LLM-Orchestrated Digital Twins for Safe, Human-Centered Decision Support in Precision Agriculture*

Charles Cao[1], Sajal K. Das[2], Jie Zhuang[1], Robert Davis[3]

[1]University of Tennessee
Email: cao@utk.edu, jzhuang@utk.edu
[2]Missouri University of Science and Technology
Email: sdas@mst.edu
[3]University of Tennessee Health Science Center
Email: rdavis88@uthsc.edu

Abstract

Digital technologies in agriculture (drones, IoT sensors, satellite imagery, and precision machinery) have created an opportunity for AI systems that not only forecast outcomes but also recommend and safely execute field-level interventions. We present a digital twin-based decision support framework in which a tool-using Large Language Model (LLM) agent operates over a dual twin of crop–soil dynamics and farm operations. Farmers express high-level goals and constraints in natural language (e.g., "reduce fungicide use by 10% without risking major outbreaks"), which are compiled into machine-readable *Field Management Cards* that encode multi-season objectives, safety constraints, and regulatory limits. The LLM agent plans candidate interventions (e.g., modified spray schedules, irrigation adjustments) by invoking twin services and farm asset APIs (sprayers, drones, sensors) through a standardized tool interface. Digital twins enforce feasibility and safety before recommendations are surfaced or actions are dispatched. Through comprehensive evaluation on 85 test cases, we demonstrate that the system achieves 87.5% accuracy in mapping natural language queries to appropriate tool calls and 88% success rate in enforcing safety constraints such as pre-harvest intervals and irrigation limits. Critically, the digital twin successfully repairs 100% of detected violations, ensuring fail-safe operation. Our results illustrate a path toward robust, human-centered AI assistants that connect agricultural data, domain models, and actuation in a single decision loop while maintaining safety and regulatory compliance.

Introduction

Agriculture faces unprecedented challenges from climate variability, increasing pest pressures, and the need to optimize resource use while maintaining productivity (Zhuang et al. 2023; Sun et al. 2022). Digital technologies, including IoT sensors, drones, satellite imagery, and mobile platforms, generate vast streams of heterogeneous, temporally rich data that offer opportunities for AI-enabled precision farming (Sarkar et al. 2024; Gupta et al. 2024). However, many existing AI solutions for agriculture remain narrow in focus, difficult to generalize, and often lack the transparency and human-centered design necessary for farmer adoption.

Current decision support systems typically focus on prediction (yield forecasting, disease detection) without closing the loop to safe, explainable actuation. Farmers need systems that can understand high-level goals expressed in natural language, reason about complex tradeoffs between yield, resource use, and environmental impact, and recommend interventions that respect safety constraints and regulatory requirements. This requires bridging the semantic gap between long-horizon objectives ("minimize pesticide use this season") and short-horizon actions ("reduce spray rate by 15% on field A3 tomorrow").

In this work, we present a digital twin-based framework that addresses these challenges through three key contributions. First, we develop an architecture combining dual digital twins that model both crop–soil dynamics and farm operations, orchestrated by an LLM agent that translates natural language goals into validated interventions through a standardized tool interface. Second, we design a user-facing interface enabling plan–compile–execute–audit workflows with graduated autonomy levels, allowing farmers to choose between autonomous execution with checkpoints, human-in-the-loop approval, or advisory-only modes. Third, we provide comprehensive evaluation demonstrating 87.5% accuracy in natural language intent grounding and 88% success rate in safety constraint enforcement, with 100% repair rate for detected violations, validating the feasibility of safe LLM-based agricultural decision support.

Background and Related Work

Digital Twins in Agriculture. Digital twins (virtual replicas of physical systems) have been successfully deployed in manufacturing and aerospace for predictive maintenance and optimization (Verdouw et al. 2021). In agriculture, digital twin concepts have been applied to greenhouse control and irrigation management (Roy et al. 2020), but most focus on single aspects rather than integrated farm-level decision making. Our dual-twin approach combines crop–soil biophysical models with operational constraints of farm machinery and actuators.

*Accepted at the First International Workshop on AI in Agriculture (Agri AI), co-located with AAAI 2026.

AI Decision Support in Precision Agriculture. Machine learning has been widely applied to crop yield prediction, disease detection, and variable-rate application (Betti Sorbelli et al. 2024; Feng et al. 2022). However, these systems typically operate in isolation without considering multi-objective tradeoffs or safety constraints. Recent work on reinforcement learning for irrigation shows promise but lacks the interpretability needed for farmer trust.

LLM Agents and Tool Use. Large Language Models have demonstrated capability in tool use and planning tasks (Li et al. 2024; Cao, Zhuang, and He 2024). Recent frameworks like Model Context Protocol (MCP) enable LLMs to interact with external systems through standardized APIs. However, deploying LLM agents in safety-critical cyber-physical systems like farms requires careful constraint enforcement; our twin-based validation addresses this gap.

