



Creating Intelligent Cyberinfrastructure for Democratizing AI



<http://icicle.ai>

Beth Plale
Indiana University



Follow us
@icicleai

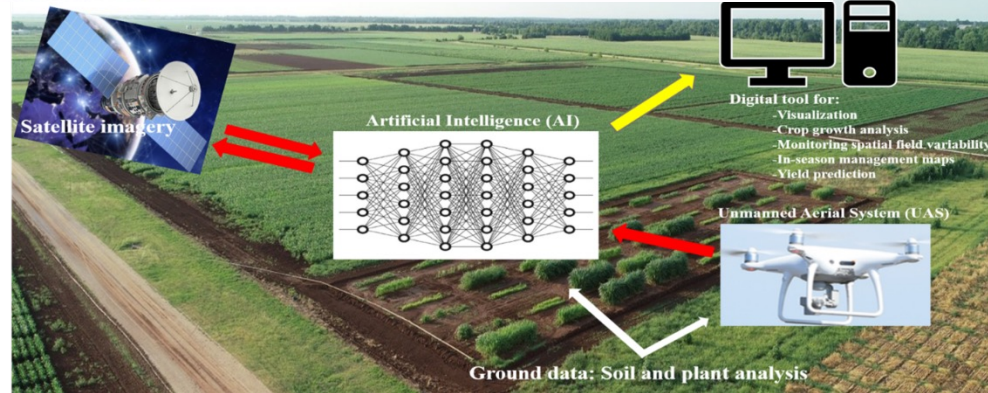
Outline

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

Societal Challenge (Example #1): Agriculture

- **Food security/sustainability in 2050**
 - 9.8B people, climate; 0.5x arable land per cap vs 1985
 - Wide gains in crop management needed (typical yields fall 3X below best practice)
- **Sustainable agricultural workforce**
 - The next generation of agriculture professionals will include engineers, computer scientists, data scientists
- **Democratization of digital agriculture capabilities**
 - Autonomous unmanned aerial vehicles, self-driving tractors and sprayers, fertilizer and seed recommendations
 - Big and small farms, staple and specialty crops, underrepresented communities
 - Privacy and ethical considerations

