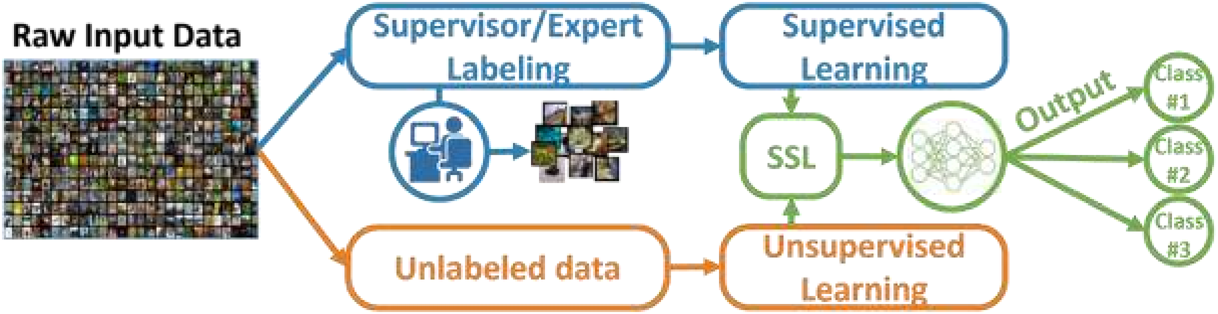
The Data Labeling Challenge

Challenge:



- Data samples need to be fully labeled by an expert for training and evaluation.
- Datasets may be collected frequently and in large volumes (millions of unlabeled images).
- Labeling data by experts is a significant bottleneck.
- Supervised learning can be time-consuming, costly, and infeasible for certain applications

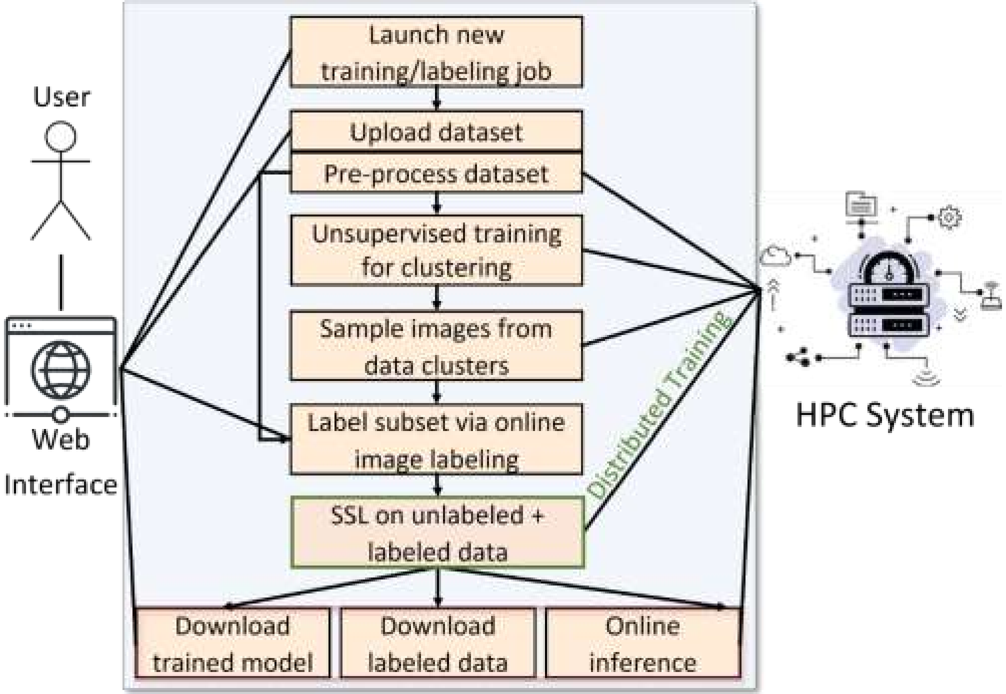
Semi-Supervised Learning (SSL) for Digital Agriculture



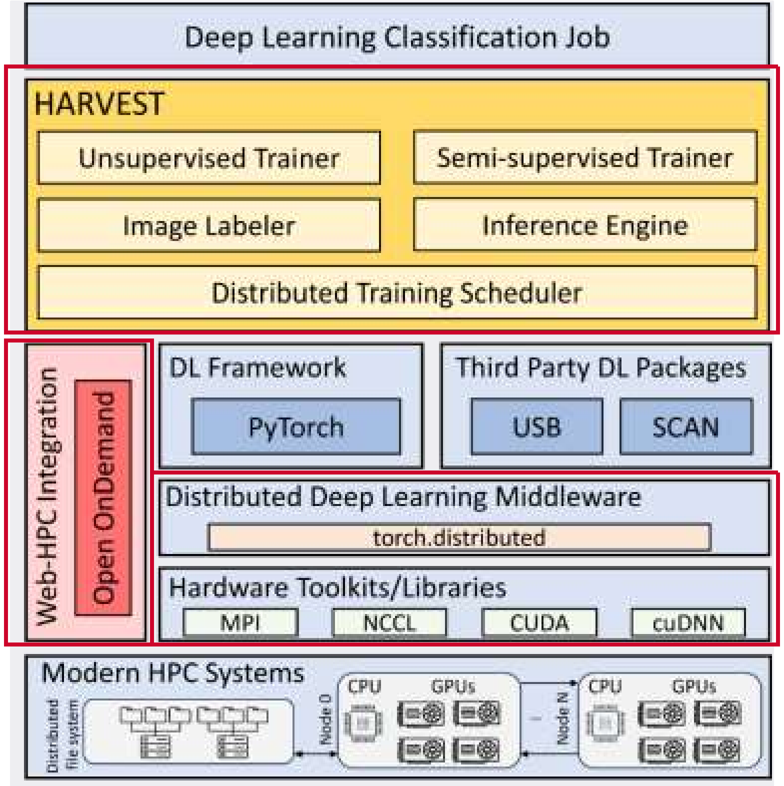
- Only requires a subset of the training dataset to be labeled (less than 1% or few hundreds).
- Achieves high accuracies by training on the rest of the unlabeled data.

