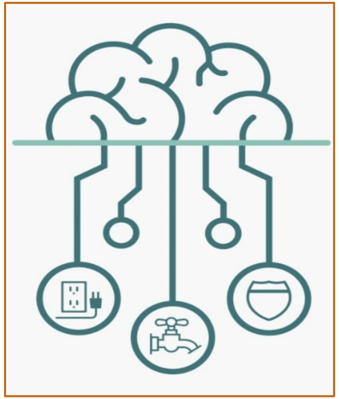
Follow us
@icicleai

http://icicle.ai

NSF-Funded AI Institute
($20M USD for Five Years)



Intelligent CyberInfrastructure With Computational Learning in the Environment (ICICLE)

PI: Dhabaleswar K (DK) Panda
CSE, OSU

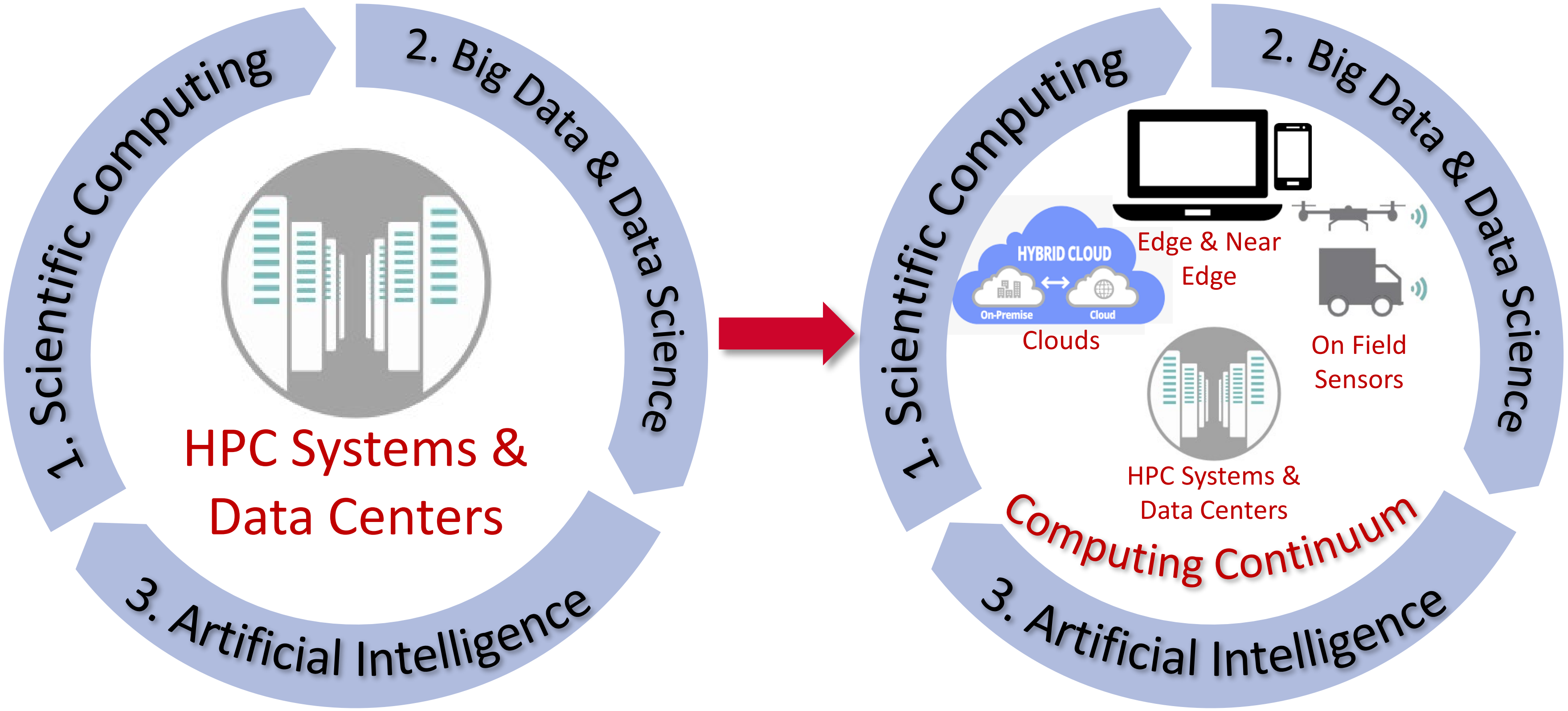
Outline

- Brief Overview of the ICICLE Institute
- Organizational Infrastructure and Management
- Strategic and Implementation Plan
- Overall Research Directions
- Example Accomplishments (so far)
  - AI Foundations
  - CI4AI
  - Digital Agriculture
- Conclusion

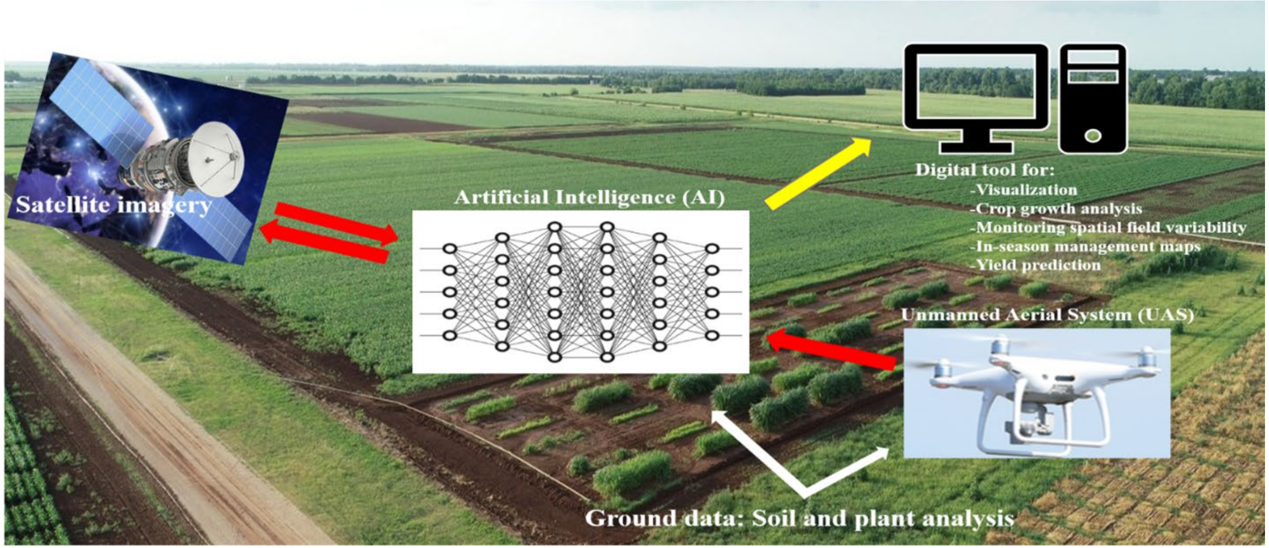
Computing has been evolving over the last three decades with multiple **phases:**

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

Phase 4 (2015 -): Emergence of the Computing Continuum



AI-Driven Digital Agriculture



https://ccag.tamu.edu/research-project/digital-agriculture/

https://medium.datadriveninvestor.com/artificial-intelligence-in-agriculture-62f71f8f6ae6

The Vision

A **national infrastructure** that enables AI at the flick of a switch, ICICLE will:

- **Democratize AI** through **integrated plug-and-play AI**.
- Catalyze **foundational AI/CI** and transform application domains.
- **Transparent and trustworthy** infrastructure for AI-enabled future,
- Address **societal problems** (conservation, food insecurity) and **national priorities**
- Grow **new generations of workforce** and Incubate **sustainable and inclusive** communities

Objectives: Intelligent CyberInfrastructure for Computing Continuum

Use Inspired Science Domains

Smart Foodsheds

Animal Ecology

Digital Agriculture

ICICLE: Intelligent CyberInfrastructure with Computational Learning in the Environment