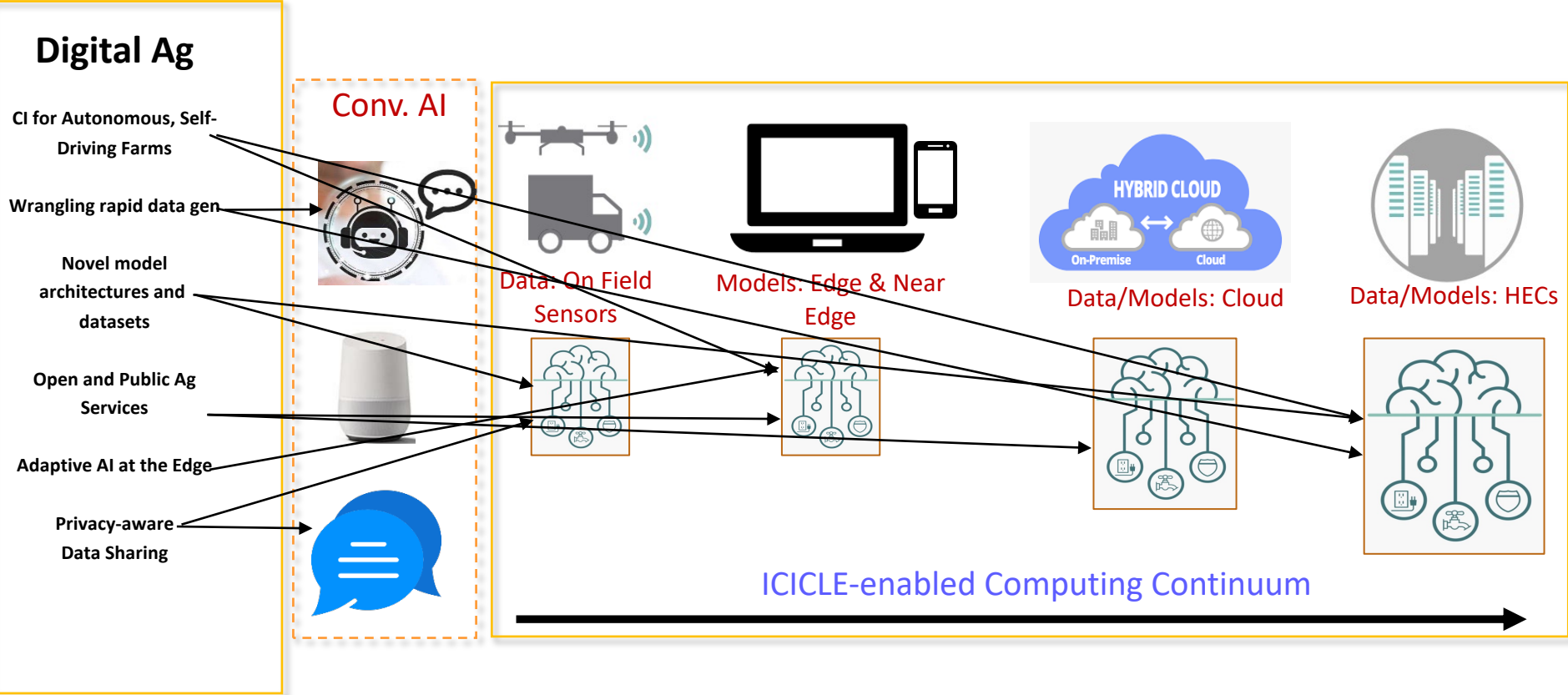
AI-Driven Digital Agriculture



<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

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

How AI can Help the Users of these Science Cases?

For the three use-inspired science cases:

- Massive and ever-growing gap between AI and its accessibility to the 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.

CI's complexity to deploy AI impedes research discoveries and innovations!

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?**

The Vision

A national infrastructure that will:

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

Objectives: Intelligent CyberInfrastructure for Computing Continuum



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

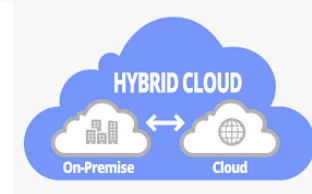
AI for "CI for AI"