HARVEST (High-Performance ARtificial Vision Framework for Expert Labeling using Semi-Supervised Training)

- Design a workflow for domain experts with no prior DL or HPC experience.
- Employ state-of-the-art SSL solutions for computer vision applications.
- Train accurate DL models using only a small fraction of labeled data.
- Accelerate training using distributed training on HPC systems.
- Enable an intuitive and user-friendly interface linked to HPC systems.
- Support any user-defined use case.
- We plan to release HARVEST a service that can be deployed on Cloud/HPC systems.

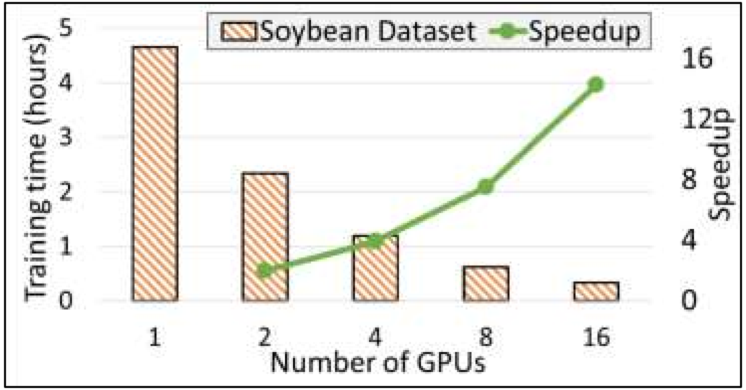


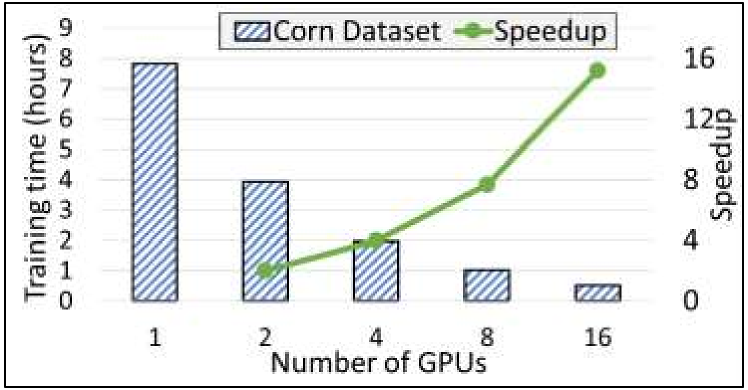
HARVEST Architecture Overview



HARVEST: Evaluation on Digital Agriculture Use Cases

- Use case: Plant stress identification for protecting crops through the growing season.
- Datasets: 1) Corn crops (9558 samples, 12 classes)* 2) Soybean crops (5636 samples, 6 Classes)*
- Achieved 97% and 93% accuracies for the Corn and Soybean datasets using only 80 labeled samples per class.
- Accelerated the training by 15.19x on 16 NVIDIA A100 GPUs reducing the training time from 7.8 hours to 31 minutes.





Dataset	Accuracy	Precision	Recall	F1 Score
Corn Dataset	97.08%	91.77%	95.43%	92.61%
Soybean Dataset	93.07%	88.64%	92.40%	89.61%

N. Alnaasan, M. Lieber, A. Shafi, H. Subramoni, S. Shearer, and DK Panda, "HARVEST: High-Performance Artificial Vision Framework for Expert Labeling using Semi-Supervised Training", 2023 IEEE International Conference on Big Data, December 2023

Demo: Semi-Supervised Learning

https://youtu.be/EYzAZWGvyJI

Multiple Challenges

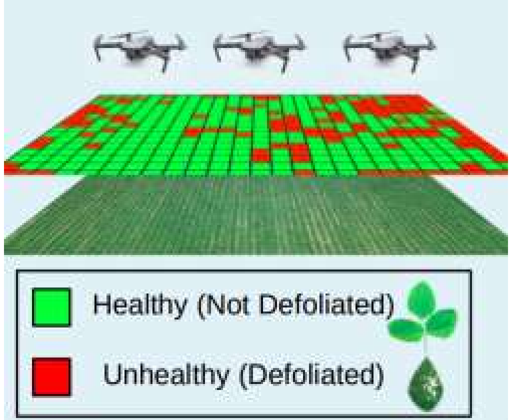
- Application Domain
- Data Labeling
- Distributed Training with Semi-Supervised Learning
- **Quantization on Edge Devices**
- **Aerial Crop Scouting**
- **End-to-end CI**

Quantization on Edge Devices

- Edge devices have limited computing power
- Explore the use of various quantization techniques – based on INT8/FP16 and static/dynamic strategies – on a range of DL inference frameworks, including OpenVINO, PyTorch, TFLite, and ONNX.
- The performance evaluation is done on Intel CPUs (Cascade Lake and Skylake) and a Raspberry Pi 4B equipped with an ARM processor.
- The characterization study uses a range of popular DL models – including MobileNetV2, VGG-19, and DenseNet-121. We found that OpenVINO and TFLite are the most optimized frameworks for Intel CPUs and Raspberry Pi 4B device, respectively.
- The performance characterization reveals that the size of original models is reduced by a quarter for INT8-based models without losing accuracy except the slight accuracy reduction of static quantization.

H. Ahn, T. Chen, N. Alnaasan, A. Shafi, M. Abduljabbar, H. Subramoni, and DK Panda,
Performance Characterization of using Quantization for DNN Inference on Edge Devices ,
7th IEEE International Conference on Fog and Edge Computing, May 2023

Digital Agriculture: Aerial Crop Scouting



- **Aerial Crop Scouting:** In this workload, we seek to create *heat maps* that describe crop health for a field
    - *Inform* self-driving tractors and sprayers to reduce the application of pesticide and fertilizer
    - *Predict* crop yields for harvest and market timing
    - *Identify* trends across farms, such as the introduction of resistant weeds
- **Technology:** Unmanned aerial vehicles (UAV) capture high resolution images
    - Flying low (15 ft above ground): 1 pixel -> mm
- **Transformative:** At mm-granularity, stakeholders can detect biological phenomena invisible to satellites
    - Soybean leaf defoliation caused by Japanese beetle
- **Software Pilot** (https://pypi.org/project/SoftwarePilot/)
- **OpenPass** (http://149.165.155.188:2298/)

Courtesy of LaRue Farms Inc.