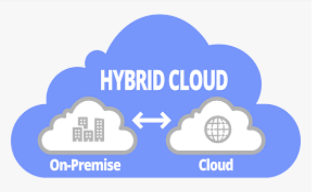
Systems AI Foundational Research for CI

Intelligent Cyber Infrastructure

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

On Field Sensors

Edge & Near Edge



Clouds

HPC Systems & Data Centers

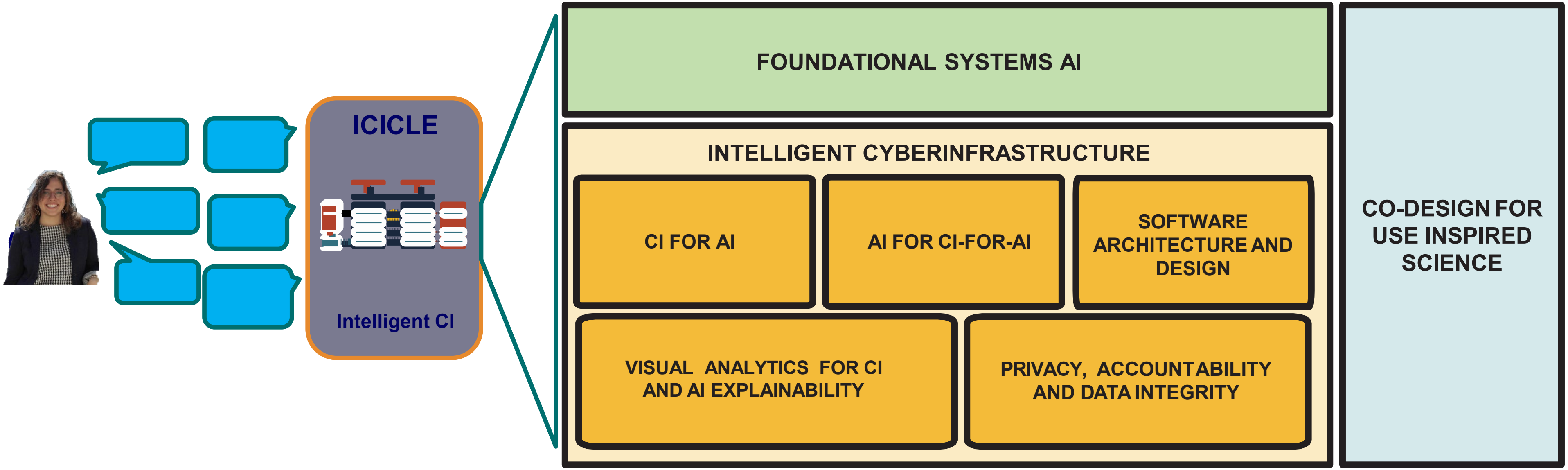
Emerging Computing Continuum

Integrating a broad range of

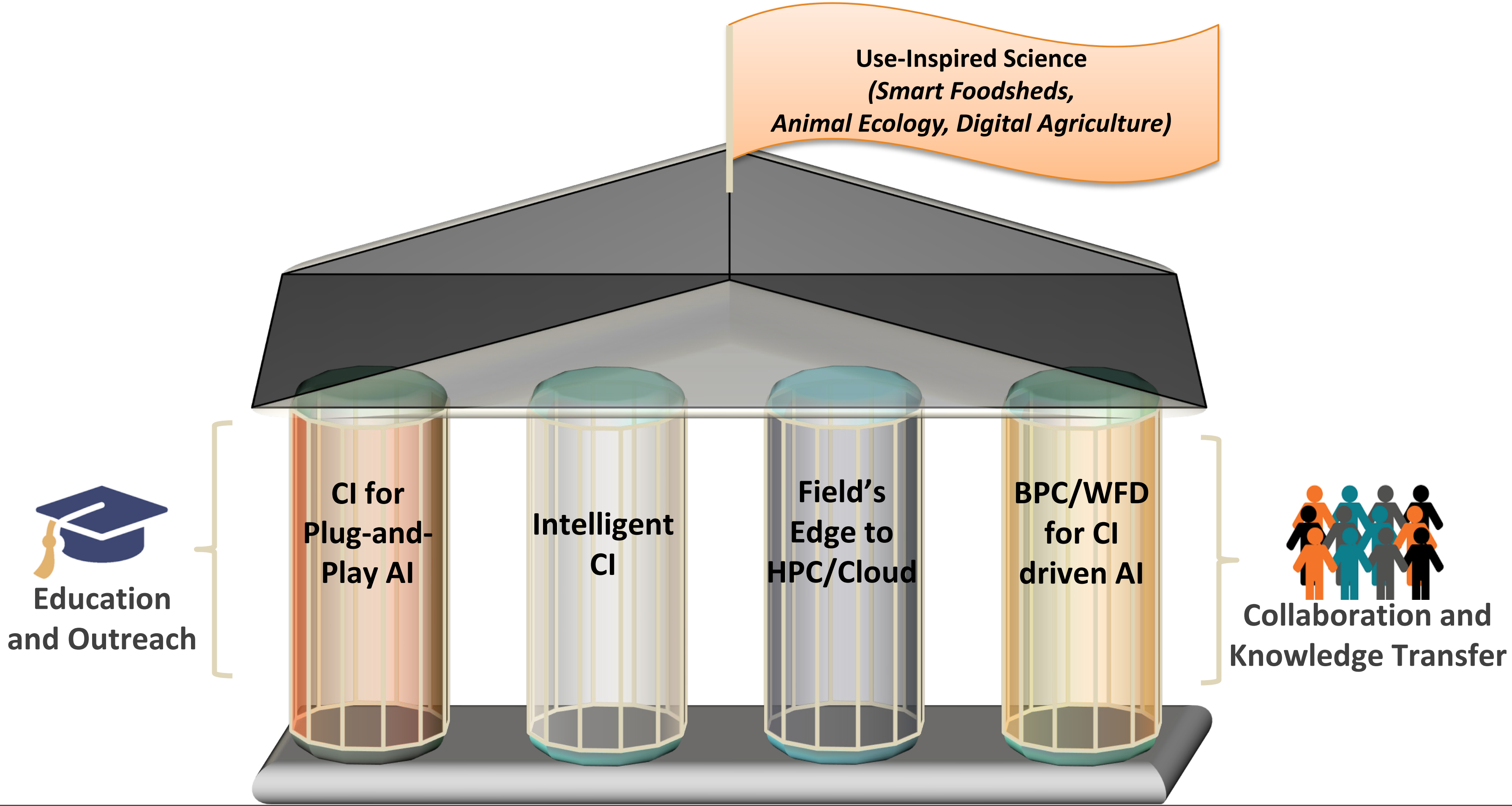
- Scientists-in-the-field
- Engineers
- Educators
- Collaborative partners
- Institutions

under one roof enables **democratized, adaptable, plug-and-play AI** and **long-tail science.**

Research Plan: Overall Vision



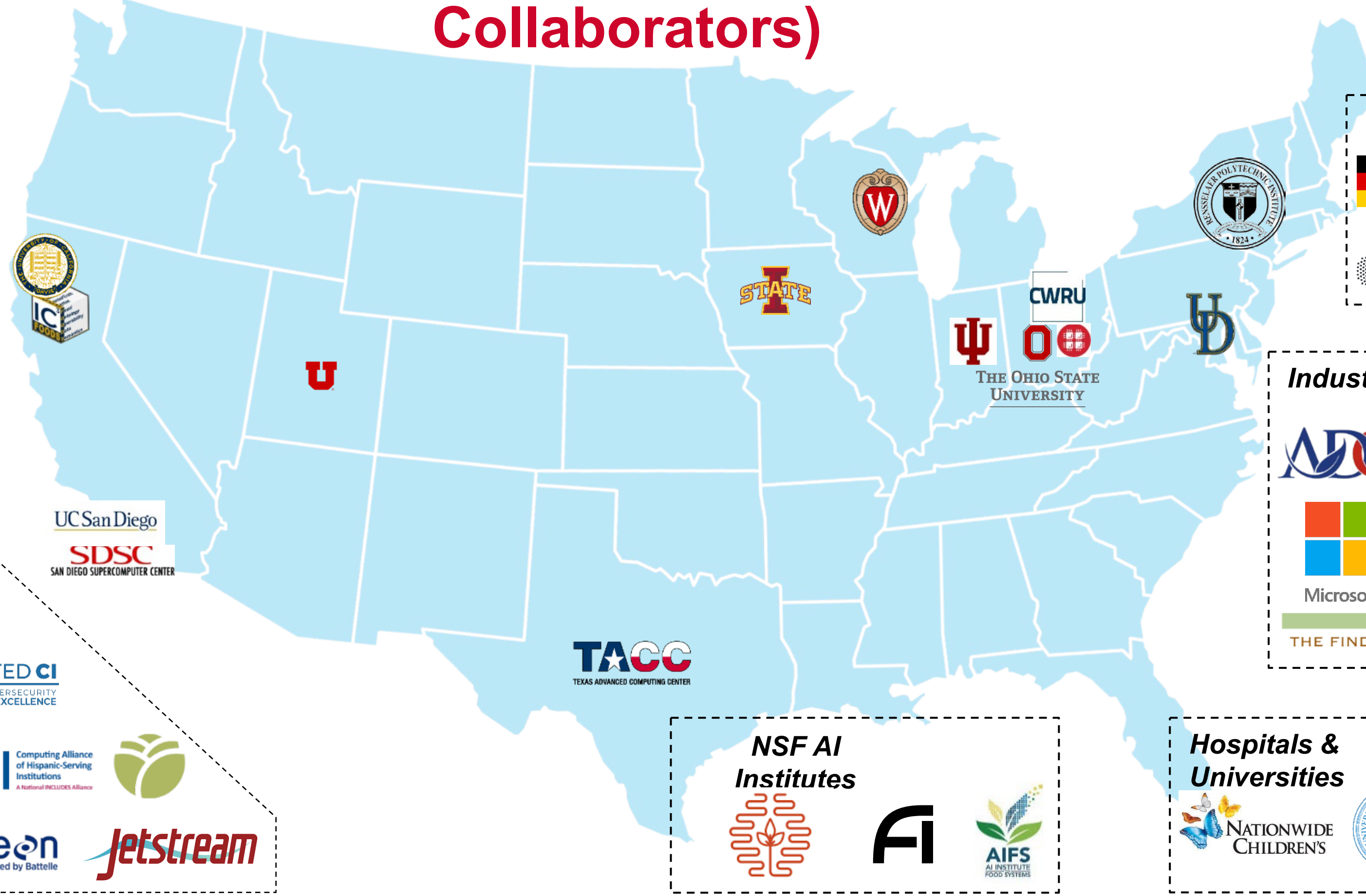
ICICLE As A Whole



Outline

- Brief Overview of the ICICLE Institute
- Organizational Infrastructure and Management
- Strategic and Implementation Plan
- Overall Research Directions
- Example Accomplishments (so far)
  - AI Foundations
  - CI4AI
  - Digital Agriculture
- Conclusion

Participation (14 Organizations, 46 Investigators, and many Collaborators)



Govt. Agencies & National Labs

Lawrence Livermore National Laboratory

U.S. Fish & Wildlife Service

Brookhaven National Laboratory

Fermilab

Research Institutes

national center for women & information technology

MolSSI

SGCI Science Gateways Community Institute

Trusted CI The NSF Cybersecurity Center of Excellence

ECEP

CAHSI Computing Alliance of Hispanic-Serving Institutions

SCEC An NSF+USGS Center

NSF

neon Operated by Battelle

Jetstream

International

Max-Planck-Gesellschaft

University of Stuttgart Germany

Industry

CNH Industrial

ADC

WILDME

Microsoft

FARM 2 FACTS

The Findings Group, LLC Research & Evaluation

NSF AI Institutes

AIFS AI Institute Food Systems

Hospitals & Universities

Nationwide Children's

ICICLE Thrust Leaders

EDUCATION & WORKFORCE DEVELOPMENT
Thomas (SDSC)

BROADENING PARTICIPATION IN COMPUTING & DIVERSITY
Biggers (IU)
Plale (IU)

FOUNDATIONAL SYSTEMS AI
Fosler-Lussier (OSU)
Su (OSU)

INTELLIGENT CYBERINFRASTRUCTURE

CI FOR AI
Blanas (OSU)
Zhang (TACC)

AI FOR CI-FOR-AI
Chaudhary (CWRU)
Eigenmann (UDel)

SOFTWARE ARCHITECTURE AND DESIGN
Stubbs (TACC)
Zhang (TACC)

VISUAL ANALYTICS FOR CI AND AI EXPLAINABILITY
Machiraju (OSU)
Shen (OSU)

PRIVACY, ACCOUNTABILITY AND DATA INTEGRITY
Ayday (CWRU)
Plale (IU)

CO-DESIGN FOR USE INSPIRED SCIENCE
Berger-Wolf (OSU)
Huber (UC-Davis)
Shearer (OSU)

KNOWLEDGE TRANSFER, COLLABORATION, & PARTNERSHIP
Lange (IC-FOODS)

SHARED CI DEVELOPMENT AND INTEGRATION
Majumdar (SDSC)
Tomko (OSU)

OUTREACH & COMMUNICATION
Savardekar (MD) (OSU)

SEMINAR PROGRAM
Morales (UW-Madison)

TECHNOLOGY TRANSFER & COMMERCIALIZATION
Chaudhary (CWRU)
Ramnath (OSU)

ICICLE Leadership Team


Panda (PI)
(OSU) [CI]

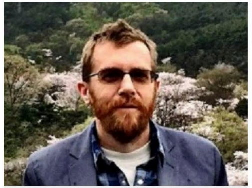

Chaudhary (Co-PI)
(CWRU) [CI]


Fosler-Lussier (Co-PI)
(OSU) [AI]


Machiraju (Co-PI)
(OSU) [AI]


Plale (Co-PI)
(IU) [BPC, CI]


Eigenmann (UDel)
[CI]


Huber (UC-Davis)
[Smart Foodsheds]


Lange (IC-FOODS)
[Smart Foodsheds]


Majumdar(SDSC)
[CI]


Morales (UW-Madison)
[All Applications]


Ramnath (OSU)
[BPC, CI]


Sadayappan (Utah)
[CI]


Savardekar (MD) (OSU)
[Management]


Stewart (RPI)
[Animal Ecology]


Stubbs (TACC)
[CI]

Tomko (OSC)
[CI]

Zhang (Iowa)
[AI, CI]

External Advisory Board (EAB)

Ewa Deelman
Univ. of Southern California
Cyberinfrastructure, Academia

Healy Hamilton
NatureServe
(Biodiversity, Non-profit)

Vipin Kumar
University of Minnesota
Cyberinfrastructure, Academia

Ted Schmitt
Allen Institute for AI
Applications, Non-profit

Sergio Soares
CNH Industrial
Use-Inspired Science, Industry

Dan Stanzione
University of Texas, Austin
Cyberinfrastructure, Lab/HPC

Valerie Taylor
Argonne National Laboratory
WFD/BPC, Lab/HPC

Tiffani Williams
Univ. of Illinois, Urbana-Champaign
WFD/BPC, Academia

Luke Zettlemoyer
Meta and Univ. of Washington
Artificial Intelligence, Industry

Outline

- Brief Overview of the ICICLE Institute
- Organizational Infrastructure and Management
- **Strategic and Implementation Plan**
- Overall Research Directions
- Example Accomplishments (so far)
  - AI Foundations
  - CI4AI
  - Digital Agriculture
- Conclusion

The Credo of ICICLE

NSF CI2030 report "... the cyberinfrastructure that will be available to support research in the 2020s will be vastly different from what exists today and will have capabilities that are not envisioned ..."