On Field Sensors



Edge & Near Edge



Clouds

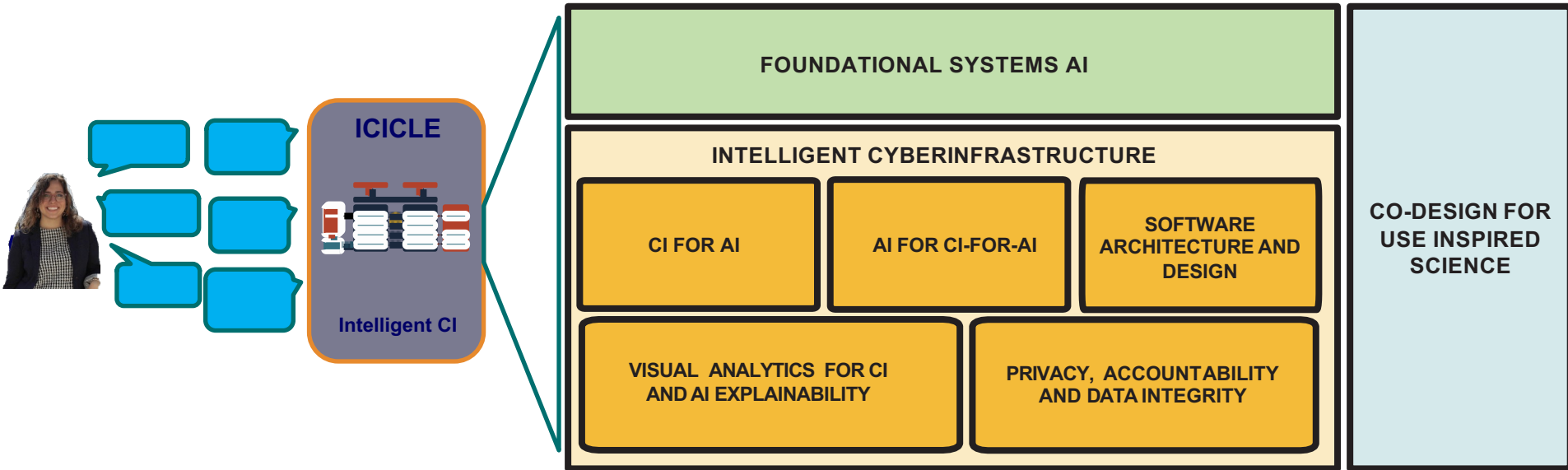


HPC Systems & Data Centers

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

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

Knowledge Graphs

- Multimodal KG to encode & reason rich data modalities (e.g., camera trap)
- Auto construction
- Interplay with LLM and knowledge-based QA

Model Commons

- MINT to support ICICLE use cases, KG, and models
- Precise profiling
- Flex composition
- Versioning and provenance

Adaptive AI

- Context-aware
- Efficient update
- User-friendly adaptation process
- Adaptation of foundation models, conversational AI

Federated Learning

- Heterogeneity
- Context-aware
- Privacy-preserving and robustness
- Going beyond classification (GNN, foundation models)

Conversational AI

- KG- and model-commons-aware
- LLM-powered
- Grounding LLMs to the context
- Hallucination reduction
- Complex reasoning