Demo: Cloud-to-Edge Middleware

https://youtu.be/M6o0NVQXny0

Goal: Towards Designing End-to-end Digital Agriculture CI Solutions and make these available as Services for various Stakeholders

Use-Inspired Cyberinfrastructure

Crop Health Issues

HPC-Enabled Data Harvesting & Preparation

HPC for Ag AI

From HPC to the field: Computing Continuum

Crop Management Interventions

End-to-end software workflows that democratize access, use, and benefits of AI-driven digital agriculture

Outline

- ICICLE Vision and Goals
- Research Challenges being Addressed
- **Selected Accomplishment Highlights**
    - CI/Software Released
    - Digital Agriculture
    - **Animal Ecology**
    - Smart Foodsheds
    - Intelligent Scaling and Scheduling
- How to Get Engaged?
- Conclusions

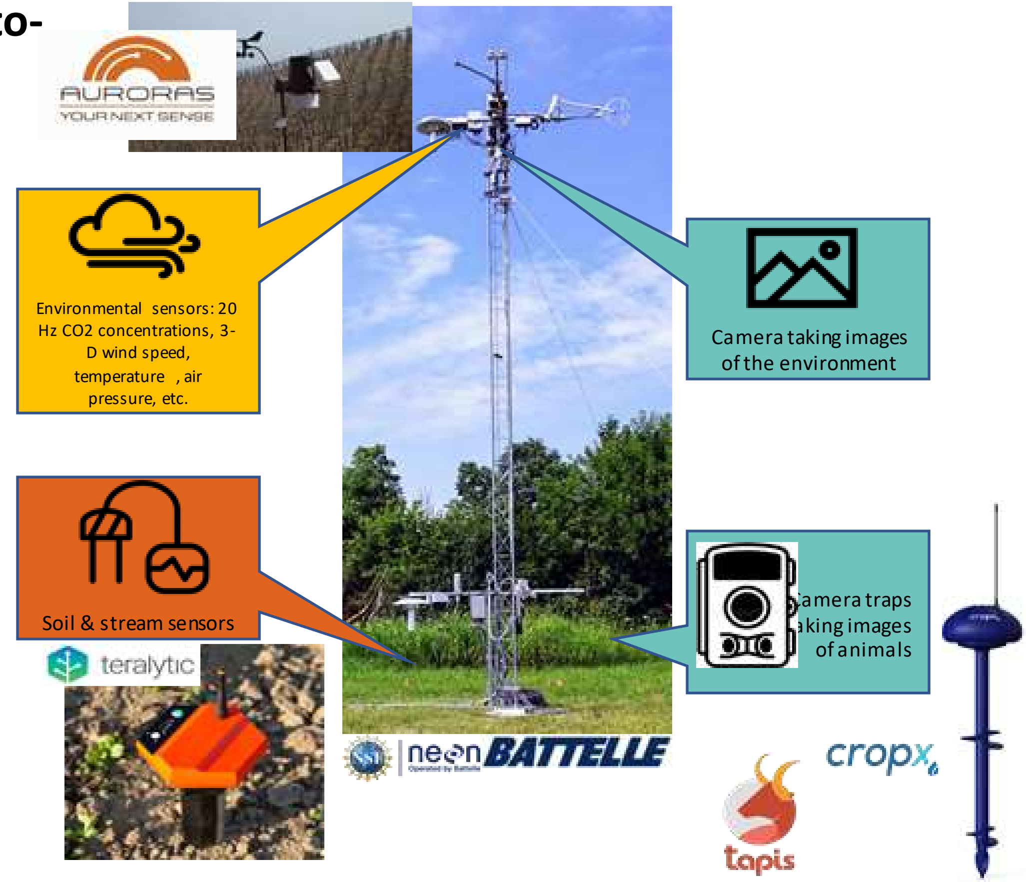
Edge-to-Center ML for Camera Traps

Optimized ML Architecture and Edge-to-Center Infrastructure for:

- Species detection in images taken in wildlife habitats
- Little to no network bandwidth
- Power conservation
- CPU and Memory constraints

Techniques:

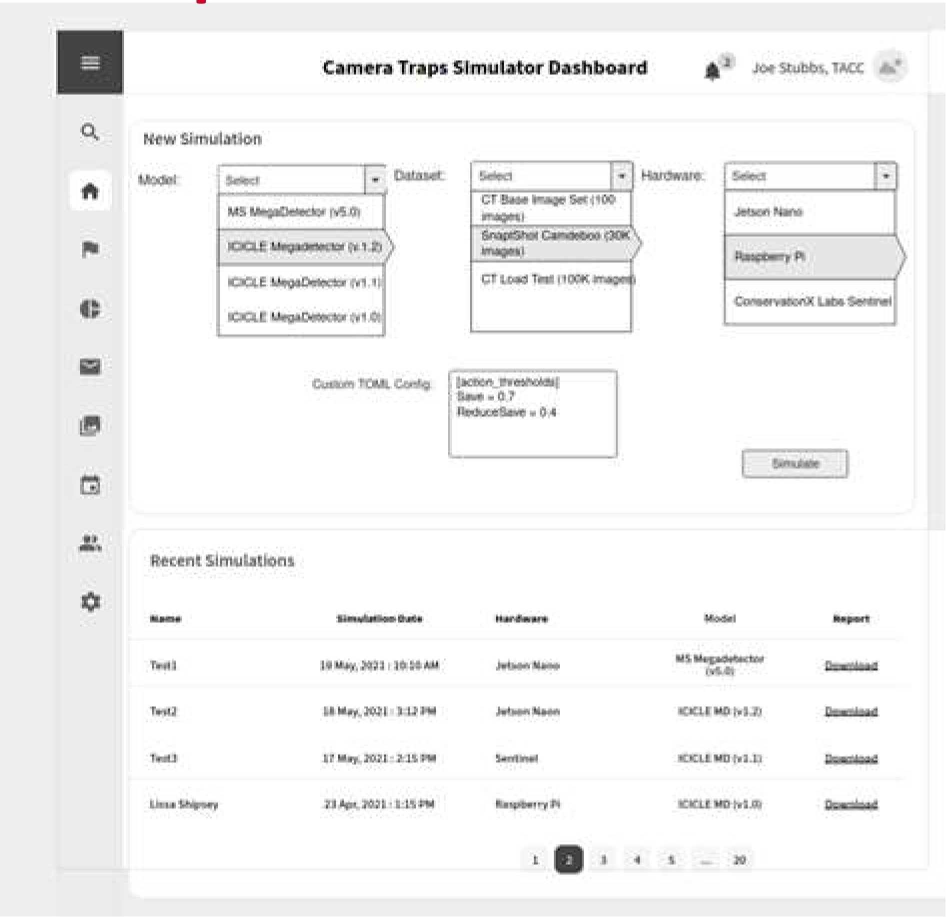
- Computer Vision
- Adaptive AI
- OOD Detection
- Continuous Learning
- Neural Architecture Search
- MLOps for the Computing Continuum