The ICICLE Institute is

- ushering in a new generation of AI-informed or "smart" CI to be drawn from a convergence of critical societal use cases and their AI requirements.
- the first and foremost *edge-to-center* AI-as-a-service enterprise to **advance foundational AI research**, benefiting a variety of experts in the field and provide means to democratize AI.

ICICLE is responding to the needs of all users

- Massive and ever-growing gap between AI and its accessibility to all users

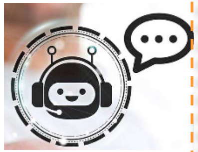
- Existing AI applications are developed largely ad-hoc and lack coherent, standardized, modular, and reusable infrastructure

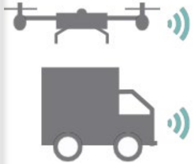
- Successful AI solution(s) for one use case rarely generalize to other use cases, or even the same use case even with slightly different context.

<u>CI's complexity to deploy AI impedes research discoveries and innovations!</u>

Co-Designing with use-inspired domains

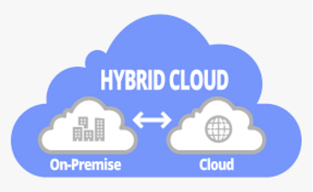
Conv. AI



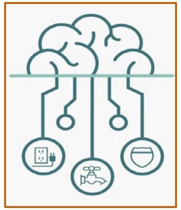


Data: On Field Sensors

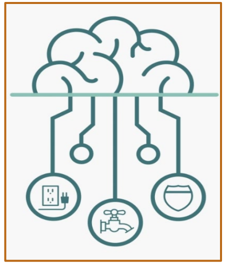
Models: Edge & Near Edge



Data/Models: Cloud



Data/Models: HECs

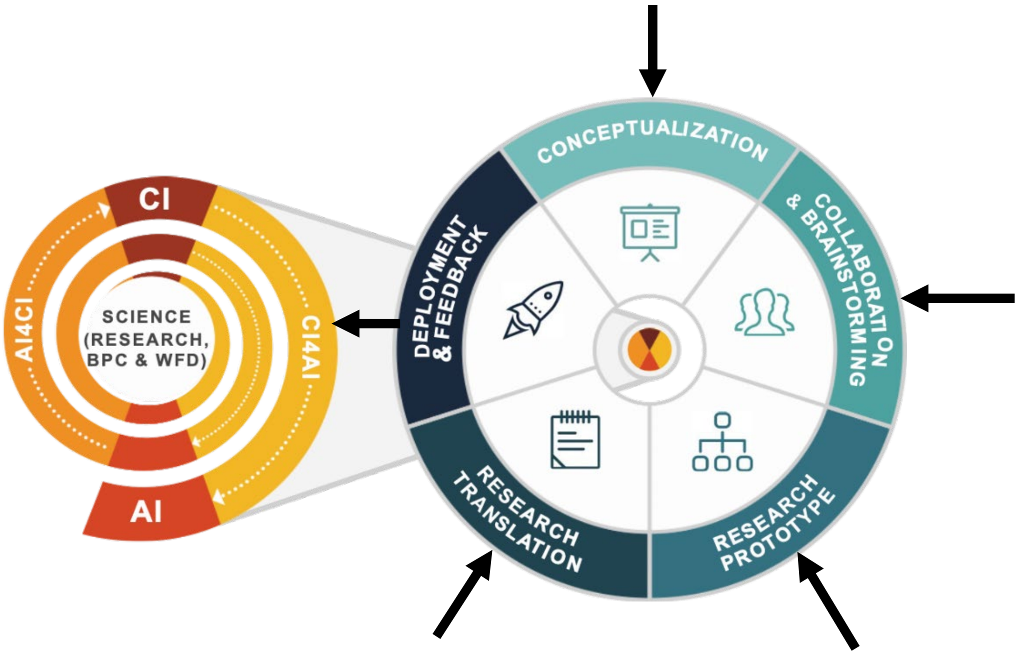


ICICLE-enabled Computing Continuum

End-users will keep data private, choose site and computation based on field constraints!

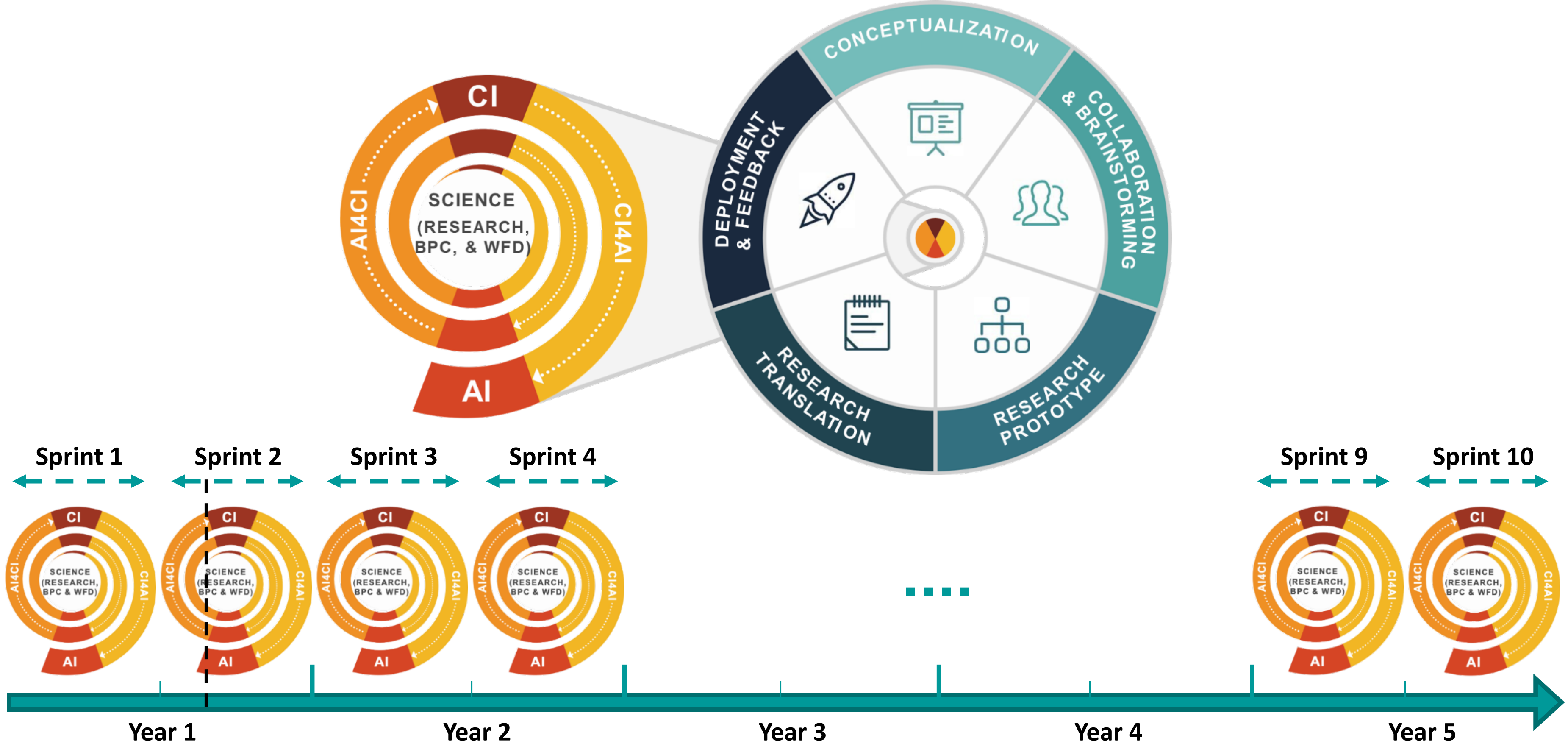
Virtuous Cycle realizes ICICLE's Co-Design

The virtuous cycle: today's AI is tomorrow's CI



<u>**Generalized AI-smart infrastructure will provide valuable research support to all users!**</u>

ICICLE's Smart CI will be built in 10 sprints



Strategic and Implementation Plan (SIP) Organization

Research Thrusts (Sec. IV)

FOUNDATIONAL SYSTEMS AI

INTELLIGENT CYBERINFRASTRUCTURE

- CI FOR AI
- AI FOR CI-FOR-AI
- SOFTWARE ARCHITECTURE AND DESIGN
- VISUAL ANALYTICS FOR CI AND AI EXPLAINABILITY
- PRIVACY, ACCOUNTABILITY AND DATA INTEGRITY

Use-Inspired Science (Sec. V)

CO-DESIGN

Broader impacts backbone network (sec. VIII)

- BROADENING PARTICIPATION IN COMPUTING
- WORKFORCE DEVELOPMENT
- COLLABORATION AND KNOWLEDGE TRANSFER
- DIVERSITY EQUITY & INCLUSION