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

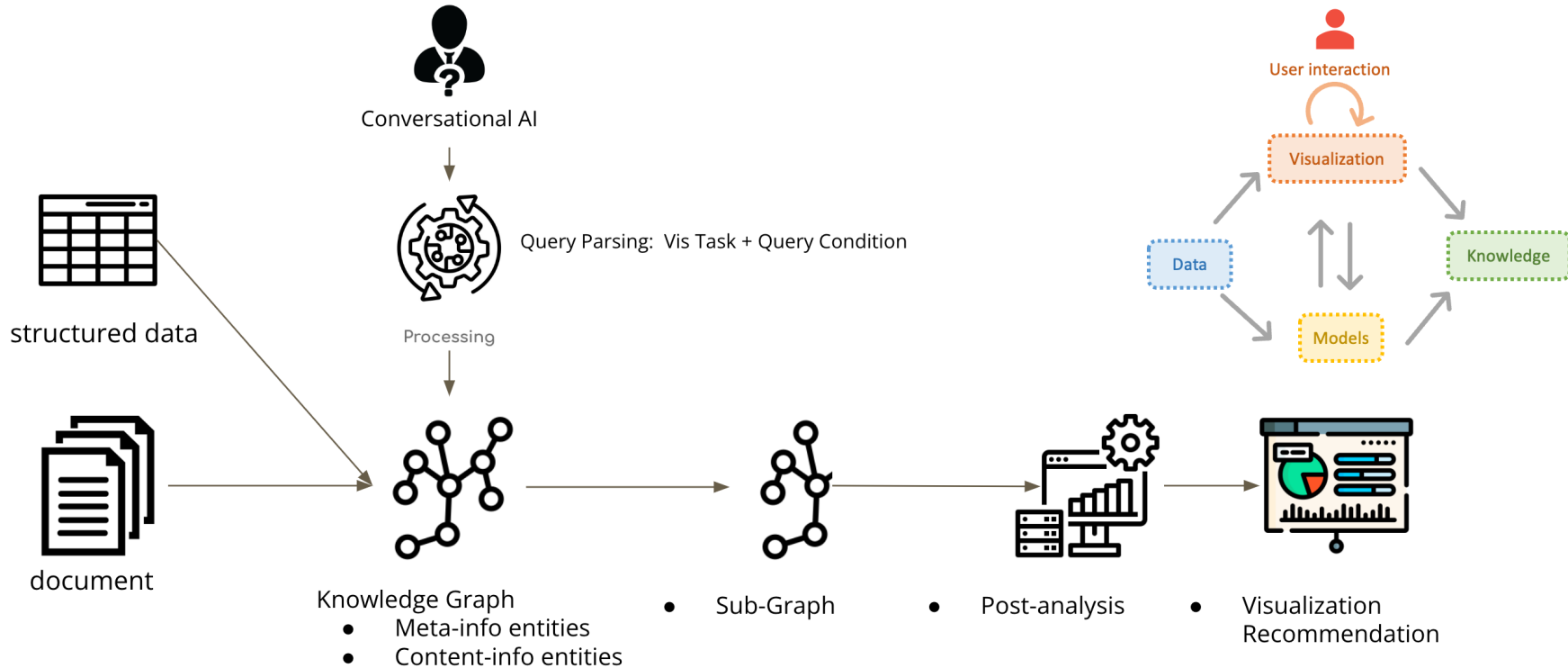
Systems

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

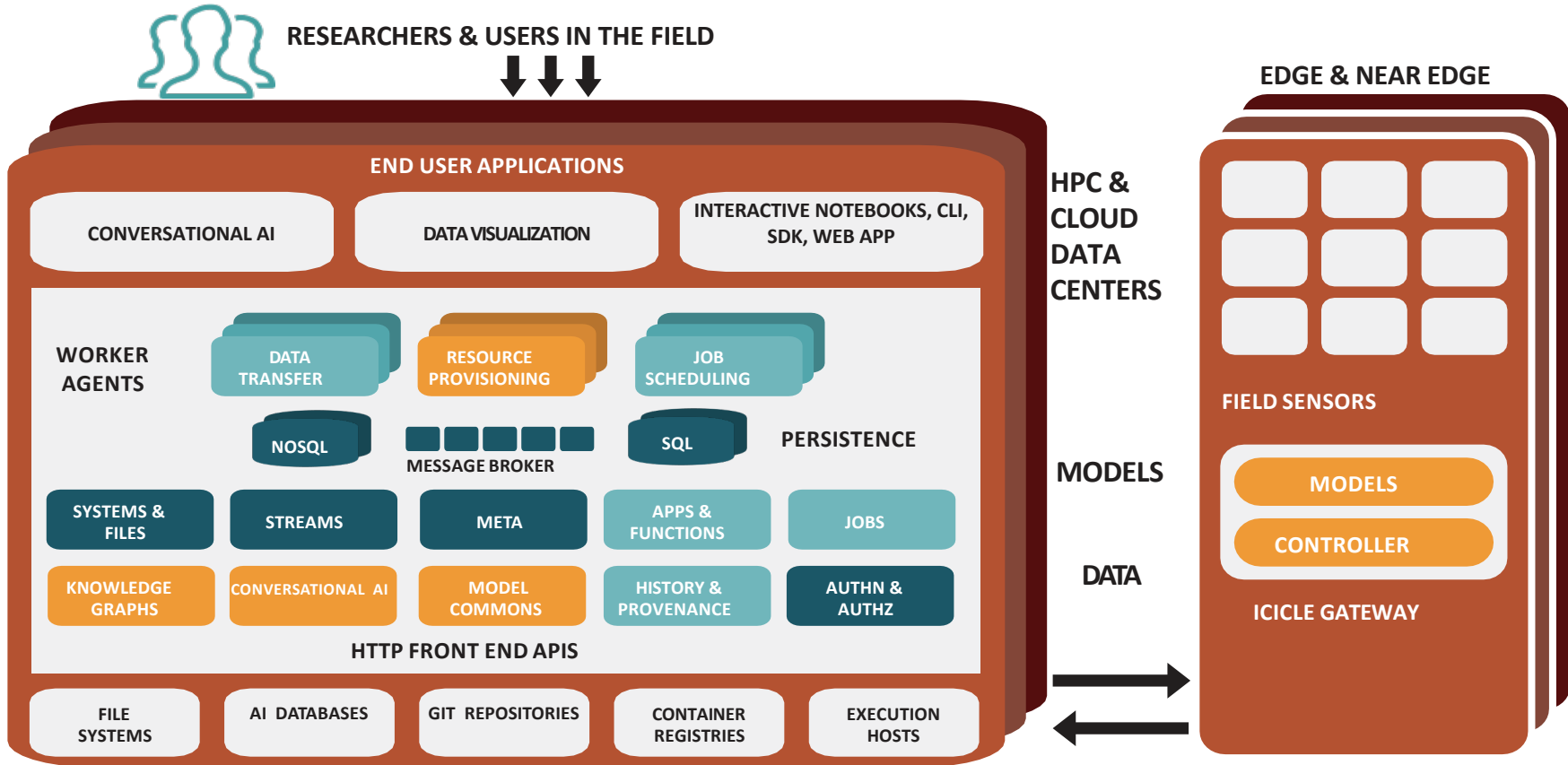
Thrust: Privacy, Accountability and Data Integrity (PADI)