Edge-to-Center ML for Camera Traps

Camera Traps Y3 Objectives:

- Automatically deploy Camera Traps software across dynamically provisioned edge hardware
- Study performance of different model architectures
- Understand tradeoffs between computation at the edge and the center
- Adapt AI models to new environments: novel species detection, environmental shifts, etc.
- Compare hardware platforms and plan capacity requirements for field deployments



Outline

- ICICLE Vision and Goals
- Research Challenges being Addressed
- **Selected Accomplishment Highlights**
  - CI/Software Released
  - Digital Agriculture
  - Animal Ecology
  - **Smart Foodsheds**
  - Intelligent Scaling and Scheduling
- How to Get Engaged?
- Conclusions

Smart Foodsheds

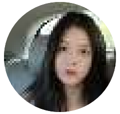
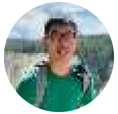
The Challenges

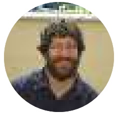
- Food system lacks resilience (highlighted by the pandemic)
- Food system actors are diverse, work in silos
- Access to data is difficult as is reconciling across data sources
- Need a common framework to organize, share, visualize, and deploy datasets and workflows

The Strategies

- Develop relationships between ICICLE and private partners to empower stakeholders to access, interpret, and utilize food systems processes, trends, and outcomes
- Use knowledge graphs to link domain knowledge of the environment, agriculture, food, diet, and health
- Develop PPOD, a schema that describes the attributes and relationships between **Persons, Projects, Organizations and Datasets** and instantiate it with real data from California and Ohio as a first use case.

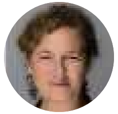
Interactive Knowledge Learning & Environment (IKLE) for Smart Foodshed

Yamei Tu *Xiaoqi Wang* *Rui Qiu* *Han-Wei Shen* *Patrick R Huber* *Allan D Hollander*

The Ohio State University *University of California Davis*


Matthew Lange *Michelle Miller* *Jinmeng Rao* *Song Gao* *Alfonso Morales*

International Center for Food Ontology Operability Data and Semantics (IC-FOODS)

University of Wisconsin-Madison


Christian R Garcia *Joe Stubbs*

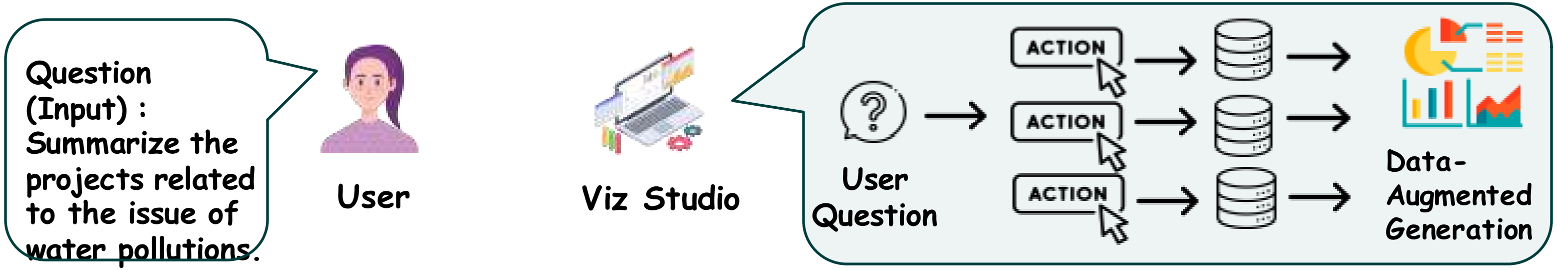
The University of Texas at Austin Texas Advanced Computing Center

Demo : Smart Foodsheds + Visual Analytics (IKLE)

https://youtu.be/Q5_i9PbefHw

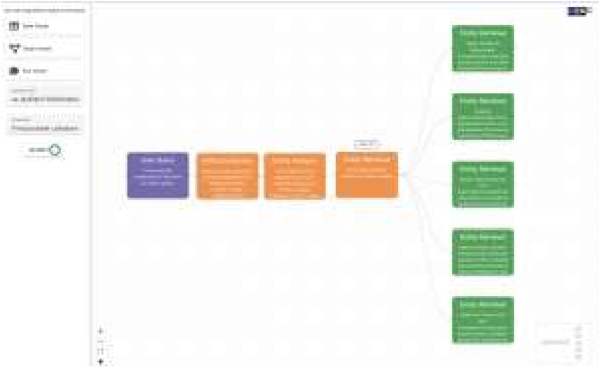
VizStudio

- Motivation: Enhance decision-making with tailored data-driven solutions. Our approach specifically addresses domain-specific challenges, providing users with actionable insights for real-world applications.
- Goals:
  – Transforming high-level questions into specific data analysis tasks using a what-why-how framework.
  – Merging the power of language models with a knowledge graph that anchors domain-specific insights.
  – Demonstrating each phase of the process through comprehensive visualization, actively incorporating user feedback to refine.