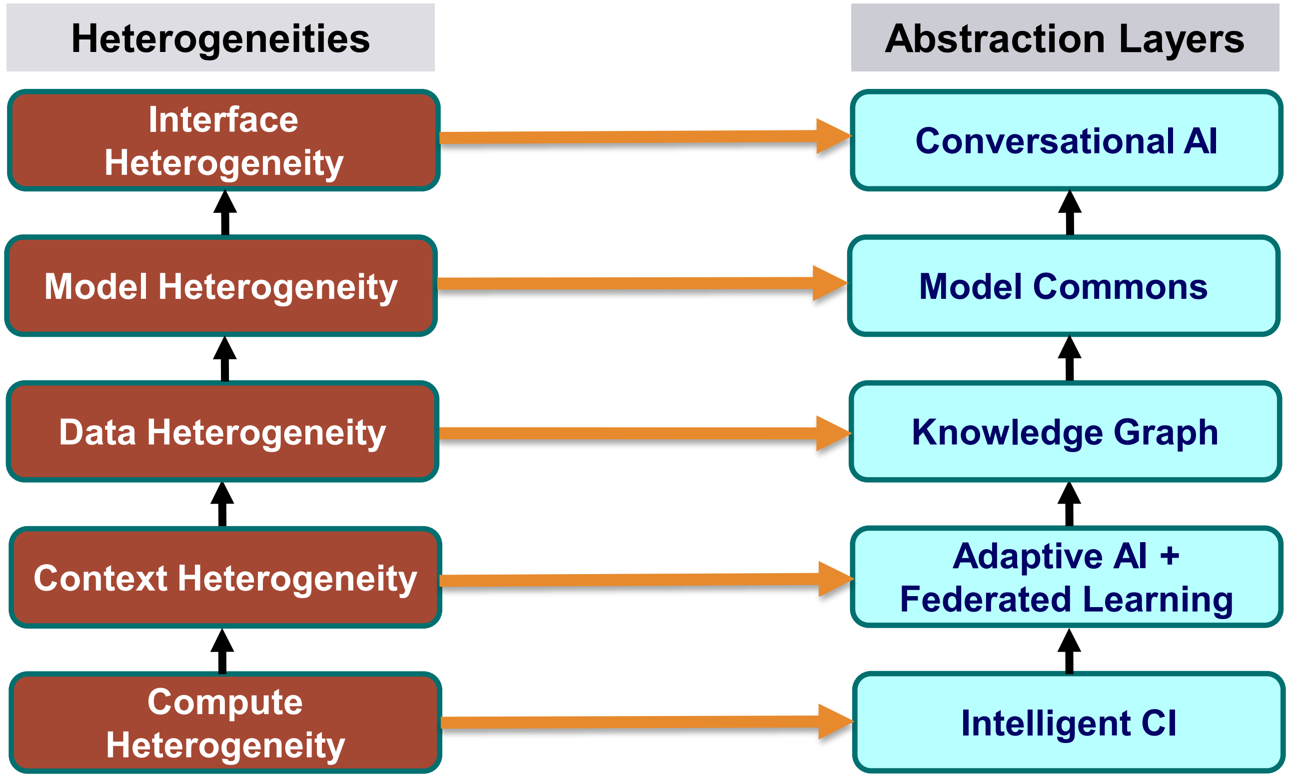
Strategic Projects (Sec. VI)

- SMART CAMERA TRAPS & TRADEOFFS
- VISUAL ANALYSIS OF FOOD SYSTEMS KNOWLEDGE GRAPHS
- DIGITAL AGRICULTURE USE-INSPIRED RESEARCH
- GROCERY STORE CLOSURE & COMMUNITY HEALTH
- “Hello ICICLE”
- REFERENCE SOFTWARE ARCHITECTURE

Outline

- Brief Overview of the ICICLE Institute
- Organizational Infrastructure and Management
- Strategic and Implementation Plan
- **Overall Research Directions**
- Example Accomplishments (so far)
  - AI Foundations
  - CI4AI
  - Digital Agriculture
- Conclusion

Plug-and-Play AI in the face of problematic heterogeneities



Enables standardization and generalization across use-cases!

Thrust: Foundational Systems AI

Components address CI complexity and heterogeneity for plug-and-play

Knowledge Graphs
- Multimodal
- Spatio-temporal
- Auto construction
- Knowledge-based reasoning and pre-training

Model Commons
- KG-supported
- Precise profiling
- Flex Composition
- Versioning and provenance

Adaptive AI
- Context-aware
- Interactive
- Continual learning
- Distillation-based compression

Federated Learning
- Heterogeneity
- Applicability to a variety of models
- Context-aware
- Privacy-preserving and robustness

Conversational AI
- KG- and model-commons-aware
- Bootstrapping and adaptivity
- Multimodal contextual response

Thrust: CI4AI

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

High Perf. Training
- Deep Learning Communication Optimization
- Deep Learning I/O Improvement
- Multi-level data/model/spatial parallelism

High Perf. Data Management
- Unified storage of data, model and hyperparameters
- Model lifecycle management for AI orchestration
- Data location transparency with migration

Edge Intelligence
- Adaptive Training/Inference and FL on Edge
- Novel Edge Offloading/Caching Orchestration
- Intelligent Anomaly Detection to improve QoS

AI-Adaptive Edge Wireless
- Adaptive, Predictable Comm. Capacity Allocation
- Predictable Wireless Comm. via Rateless-Coding & Multi-Modal/Path

Control and Coordination
- App/CI Interface Design
- Tapis Integration
- Production-ready Service Hardening and Optimization

Thurst: AI4CI

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

KGs & Model Commons for CI

- Conduct survey and literature review of available datasets that enable optimization of existing CIs

Intelligent Modeling and Optimization

- Collection of baseline x86 and ARM platforms for ResNet image pipeline by Digital Agriculture project

Applications

- Develop a "self-driving" AI middleware prototype for ML training

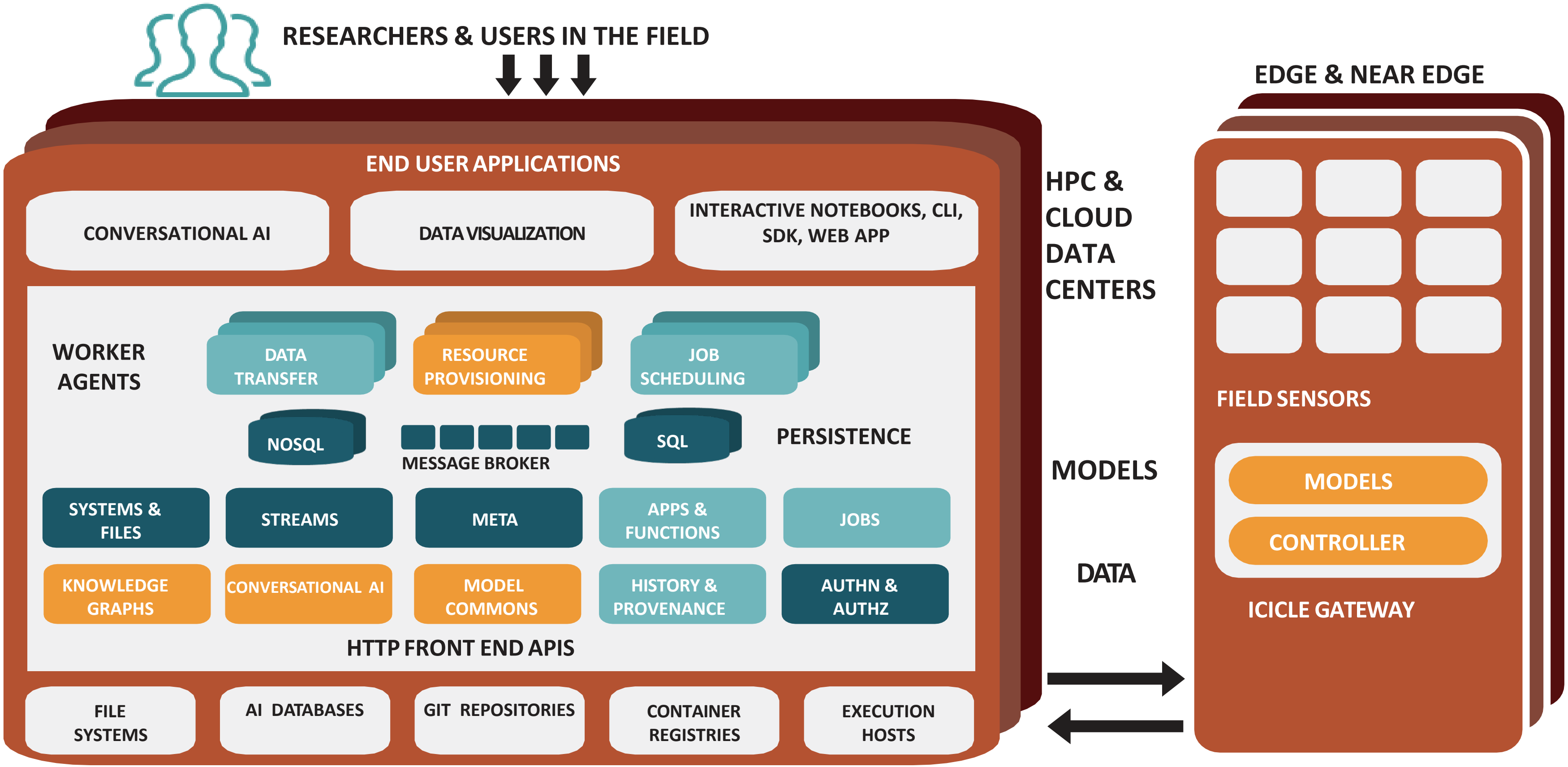
Middleware

- Develop a set of intelligent linear algebra kernels for sparse-matrix operations

Systems

- Resource allocation optimizer for ML training
- Intelligent edge wireless via mean-field reinforcement learning

The Deliverable: The ICICLE Software Stack

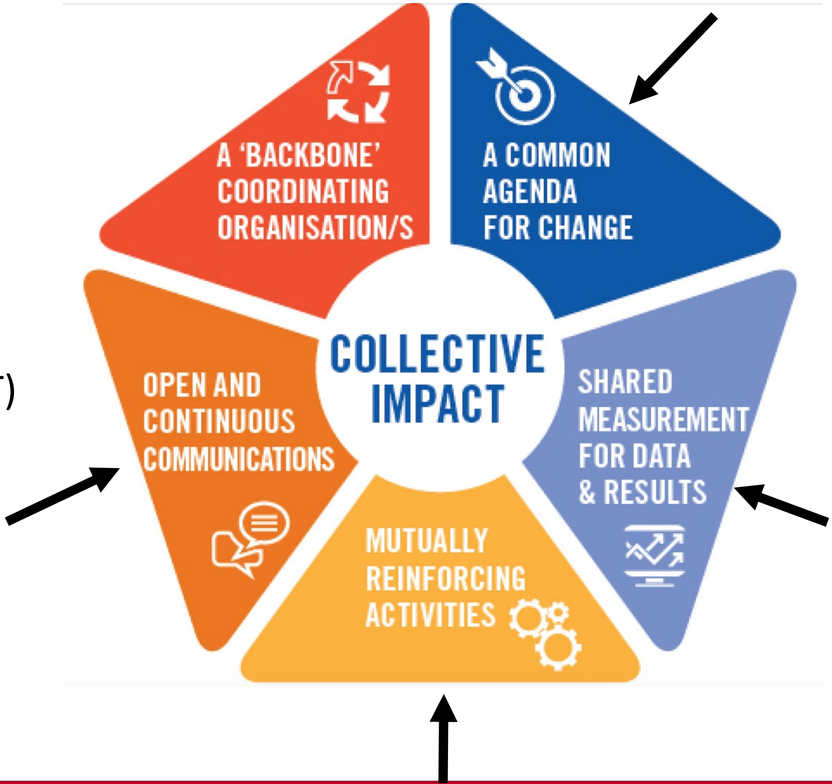


Broader Impacts Backbone Network (BIBN)

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

- Brief Overview of the ICICLE Institute
- Organizational Infrastructure and Management
- Strategic and Implementation Plan
- Overall Research Directions
- **Example Accomplishments (so far)**
  - **AI Foundations**
  - CI4AI
  - Digital Agriculture
- Conclusion