- 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



The Deliverable: The ICICLE Software Stack

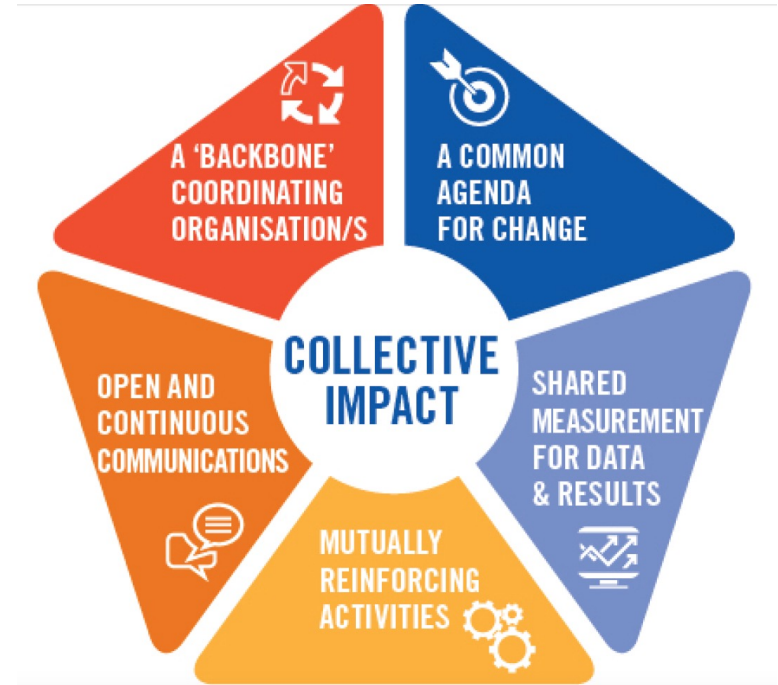


Broader Impacts Backbone Network (BIBN)

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 (demo)**
 - **Smart Foodsheds (demo)**
 - **Grocery Store Closure (demo)**
 - **AI4CI**
 - **BIBN**
- 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

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

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

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?

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.

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.

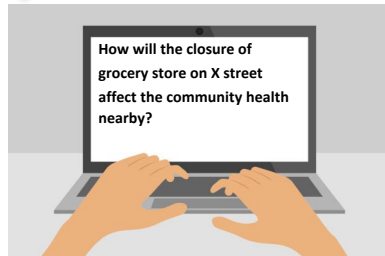
SCENARIO



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.

Selected Accomplishments from Broader Engagement

- BPC
 - Inclusive environments initiative: **ICICLE NextGens** group, **ICICLE Code of Conduct**
 - Building awareness: bi-weekly **Ally tips (bias)**; AI Ethics tips purposed for Indiana Univ **K-12 summer camp**
- WFD
 - **Hello ICICLE**: clients (Notebooks, command line, python, Web client) for testing software
 - Summer 2023 launch of **ICICLE AI Ethics tips** series of 6 videos
 - Consolidation and organization of **ICICLE Publication and Training Resources** (with WFD and HelloICICLE)
- KT
 - ICICLE Seminar Series
 - **Partnership Agreements** for stakeholders to engage with ICICLE. (Students, Academic Scholars, Organizations, Industry Sponsored, and Stakeholder Roundtable)
 - Engaging stakeholders, including through 2023 class of 5 **ICICLE Educational Fellows**

Outline

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

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/mailling-lists>)
 - icicle-announce
 - icicle-discuss

Thank You!