Conversational Interface

- **Interface Design:**
  - Developed an interface that accepts queries in the form of questions.
- **Architecture Development:**
  - **Front-end**: Implemented with Vue 3 and VueFlow for dynamic user experiences.
  - **Back-end**: Utilizes Django with Uvicorn to enable asynchronous communication between the front-end and back-end.
- **Functionality:**
  - **Question Pre-Processing**: Utilized Large Language Models to analyze and process user queries.
  - **Information Retrieval**: Integrated with the PPOD-KG to fetch relevant information based on user queries.
  - TO DO **Interactive Visualizations**: Enhanced user engagement through graph visualizations, allowing for interactive exploration of follow-up queries.



Demo

You can drag these nodes to the panel.
Table Viewer
Graph Viewer
Text Viewer
Scatterplot
OpenAI API
sk-QUE8k17US0Sr4tA4
Question
QUERY

GROCERY STORE CLOSURE & COMMUNITY HEALTH

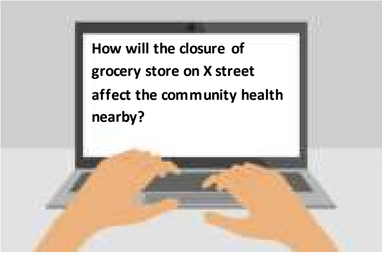
Pain points

- In public health and food systems, computer models are not used or have limited impact because decision-makers are not able to access them in a practical and timely manner.

SCENARIO

A food retail company announces plans to **close a grocery store** in a Columbus, Ohio neighborhood with **very high % of food-insecure households**.

Now the health commissioner wants to know how the grocery store closure will affect community health so they can lobby the food retail company to not close the grocery store or set up emergency food supply to reduce the impact on community health.



Our Solution
(and use case in ICICLE)

A conversational AI-enabled web interface that allows decision-makers to run "What if?" scenarios based on an agent-based model for food insecurity.

Use Case

Objective: Help food system leaders quickly evaluate the impact of a food store closure on household food insecurity

Significance: Improving access to community-informed computational models empowers communities to use models to make better decision involving complex systems, such as the local foodshed.

Grocery Store Closure Team

Harsh Panday | *Amad Hussain* | *Erika Goetz* | *Carlos Guzman* | *Ayaz Hyder* | *Huan Sun* | *Eric Fosler-Lussier*

The Ohio State University
College of Public Health / Dept. of Computer Science & Engineering

Demo: GROCERY STORE CLOSURE & COMMUNITY HEALTH

https://youtu.be/GYjMeaE74sk

Outline

- ICICLE Vision and Goals
- Research Challenges being Addressed
- **Selected Accomplishment Highlights**
  - CI/Software Released
  - Digital Agriculture
  - Animal Ecology
  - Smart Foodsheds
  - **Intelligent Scaling and Scheduling**
- How to Get Engaged?
- Conclusions

Project outline and goals

- Reference implementation and demonstration of end-to-end optimizations across the edge to HPC continuum.
- Integrating and interoperating set of knowledge/intelligence enhanced CI components.
- The end-to-end computational capability optimizes for, e.g. performance, model accuracy, or energy usage applied to the operation of AI assets.
  - Example flow for a model update

    Monitor model performance at the edge (animal ecology use-case) -> trigger model replacement and re-scheduling (FaaS) -> retrain model (smart scheduling).
  - Other flows: React to dynamic latency/energy/cost/resource-availability

AI4CI: HARP – HPC Application Runtime Predictor

Swathi Vallabhajosyula

Rajiv Ramnath

Carlos Guzman

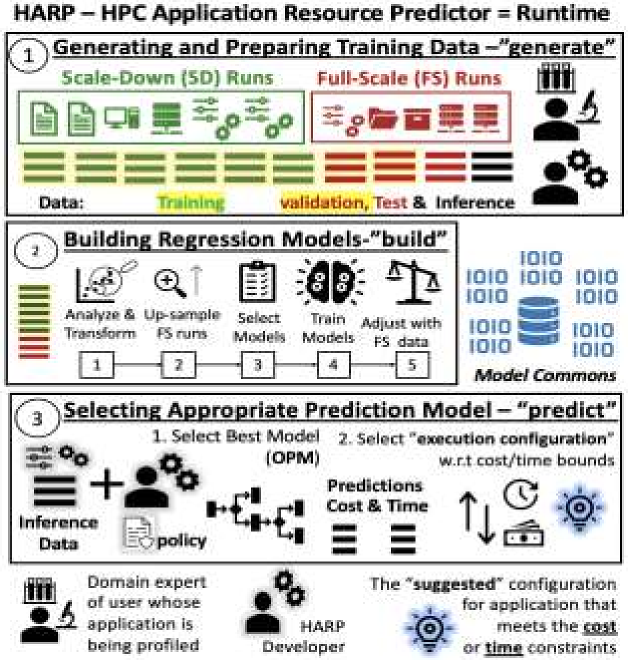
Joe Stubbs

The Ohio State University
Dept. of Computer Science & Engineering

The University of Texas at Austin
Texas Advanced Computing Center

Cost ↓ ~30%

n-tasks-per-node	Walltime (mins)	Cost Per job ($)
10	8.5954	0.01719
14	8.5768	0.01886
20	8.5852	0.02189
28	8.5931	0.02492