AI Foundations: Core Team & Project Collaborators

Investigators:

- **Yu Su** (OSU)
- **Eric Fosler-Lussier** (OSU)
- Erman Ayday (CWRU)
- Wei-Lun Chao (OSU)
- Song Gao (UW)
- Matthew Lange (IC-FOODS)
- Raghu Machiraju (OSU)
- Beth Plale (IU)
- Ellen Riloff (Utah)
- Christopher Stewart (OSU)
- Huan Sun (OSU)

Students/Staff:

- Jany Chan (OSU)
- Hong-You Chen (OSU)
- Erika Goetz (OSU)
- Yu Gu (OSU)
- Amad Hussain (OSU)
- Tianyu Jiang (Utah)
- Vardaan Pahuja (OSU)
- Jinmeng Rao (UW)
- Dhruv Venkataraman (OSU)
- Clay Washington (OSU)
- Xiang Yue (OSU)
- Zichen Zhang (OSU)

Smart Foodsheds

- Patrick Hubber (UC Davis)
- Allan Hollander (UC Davis)
- Ayaz Hyder (OSU)
- Michelle Miller (UW)
- Alfonso Morales (UW)

Animal Ecology

- Tanya Berger-Wolf (OSU)
- Chuck Stewart (RPI)

Digital Agriculture

- Christopher Stewart (OSU)
- Matthew Lange (IC-FOODS)

Reference Architecture

- Joe Stubbs (TACC)
- Zhao Zhang (TACC)
- Rajiv Ramnath (OSU)

Research Goals and Challenges

- **Conversational AI** to provide unified and user friendly interfaces for human-machine interaction and improve the accessibility and usability of the entire ICICLE system
- **Model Commons** to provide abstraction for addressing model heterogeneity and improve discoverability, matchability, and interoperability of AI models
- **Knowledge Graphs** as the knowledge backbone to provide semantically-rich abstraction for addressing data heterogeneity
- **Adaptive AI** to enable in-situ adaptation of AI models at the edge
- **Federated Learning** to support edge-to-center, decentralized, and privacy-preserving learning

Research in Y1

Knowledge Graphs

- Multi-modal KG formalism for camera traps
- KG construction for Smart Foodsheds
- New ontology and KG for commodity flow

Federated Learning

- Federated learning with pre-trained models (under review)
- Personalized federated learning (ICLR'22)

Conversational AI

- Novel Strongly generalizable question answering over large-scale KGs (under review)
- Conversational interface for Smart Foodsheds

Model Commons

- Survey existing model and data commons for feature comparison
- Started software dev based on OSU's Data Commons

Adaptive AI

- Question synthesis for domain adaptation of question answering models (ACL'22)

AI Foundation – Projects Impacted

- Knowledge Graphs
- Federated Learning
- Conversational AI
- Model Commons
- Adaptive AI

- Smart Camera Traps & Tradeoffs
- Digital Agriculture Use-Inspired Research
- “Hello ICICLE”
- Reference Software Architecture
- Visual Analysis of Food Systems Knowledge Graphs
- Grocery Store Closure & Community Health

Animal Ecology


Digital Agriculture

Smart Foodsheds

Research in Y1

Knowledge Graphs

- Multi-modal KG formalism for camera traps
- KG construction for Smart Foodsheds
- New ontology and KG for commodity flow

Federated Learning

- Federated learning with pre-trained models (under review)
- Personalized federated learning (ICLR'22)

Conversational AI

- Novel Strongly generalizable question answering over large-scale KGs (under review)
- Conversational interface for Smart Foodsheds

Model Commons

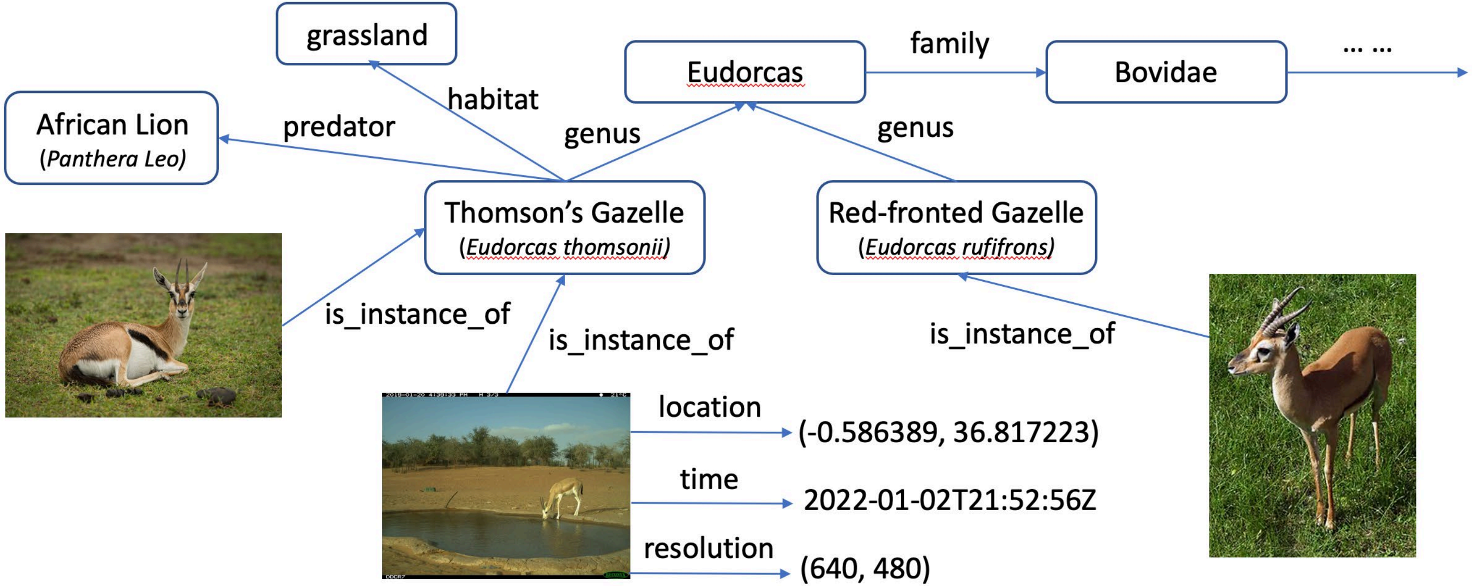
- Survey existing model and data commons for feature comparison
- Started software dev based on OSU's Data Commons

Adaptive AI

- Question synthesis for domain adaptation of question answering models (ACL'22)

KG Formalism: Multimodal KG for Camera Traps

- **Use-inspired Goal**: Develop new KG formalisms to represent and reason with multimodal data
- **Progress**: Co-design formalism with Animal Ecology, identified and analyzed public data sources, investigated KG tooling and decided on using Neo4j
- **Applications**: Semantic search, visualization, multimodal reasoning

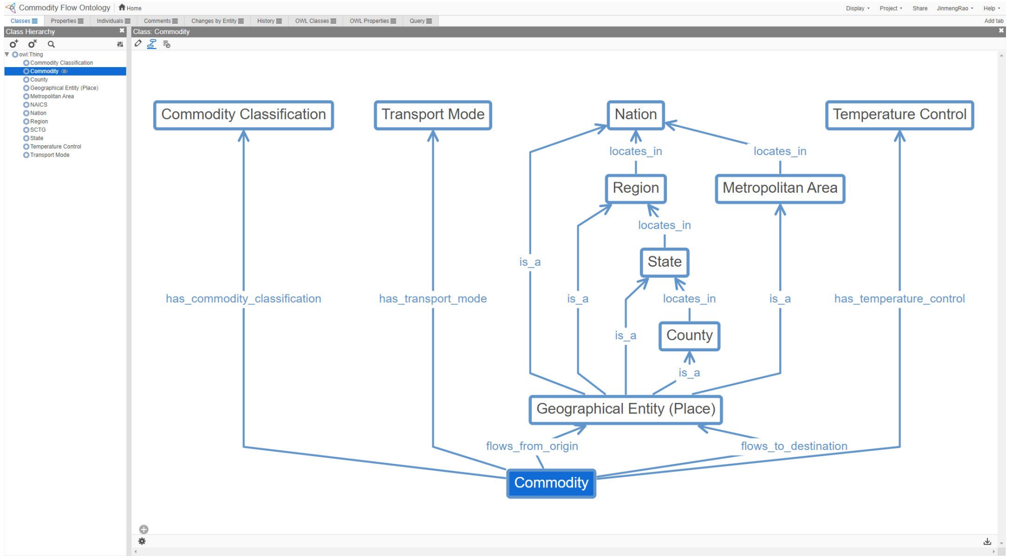


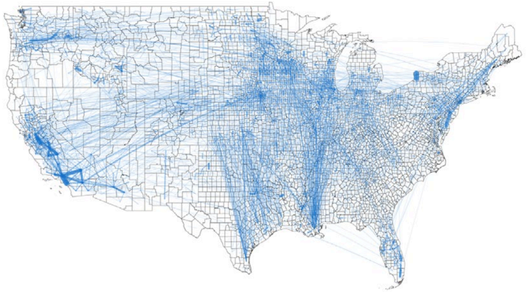
KG Construction: Learning Organization Knowledge for Smart Foodsheds

- **Problem Statement**: automatically assign organizations to categories in the Smart Foodsheds ontology to support rapid knowledge graph population.
- **Current Approach**: apply natural language processing and machine learning to categorize organizations based on WWW data (web pages and search engine results).
- **Challenges**: large multi-class and multi-label NLP-based categorization task; very small amount of human-labeled training data.

	Count	Pre	Rec	F1
Water	478	91.3	84.6	87.8
Physical Infrastructure	458	94.2	81.9	87.6
Wastes & Pollution	371	79.1	76.5	77.8
Biodiversity	343	79.8	77.9	78.8
Land & Soil	234	77.3	56.7	65.4
Food Production	221	66.7	56.8	61.3
Governance	196	100	92.6	96.2
Institutions	127	53.2	56.8	54.9
Common Pool Resources	93	84.6	51.2	63.8
Sociocultural Systems	73	100	87	93
Air & Climate	64	64.3	45	52.9
Agricultural Sector	60	54.5	35.3	42.9
Disasters	49	72.7	50	59.3
Educational Resources	45	92.9	68.4	78.8
Energy	44	50	30.8	38.1
Micro AVG		82.8	72.4	77.3
Macro AVG		77.4	63.4	69.2

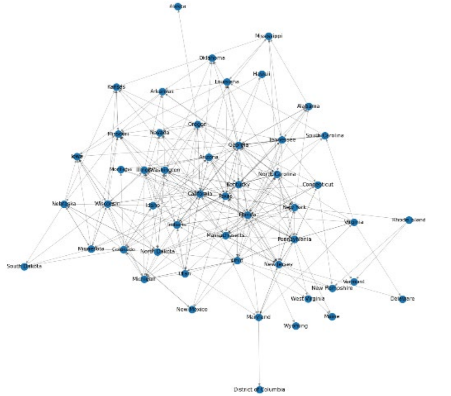
KG Construction: Commodity Flow Ontology and Network Analysis





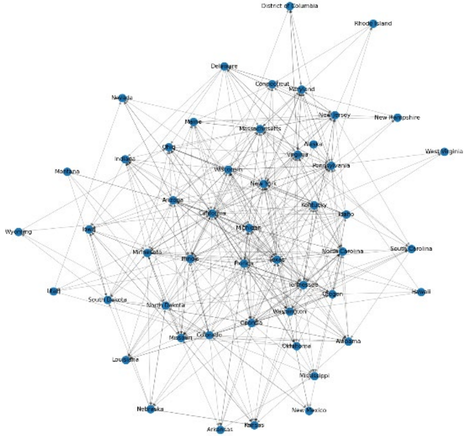
Year 2012

Resilience Score: 0.907
Efficiency Cost: 0.038



Year 2017

Resilience Score: 0.917
Efficiency Cost: 0.030



Research in Y1

Knowledge Graphs

- Multi-modal KG formalism for camera traps
- KG construction for Smart Foodsheds
- New ontology and KG for commodity flow

Federated Learning

- Federated learning with pre-trained models (under review)
- Personalized federated learning (ICLR'22)

Conversational AI

- Novel Strongly generalizable question answering over large-scale KGs (under review)
- Conversational interface for Smart Foodsheds

Model Commons

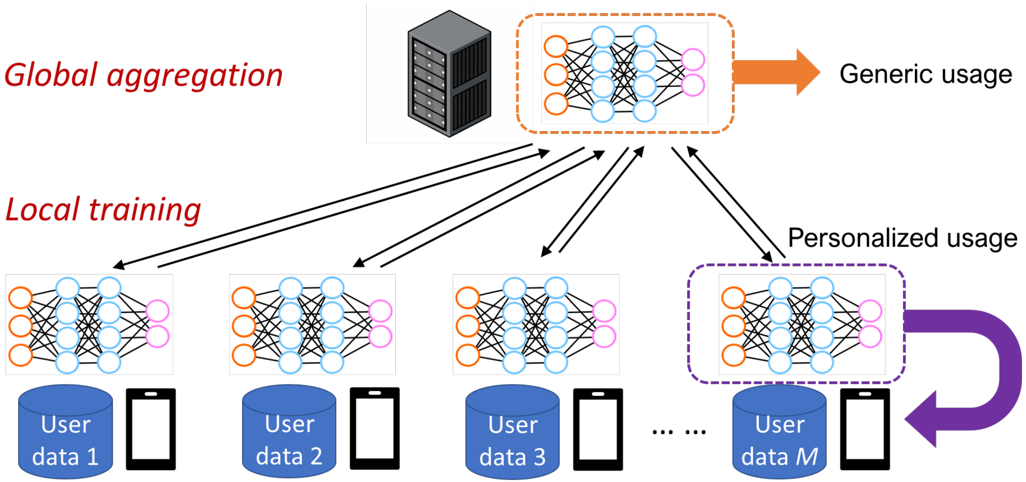
- Survey existing model and data commons for feature comparison
- Started software dev based on OSU's Data Commons

Adaptive AI

- Question synthesis for domain adaptation of question answering models (ACL'22)

* Foundational AI

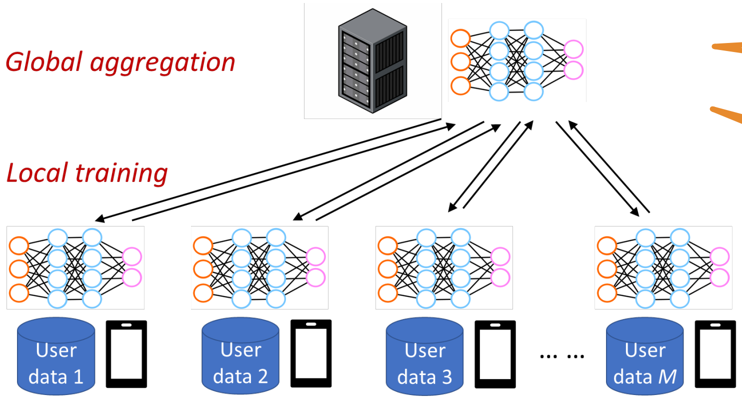
Federated Learning



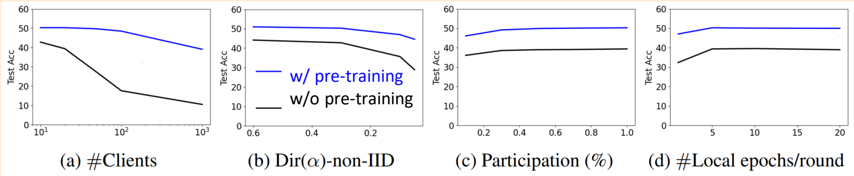
- Robust to users' heterogeneity
- Applicable to state-of-the-art AI models
- Context-aware and user-aware

	Existing literature	**ICICLE scenarios**
Model pre-training	No	**Yes (model commons)**
Server data	No	**Yes (data commons)**
Applicability	Generic OR Personalized	**Generic (updated model common)** **AND** **Personalized (individual users)**

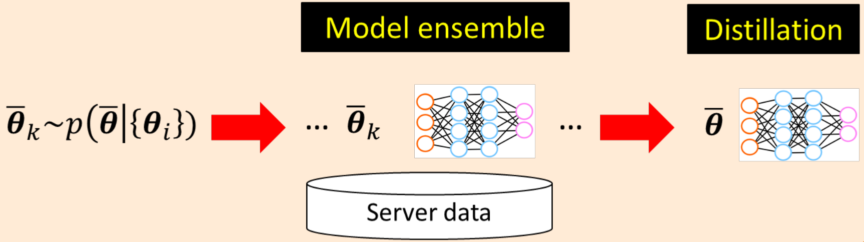
Federated Learning



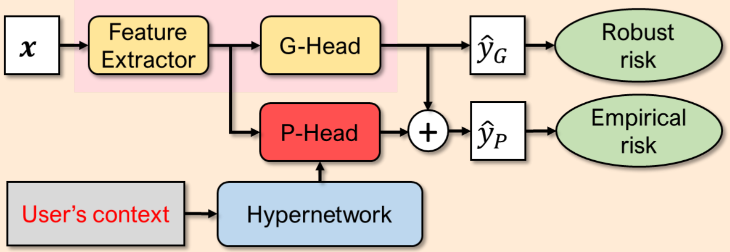
(1) Pre-training for FL



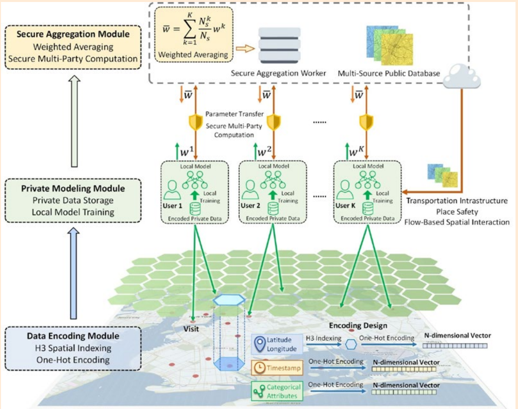
(2) Bayesian ensemble



(3) Decoupled local training



(4) Private & public data integration across geo-locations



Research in Y1

Knowledge Graphs

- Multi-modal KG formalism for camera traps
- KG construction for Smart Foodsheds
- New ontology and KG for commodity flow

Federated Learning

- Federated learning with pre-trained models (under review)
- Personalized federated learning (ICLR’22)

Conversational AI

- Novel Strongly generalizable question answering over large-scale KGs (under review)
- Conversational interface for Smart Foodsheds

Model Commons

- Survey existing model and data commons for feature comparison
- Started software dev based on OSU’s Data Commons

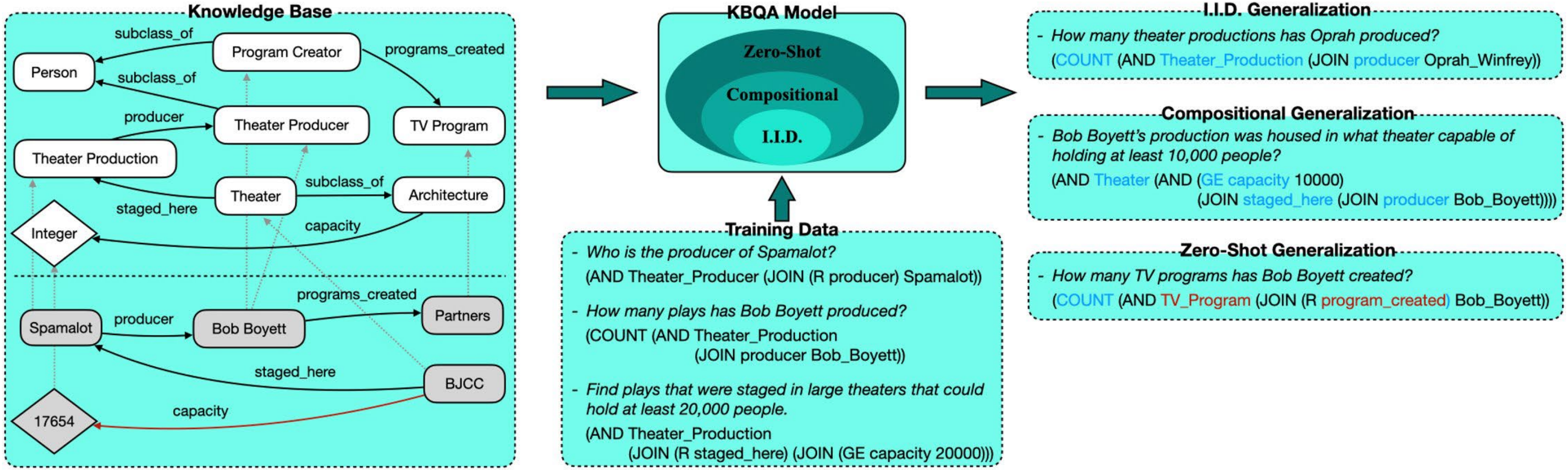
Adaptive AI

- Question synthesis for domain adaptation of question answering models (ACL’22)

* Foundational AI
* Democratization

Conversational AI (for Knowledge Graphs)

- Novel method for strongly generalizable and efficient question answering on large-scale KGs (**45 million entities** and **3 billion facts**)
- **One order of magnitude faster** than existing methods
- Achieve **73.7%** F1 score and can work on **entirely new domains** without any training data



ArcaneQA: Dynamic Program Induction and Contextualized Encoding for Knowledge Base Question Answering
Yu Gu and Yu Su. arXiv:2204.08109, 2022

Conversational AI (for Smart Foodsheds)

- Long-term goal: Rapid Conv AI prototyping using domain expertise
- Immediate project: Allow for a conversational interface which modifies parameters for Agent-based foodshed model
- NextGen: Report on simulation with generated language that compares the key metrics before and after the simulation update

Adapting implementation of few-shot semantic parsing* that converts NL utterances to parameter changes

- Uses a "canonical utterance" to bridge between NL variants and control language
- Conversion to control language using rule-based grammar

What happens if you increase car ownership in the model?

This translates to a canonical utterance (via large LMs) →

increase cars by 10%

↓ parse with grammar

```
:set param cars+=10
```

*Shin et al., Constrained Language Models Yield Few-Shot Semantic Parsers, EMNLP 21.