Goal

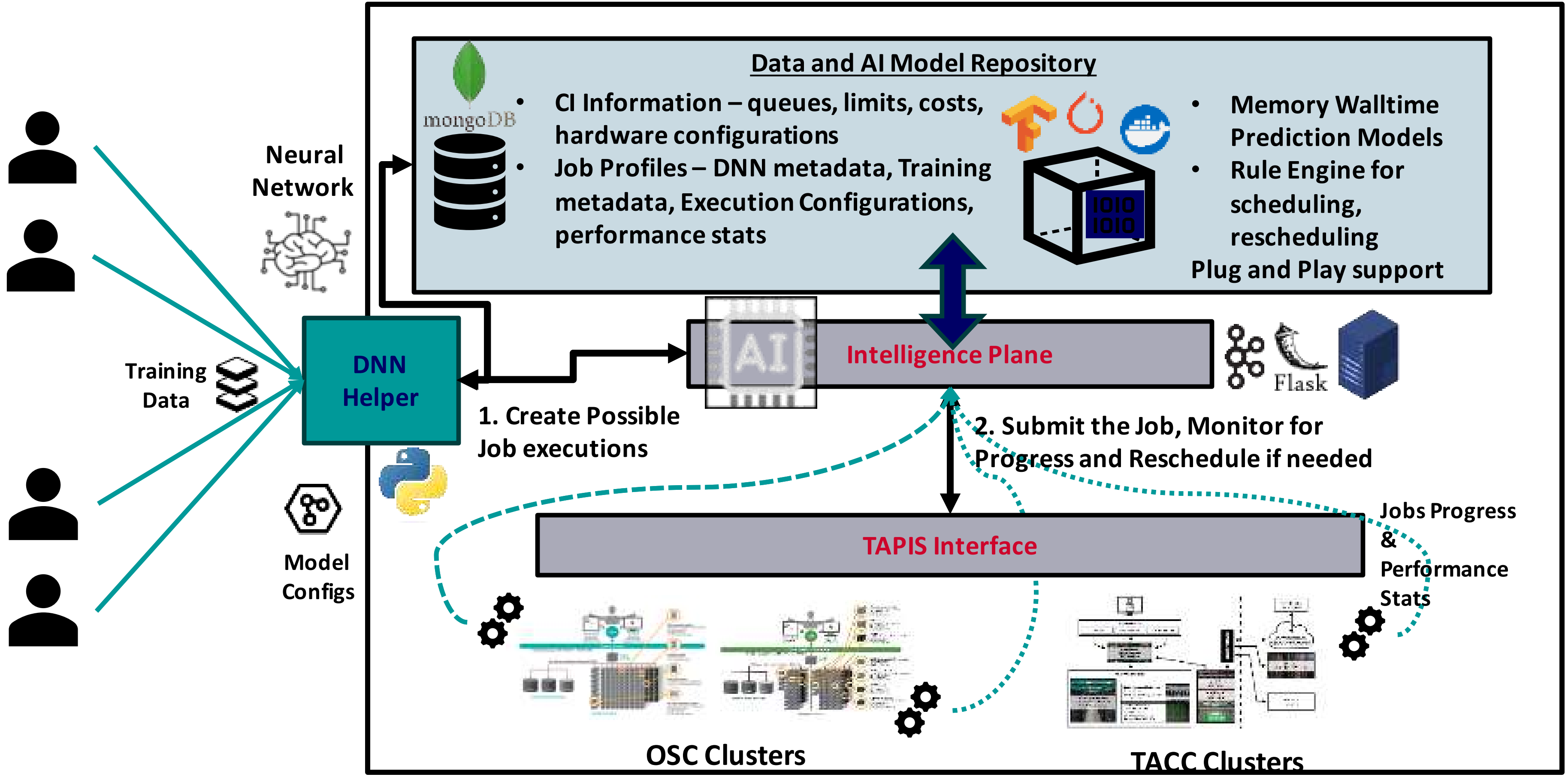
Estimating the resource requirements to execute an application on shared cyber infrastructures to aid recommendation systems or smart job allocations.

Accomplishments

- Understanding the allocation behavior of different users against different systems and ways for optimizing the allocations.
- Establishing an end-to-end application-independent framework called HARP (HPC Application Runtime Prediction) that can emulate the application executions, profile them, and estimate the resource requirements against targeted environments with cost/time constraints.

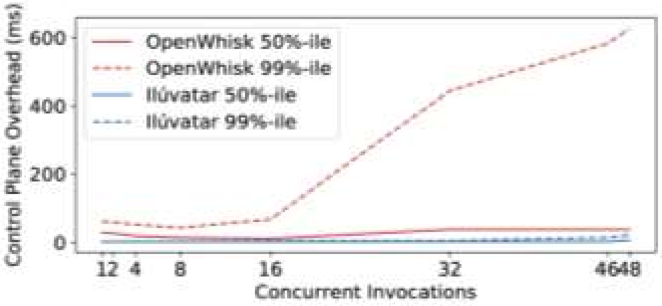
S. Vallabhajosyula and R. Ramnath, " Insights from the HARP Framework: Using an AI-Driven Approach for Efficient Resource Allocation in HPC Scientific Workflows", PEARC 2023

Smart Scheduler

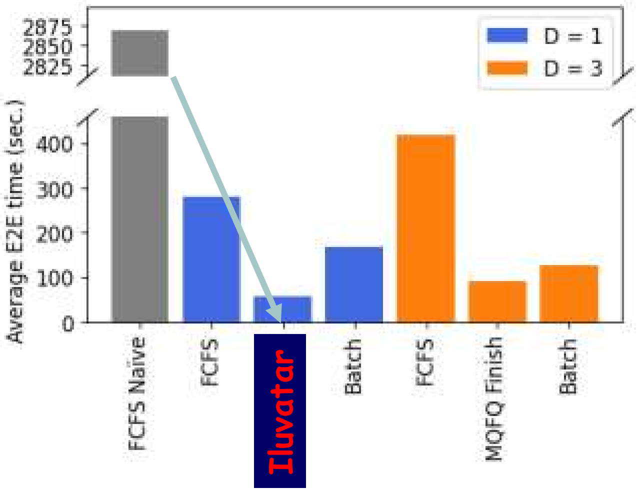


High Performance Functions as a Service

- Iluvatar: A Control-plane for Functions as a Service [HPDC '23
  - Runs on edge, cloud, and HPC . **100x faster than OpenWhisk.**
  - FaaS is a common abstraction for many services
    - ML inference, data processing, …
    - Iluvatar uses a new Rust-based implementation architecture: allows for flexible queuing policies



- Recent work: GPU support for functions
  - Fair queueing policies and GPU multiplexing mechanisms
  - Reduce latency by 80x compared to naïve GPU usage
  - Democratize AI by running ML models and analytics in functions everywhere (cloud, edge, …)

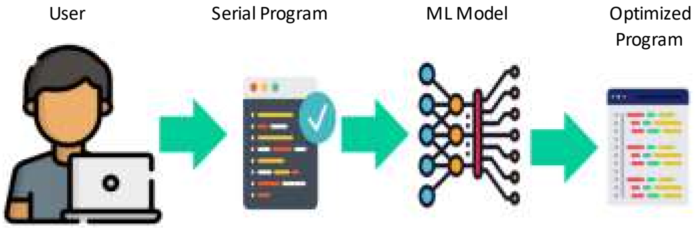


Compiler and Code-Optimization (CO)

- A source-to-source compiler framework that collects both static (source-code) as well as run-time information from applications and uses this information to generate optimized code across the ICICLE stack.

- **Problem:** Determine what optimizations to apply and where depends on the program and the target platform. The different sets of optimizations will create a vast optimization space.

- **Accomplishment:**
  - Identified a set of features that impact the performance of a target application.
  - Developed a collection mechanism that gathers both static and dynamic information of programs.
  - Developed a preliminary ML model capable of predicting the program transformation techniques.
  - Developed a preliminary Search Space Navigation System capable of choosing the best set of transformations techniques.



Outline

- ICICLE Vision and Goals
- Research Challenges being Addressed
- Selected Accomplishment Highlights
- **How to Get Engaged?**
- Conclusions

Engagement with Other Organizations

- MOU signed and Collaborative Projects going on with other AI Institutes
  - AIIRA
  - AIOpt
- On-going discussions with several other AI Institutes
  - AIFARMS
  - AgAID (Joint seminar series)
- On-going Collaboration with TIH-Mumbai along Digital Agriculture
- Interactions with industry (Digital Agriculture)
  - CNH Industrial
  - TCS, India

Multiple Levels of Collaboration and Engagement

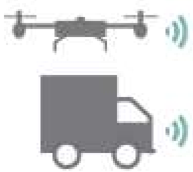
- Using the Released Software/CI components
  - Available at https://icicle.osu.edu/cyberinfrastructure/software
  - Get engaged as a member in the Stakeholder Roundtable (more details below)
- Become a part of ICICLE (multiple options)
  - Student Associate
  - Visiting Research Fellow
  - Academic Collaborator
  - Industry Partner
  - Stakeholder Roundtable Member
  - More details at: https://icicle.osu.edu/engagement/join-us
- Join the ICICLE mailing lists (https://icicle.osu.edu/engagement/mailing-lists)
  - icicle-announce
  - icicle-discuss

Outline

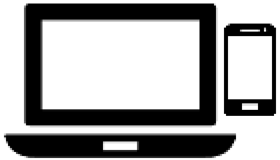
- ICICLE Vision and Goals
- Research Challenges being Addressed
- Selected Accomplishment Highlights
- How to Get Engaged?
- **Conclusions**

Designing Next-Generation CI through Co-Designing with Use-inspired Domains

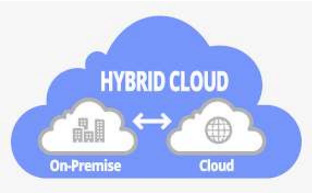
Conv. AI



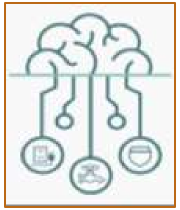
Data: On Field Sensors



Models: Edge & Near Edge



Data/Models: Cloud



Data/Models: HECs

ICICLE-enabled Computing Continuum

Potential for the ICICLE Solutions to be applied to more Verticals

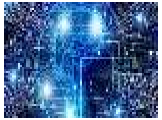
Smart Foodsheds

Digital Agriculture

Animal Ecology

Health & Medicine

Environment



Communications & Collaboration

Mobility, Machines, & Manufacturing

AI for Social Good

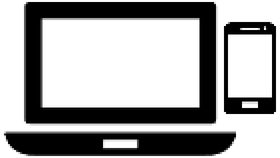
ICICLE: Intelligent CyberInfrastructure with Computational Learning in the Environment

Systems AI Foundational Research for CI

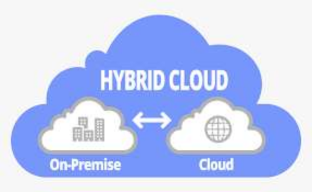
Intelligent Cyber Infrastructure

CI for AI	**AI for "CI for AI"**

On Field Sensors



Edge & Near Edge

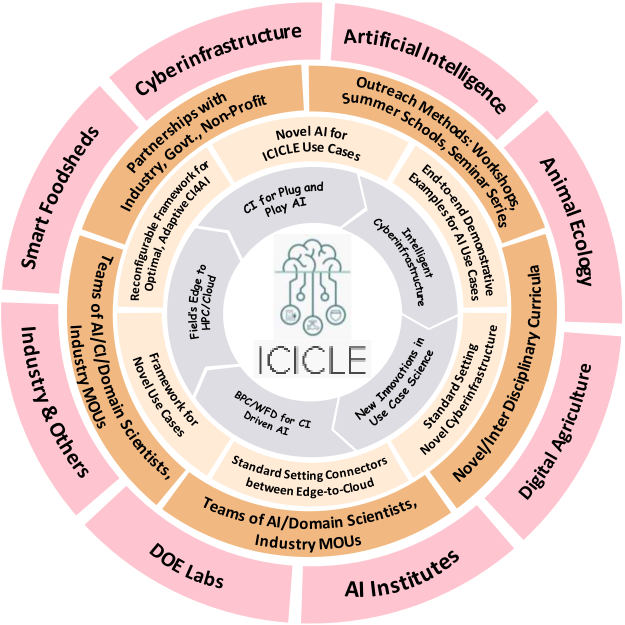


Clouds

HPC Systems & Data Centers

Emerging Computing Continuum

ICICLE Enabling Global Leadership in `Computing + AI’



Join Us!