AI Foundations Summary

- Advances in AI Foundations include:
  - Knowledge graph projects on multimodal KGs, organizational category classification, and commodity flows
  - New models for Federated Learning that allow for pre-training, personalization
  - New Conversational AI models for large-scale knowledge graphs and interface for agent-based foodshed model
  - Adaptive AI for question answering, initial development of model commons
- **New research outcomes:**
  - **10** papers published, **6** papers under review
  - **7** presentations at top AI venues & invited talks at universities/companies
  - **4** new open-source codebases on GitHub
  - Undergraduate-led ConvAI system wins award in Amazon Alexa Prize TaskBot competition

Outline

- Brief Overview of the ICICLE Institute
- Organizational Infrastructure and Management
- Strategic and Implementation Plan
- Overall Research Directions
- **Example Accomplishments (so far)**
  - AI Foundations
  - **CI4AI**
  - Digital Agriculture
- Conclusion

CI4AI: Core Team & Project Collaborators

Investigators:
- Spyros Blanas (OSU)
- Javier Duarte (UCSD)
- Marin Kandes (SDSC)
- Dhabaleswar K. Panda (OSU)
- Rajiv Ramnath (OSU)
- Aamir Shafi (OSU)
- Joe Stubbs (TACC)
- Hari Subramoni (OSU)
- Mahidhar Tatineni (SDSC)
- Karen Tomko (OSC)
- Hongwei Zhang (ISU)
- **Zhao Zhang (TACC)**

Students:
- Nawras Alnassan (OSU)
- Quentin Anthony (OSU)
- Arpan Jain (OSU)
- Weiran Huang (TACC)
- Zhibo Meng (ISU)

Digital Agriculture
- Patrick Huber (UC Davis)
- Matthew Lange (IC-FOODS)
- Christopher Stewart (OSU)

Smart Camera Traps & Tradeoffs
- Tanya Berger-Wolf (OSU)
- Chuck Stewart (RPI)

Reference Architecture
- Rajiv Ramnath (OSU)
- Deepak Suresh (OSU)

Research Plan: CI4AI

High Perf. Training

- High-Performance Model Training

High Perf. Data Management

- Investigate the use of HDF5 VOL to decouple data access and parallel I/O

Edge Intelligence

- Performance characterization of edge devices

AI-Adaptive Edge Wireless

- Leverage rateless codi ng and multi-modal, multi-path connectivity in edge wireless to enable high-throughput (Y2)

Control and Coordination

- Intelligent Resource Management with Tapis

CI4AI – Projects Impacted

- High Perf. Training
- High Perf. Data Management
- Edge Intelligence
- AI-Adaptive Edge Wireless
- Control and Coordination

- Smart Camera Traps & Tradeoffs
- Digital Agriculture Use-inspired Research
- "Hello ICICLE"
- Reference Software Architecture
- Visual Analysis of Food Systems Knowledge Graphs
- Grocery Store Closure & Community Health

Animal Ecology


Digital Agriculture

Smart Foodsheds

* Foundational CI

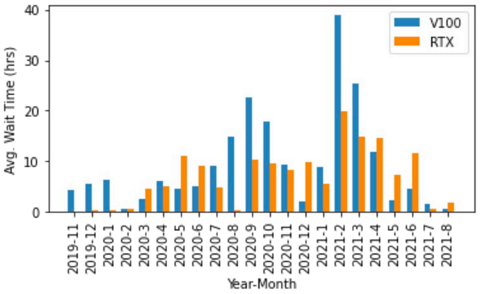
High-Performance Model Training and Support for Edge Intelligence

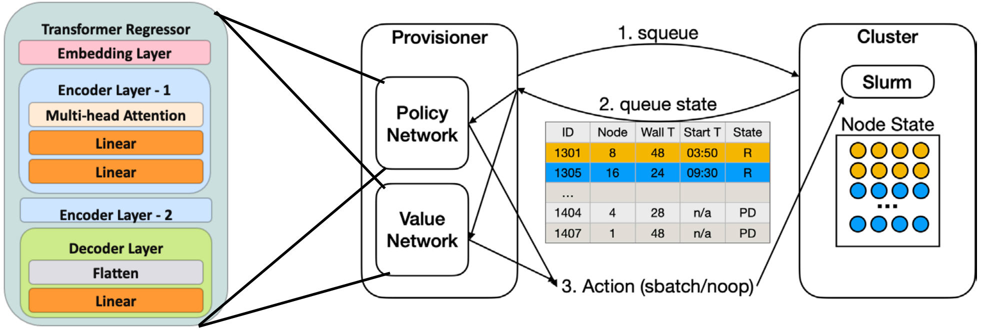
- CI4AI1: High-Performance Model Training:
  - Reduced model training time for Digital Agriculture use case with high-performance distributed DNN training using parallelism on up to 16 A100 GPUs with 11.6x improvement
  - Utilized efficient model architectures for Digital Agriculture applications including the training and evaluation of different Vision Transformer variants with various parallelism techniques
- CI4AI3: Support for Edge Intelligence:
  - Explored multiple DL frameworks, models, datasets, and benchmarks for inference characterization on edge devices including NVIDIA Jetson, Raspberry Pi 4, etc.
    - MLPerf Edge inference Benchmarks: https://mlcommons.org/en/inference-edge-20/
  - Created a benchmark using Digital Agriculture models and datasets to evaluate inference performance on edge devices:
    - Metrics: Single-stream latency, Query per seconds (QPS)

* Foundational CI

Control and Coordination for Computing

- CI4AI5.b: Intelligent Resource Management with Tapis:
  - Motivation:
    - Deep learning training (e.g., GPT-NeoX 20B) and inference service experience long interruption on batch GPU clusters due to
      - the long queue wait time when machine is heavily loaded
      - maximum wall-clock time, e.g., 48 hours
    - Predicting queue wait time then proactively submitting the next job leads to sub-optimal solutions
  - System:
    - A job provisioner that interacts with Slurm using squeue and sbatch





ID	Node	Wall T	Start T	State
1301	8	48	03:50	R
1305	16	24	09:30	R
...				
1404	4	28	n/a	PD
1407	1	48	n/a	PD

Control and Coordination for Computing

- CI4AI5.b: Intelligent Resource Management with Tapis:
  - Method:
    - Leveraging the reinforcement learning method and train a job provisioner with 20-month long job traces on TACC Frontera-RTX (RTX) and Longhorn (V100) GPU clusters
    - A transformer based neural network
    - Deep Q Learning (DQN) and Policy Gradient
  - Results:
    - Heavily loaded (avg queue wait time: [12, ∞] hours) -- enhancing the probability of interruption-free service from 0% to 25% (DQN) and 40% (Policy Gradient)
    - Medium loaded (avg queue wait time: [2, 12] hours) -- enhancing the probability of interruption-free service from 50% to 70% (DQN) and 72% (Policy Gradient)

CI4AI Summary

- Enabling fast and efficient model training for use-inspired science of digital agriculture and animal ecology
- A quantitative understanding of model inference performance on edge devices with the associated benchmark applications
- An intelligent resource provisioner on batch GPU clusters to facilitate long-running model training and inference services

Outline

- Brief Overview of the ICICLE Institute
- Organizational Infrastructure and Management
- Strategic and Implementation Plan
- Overall Research Directions
- **Example Accomplishments (so far)**
  - AI Foundations
  - CI4AI
  - **Digital Agriculture**
- Conclusion

Digital Agriculture Team Members

Investigators

Chris Stewart, OSU

Scott Shearer, OSU

Animal Ecology

Aamir Ahmad, University of Stuttgart
Tanya Berger-Wolf, OSU
Chuck Stewart, Rensselaer Polytechnic Institute

PADI

Ariful Azad, Indiana University
Vipin Chaudhary, Case Western
Beth Plale, Indiana University

CI4AI

Aamir Shafi, OSU
Hari Subramoni, OSU
P. Sadayappin, Utah

AI Foundations

Yu Su, OSU
Wei-Lun (Harry) Chao, OSU
Song Gao, University of Wisconsin
Raghu Machiraju, OSU

Smart Foodsheds

Matthew Lange, IC Foods
Alfonso Morales, University of Wisconsin

Reference Architecture

Rajiv Ramnath, OSU
Joe Stubbs, TACC
Richard Cardone, TACC

AI4CI

Aamir Shafi, OSU
Zhao Zhang, TACC

Graduate Student Members

Nawras Alnassan, OSU
Arpain Jain, OSU
Jenna Kline, OSU
Zichen Zhang, OSU
Swathi Manikya, OSU
Chris Dean, OSU
Chris Tkach, OSU
Jinghua Yan, Utah
Yufan Xu, Utah

Undergraduate Members

Seth Ockerman, Grand Valley
John Wu, OSU
Kevyn Angueirra Irizarry, OSU

Digital Agriculture – Project Impacts

Animal Ecology

Smart Foodsheds

Digital Agriculture Projects (selected)

Open-source software for sUAVs

Low latency inference at edge sites

High performance model training and model profiling

Federated learning and differential privacy for farmers

Outreach and user community

AI Foundations

CI4AI

AI4CI

Reference Architecture

* Societal Problems
* National Priorities
* Use of Plug & Play
*Democratization

Challenges and Relevance to Broad Vision

- Digital agriculture uses remote sensors, in-field equipment, and AI to boost crop yields by diagnosing and autonomously treating crop health conditions

Research vision

- Next-gen CI should enable data collection and management for large, complex, and heterogeneous data
- Next-gen CI provide edge-to-cloud support for digital agriculture

- Scientific questions for digital agriculture concern:
  - efficacy, agricultural efficiency, cost effectiveness, community-technology integration,
  - computational efficiency, data management, software-driven autonomy, and privacy

Y1 Activities

Use-Inspired AI

- **Documented 12 digital agriculture use cases. Prototype development in progress**
- **Created dataset on soybean leaf defoliation (publicly available in multiple data commons; shared in ICICLE)**
- **Created neural network architecture for leaf defoliation in soybean fields (DefoNet)**
(published in CompAg; highlighted in media [press-1, press-2], publicly avail)
- Curating multiple agricultural datasets and moving to data/model commons
- International collaborations on differential privacy for ag data and crop pose estimation for long-term studies
- Plan for outreach to agricultural experts

Y1 Activities (Cont'd)