- Integrate into the National CI Ecosystem
- Integrative and Interoperable
- Leverages existing recognized capabilities
  - Centers of Excellence, AI Institutes, Large Facilities
- Collaborative
  - Actively engaging CI experts, domain scientists,
  - AI/CI Users and developers
- Sustainable and Inclusive
  - Workforce Development, Broadening Participation, Collaboration and Knowledge Transfer
  - Benefits other institutes, large facilities, and all sciences beyond lifetime of award

Acknowledgments to all ICICLE Participants (Faculty, Students and Staffs)

Current Faculty
- E. Ayday, CWRU
- V. Chaudhary, CWRU
- A. Azad, IU
- P. Sharma, IU
- H. Zhang, ISU
- T. Berger-Wolf, OSU
- S. Blanas, OSU
- Y. Cai, OSU
- W. Chao, OSU
- E. Fosler-Lussier, OSU
- A. Hyder, OSU
- DB. Jackson-Smith, OSU
- R. Machiraju, OSU
- DK. Panda, OSU
- R. Ramnath, OSU
- S. Shearer, OSU
- H. Shen, OSU
- C. Stewart, OSU
- Y. Su, OSU
- H. Subramoni, OSU
- H. Sun, OSU
- C. Stewart, RPI
- B. Salimi, UCSD
- R. Eigenmann, UD
- A. Ahmad, Uni Stuttgart
- E. Riloff, UU
- P. Sadayappan, UU
- E. Ely-Ledesma, UW-Madison
- S. Gao, UW-Madison
- A. Morales, UW-Madison

Past Staff
- C. Campbell, IU
- S. Sanders, IU
- A. Ivanovic, OSU
- P. Rose, UCSD
- K. Pierce, TACC

Past Faculty
- C. Hoy, OSU
- T. Tomich, UC Davis
- J. Duarte, UC San Diego
- M. Norman, UC San Diego

Past Ph.D. Students
- FB Saravi, CWRU
- MK. Rahman, IU
- T. Zhang, ISU
- H. Ahn, OSU
- P. Chawla, OSU
- E. Goetz, OSU
- Y. Gu, OSU
- A. Jain, OSU
- D. Suresh, OSU
- S. Raje, UU
- H. Park, UW Madison

Current Staff
- M. Lange, IC-Foods
- T. Ruemping, IC-Foods
- D. Siedband, IC-Foods
- M. Biggers, IU
- RJ. Ping, IU
- BA. Plale, IU
- J. Wernert, IU
- M. Abduljabbar, OSU
- K. Armstrong, OSU
- J. Chan, OSU
- J. Chumley, OSU
- C. Guzman, OSU
- W. Michel, OSU
- N. Savardekar, OSU
- A. Shafi, OSU
- S. Khuvis, OSC
- S. Oottikkal, OSC
- K. Tomko, OSC
- D. Choi, SDSC
- M. Kandes, SDSC
- A. Majumdar, SDSC
- P. Rodriguez, SDSC
- M. Tatineni, SDSC
- R. Cardone, TACC
- C. Garcia, TACC
- S. Li, TACC
- J. Stubbs, TACC
- Z. Zhang, TACC
- A. Hollander, UC Davis
- P. Huber, UC Davis
- C. Riggle, IC-Foods
- P. Hoover, UCSD
- M. Thomas, UCSD
- M. Miller, UW Madison

Current International Students TIH - IITB
- A. Borkar, TIH IITB
- RM. Chitre, TIH IITB
- R. Katole, TIH IITB
- S. Khandelwal, TIH IITB
- T. Sharma, TIH IITB
- A. Thaduri, TIH IITB
- S. Zac, TIH IITB

Current International Faculty TIH – IITB
- M. Baghini, IITB
- Chalapathi G, IITB
- A. Sinha, IITB
- R. Velmurugan, IITB
- S. Paramane, TIH IITB

Current K-12 Students
- R. Estanislao, SDSC
- D. Lee, SDSC
- M. Ray, SDSC
- S. Samar, SDSC

Past Masters Students
- SR. Kalli, OSU
- H. Panday, OSU
- RR. Loka, UW Madison
- D. Sykes, UW Madison

Current Ph.D. Students
- P. Kousha, OSU
- Z. Li, OSU
- V. Pahuja, OSU
- R. Qiu, OSU
- E. Romero, OSU
- C. Tu, OSU
- Y. Tu, OSU
- S. Vallabhajosyula, OSU
- L. Waltz, OSU
- B. Wang, OSU
- X. Wang, OSU
- J. Yao, OSU
- X. Yue, OSU
- T. Zhang, OSU
- K. Zhang, OSU
- Z. Zhang, OSU
- DD. Vecchia, RPI
- M. Rosas, UD
- T. Jiang, UU
- Y. Xu, UU
- J. Yan, UU
- K. Armendariz, UW-Madison
- J. Rao, UW-Madison
- J. Kline, OSU
- G. Ubbiali, IC-Foods

Current Institute Evaluators (WFD)
- T. McKlin, TFG
- C. Wise, TFG

Past K-12 Students
- J. Karpinski, SDSC
- A. Sarin, SDSC

Past UG Students
- S. Ockerman, OSU
- KP. Sailaja, OSU
- C. Washington, OSU
- J.Kim, TACC
- C. Skevofilax, TACC
- S. Wegner, UW Madison

Current Masters Students
- R. Danhi, IC-Foods
- J. Cheng, OSU
- S. Deshmukh, OSU
- M. Han, OSU
- A. O'Quinn, OSU
- C. Wang, OSU
- J. Yang, OSU
- Q. Ding, TACC
- V. deBella, UW Madison
- M. Krempely, UW Madison
- S. Suresh, UW Madison
- G. Wilkins, UW Madison

Current Undergraduate Students
- T. Chen, OSU
- KA. Irizarry, OSU
- M. Lieber, OSU
- E. Luo, OSU
- D. Venkataraman, OSU
- S. Shah, UT Austin
- A. Karunakaran, UW Madison
- M. Kuhn, UW Madison
- Y. Qu, UW Madison
- K. Sung, UW Madison

Educational Fellows (2023)
- B. Alston, OSU
- TE. Feiten, UC
- A. Hingle, GMU
- C. Lucken, UC
- C. Okolo, CU

Thank You!