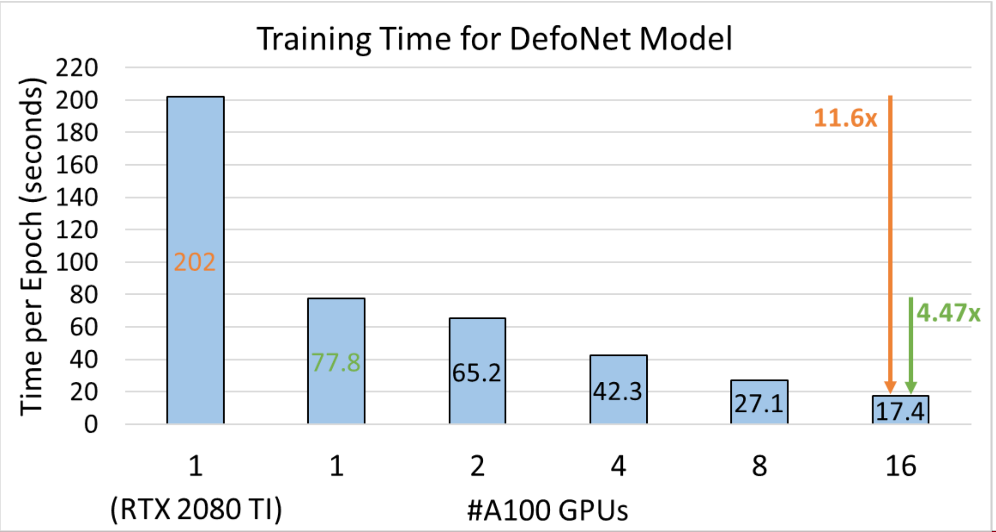
Next-Gen CI

- Low-latency inference on autonomous UAV, tractors and sprayers
- Decentralized model benchmarking for data commons
- **High Performance, Distributed Model Training for Ag Data**

Exploit high-performance distributed/parallel training algorithms [3] for ag datasets and models (e.g., DefoNet [1])

- **Edge Resource Management for Swarms of Autonomous sUAVs [2]**

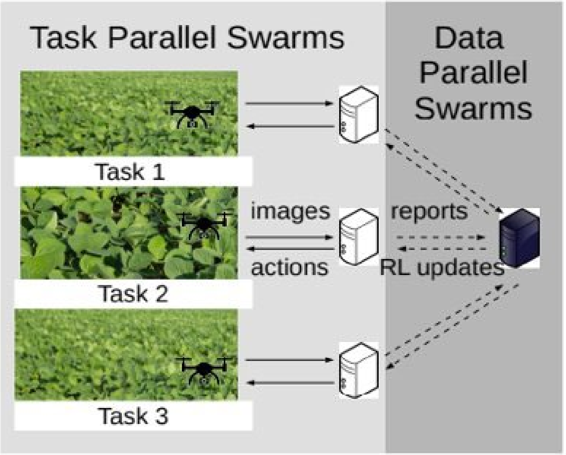
Mission goal is to create a map of severe defoliation in a soybean field using DefoNet and reinforcement learning



Y1 Activities (Cont.)

Next-Gen CI

- Low-latency inference on autonomous UAV, tractors and sprayers
- Decentralized model benchmarking for data commons
- **High Performance, Distributed Model Training for Ag Data**
  Exploit high-performance distributed/parallel training algorithms [3] for ag datasets and models (e.g., DefoNet [1])
- **Edge Resource Management for Swarms of Autonomous sUAVs [2]**
  Mission goal is to create a map of severe defoliation in a soybean field using DefoNet and reinforcement learning



Hypothesis: Should UAV in the same field communicate exchange data?

Digital Agriculture Summary and Highlights

- **21 ICICLE faculty and staff participants across 10 institutions**
- **2 graduate students in non-traditional CSE; 3 students from URG**
- **Research in progress across a range of directions**
  - Community to agricultural efficiency to data management to autonomy
- **Publicly available datasets, models, source code, and prototypes**

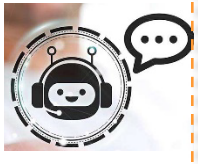
1. Zichen Zhang, Sami Khanal, Amy Raudenbush, Kelley Tilmon, Christopher Stewart, *Assessing the efficacy of machine learning techniques to characterize soybean defoliation from unmanned aerial vehicles*, Computers and Electronics in Agriculture, 2022
2. Jayson Boubin et al., *Data-Parallel Versus Task-Parallel Swarms for Small Unmanned Aerial Systems*, IEEE IOTDI, 2022
3. Arpan Jain, Tim Moon, Tom Benson, Hari Subramoni, Sam Ade Jacobs, Dhabaleswar K Panda, and Brian Van Essen, *SUPER: SUb-Graph Parallelism for TransformERs*, IEEE IPDPS 2021

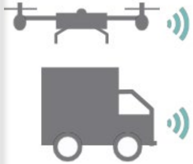
Outline

- Brief Overview of the ICICLE Institute
- Organizational Infrastructure and Management
- Strategic and Implementation Plan
- Overall Research Directions
- Example Accomplishments (so far)
  - AI Foundations
  - CI4AI
  - Digital Agriculture
- **Conclusion**

Designing Intelligent CI for Computing Continuum

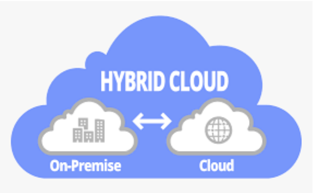
Conv. AI



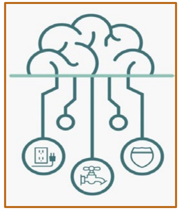


Data: On Field Sensors

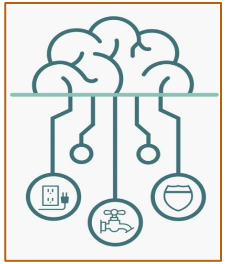
Models: Edge & Near Edge



Data/Models: Clouds



Data/Models: HECs



ICICLE-enabled Computing Continuum

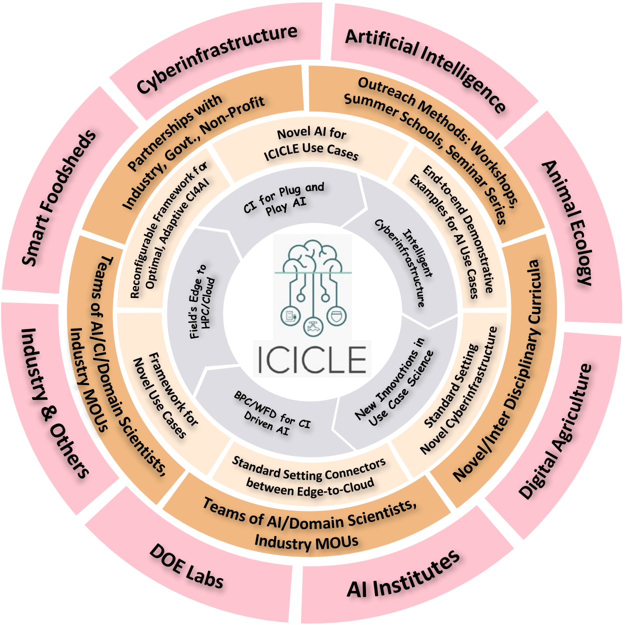
Y1 Highlights

- Significant progress in 9 months
- Engaged with the three use-inspired science domains
- Solutions to foundational and translational challenges are being worked out
  - in individual thrusts/projects
  - across thrusts/projects
- These solutions are also being integrated into CI components
- Significant progress on WFD, BPC and CKT
- On track or ahead for Y1

Themes in Y2

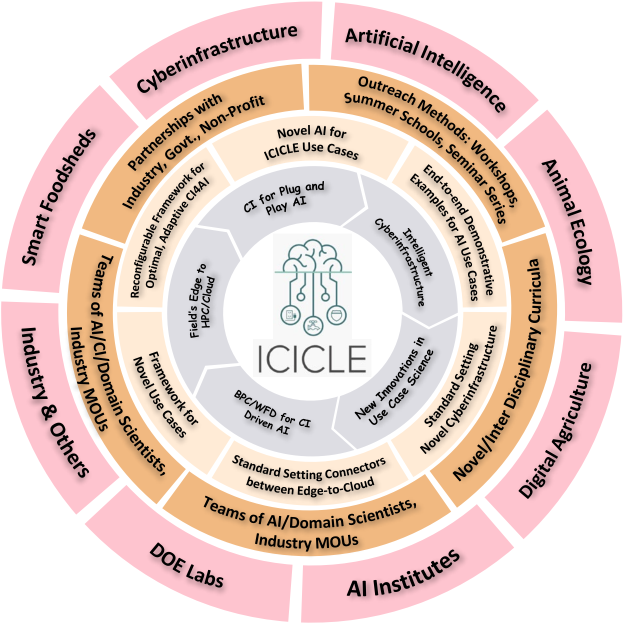
1. Converging and integrating thrusts and projects
   - Use case generalization to identify common ICICLE components
   - Development of plug-and-play components
   - Capturing in the ICICLE Reference Architecture
2. Progressing towards democratizing AI
   - Demonstrating examples of plug-and-play AI across a set of application domains
3. Broadening Impact
   - Incorporation of BPC and knowledge transfer initiatives
   - Developing a framework for democratization
4. Establishing ICICLE as a national resource
   - Processes for systematic evaluation of AI advances in the field
   - Bringing in other communities and their applications
   - Initiating partnerships with other AI institutes, ERCs and FFRDCs
   - Translation of ICICLE components to national CI

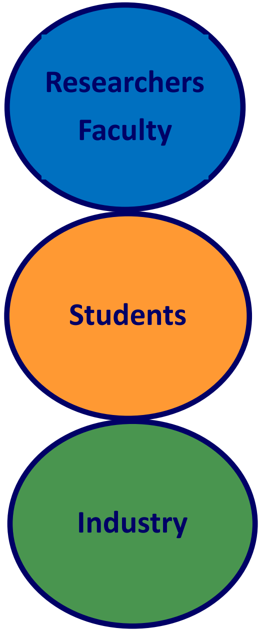
ICICLE Enabling Global AI leadership

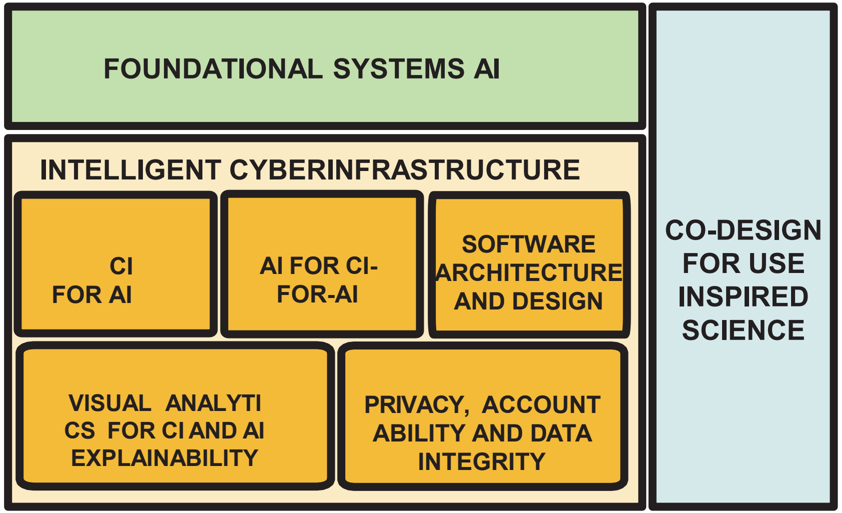


- Integrate into the National CI Ecosystem
- Integrative and Interoperable
- Leverages existing recognized capabilities
  - Centers of Excellence, AI Institutes, Large Facilities
- Collaborative
- Sustainable and Inclusive
  - Workforce Development, Broadening Participation, Collaboration and Knowledge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

Engaging With ICICLE







Co-develop & Adopt ICICLE developed CI!

Contact: panda@cse.ohio-state.edu