System Architecture

Our framework transforms the agricultural discovery loop into a disciplined workflow that compiles long-horizon goals into short-horizon, safety-ensured actuator actions. We adapt principles from automated scientific discovery to the agricultural domain across three synergistic layers. Figure 1 illustrates the complete system architecture, showing the LLM-orchestrated agent layer, dual digital twin platform, safety validation workflow, and user interaction timeline.

Farm Digital Twin Architecture

We implement a dual digital twin design that captures both the biophysical and operational aspects of farming systems:

Crop–Soil Twin. This twin encodes agronomic knowledge through integrated process models that capture the complex interactions between crops, soil, weather, and management practices. The crop growth models simulate phenology, biomass accumulation, and water stress responses based on environmental conditions and management interventions. Soil water balance and nutrient dynamics models track moisture levels, nutrient availability, and their impacts on crop development. Disease pressure models incorporate weather data, spore load estimates, and host susceptibility to predict infection risks and treatment efficacy (Cao et al. 2024). The twin also enforces critical constraints such as maximum chemical application rates, pre-harvest intervals for food safety, and environmental thresholds that prevent operations under adverse conditions.

Operations Twin. This twin models the physical capabilities and limitations of farm machinery and infrastructure that execute field operations. It captures actuator capabilities including sprayer boom widths, drone flight endurance, and irrigation system capacities that determine feasible application rates and coverage patterns. Operational constraints such as no-fly zones for drone operations, equipment availability windows, and labor scheduling requirements are encoded to ensure proposed actions can actually be executed. The twin maintains response surfaces and calibration states for each piece of equipment, tracking factors like nozzle wear that affect application accuracy. Communication latencies and potential failure modes are also modeled to ensure the system can handle real-world uncertainties in equipment performance and connectivity.

Together, these twins define the *feasible action space* for any given field state and forecast conditions.

Table 1: LLM Agent Tool Interface

API	Details
Read-Only Tools	
`query_twin()`	Retrieve current field state estimates from biophysical models
`forecast()`	Predict future conditions based on weather and growth models
`check_constraints()`	Validate proposed actions against safety and regulatory rules
`simulate()`	Run what-if analysis to predict intervention outcomes
Write Tools (Twin-Validated)	
`sched_irrigation()`	Schedule irrigation event subject to water budget constraints
`adjust_spray_rate()`	Modify chemical application rate within approved limits
`plan_drone_survey()`	Request aerial imagery collection for specified fields

LLM Agent and Tool Interface

The LLM agent interacts with the farm digital twin through a RESTful API that exposes farm capabilities as JSON-RPC style function calls, following a strict separation between read-only tools for querying and analysis versus write tools that trigger validated field-level interventions. Table 1 summarizes the available tools. All parameters are validated against predefined schemas for type safety, and return values include confidence scores and metadata enabling the LLM to reason about information quality. The design prevents direct equipment control, instead requiring all actions to flow through twin validation layers.

All write operations follow a validate-then-execute pattern where the digital twin first checks the proposed action against hard constraints (equipment limits, safety thresholds, regulatory requirements) and soft preferences (cost targets, sustainability goals). Only actions passing all validation checks enter the approval queue, where they may require human confirmation depending on the configured autonomy level. This architecture ensures that LLM hallucinations or misinterpretations cannot directly trigger unsafe farm operations, instead failing safely at the validation layer.

Natural Language Workflow Pipeline

Operations follow a four-phase pipeline that exposes AI as a service while preserving human oversight:

Plan. Users submit objectives in natural language, binding agricultural contexts (crop type, growth stage, field history) to operational constraints (budget limits, equipment availability, regulatory requirements). For example: “Re-

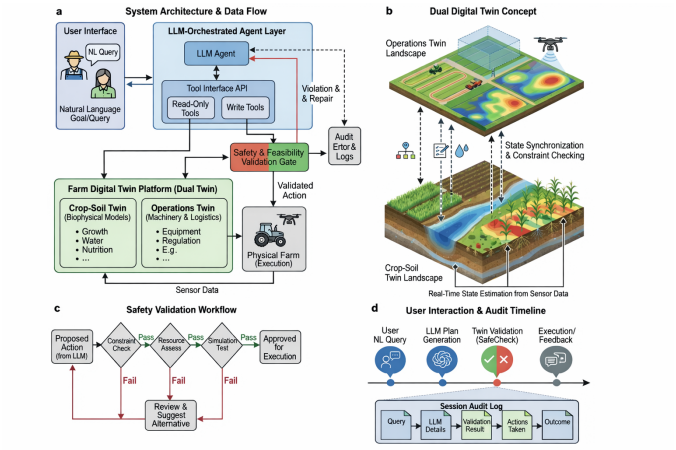

Figure 1: Complete system architecture and workflow. (a) LLM-orchestrated agent layer with tool interface API separating read-only and write tools, along with validation and repair mechanisms; (b) Dual digital twin concept showing both operations twin (landscape) and crop-soil twin (underground) with real-time state estimation from sensor data; (c) Multi-stage safety validation workflow with constraint checking, resource checking, simulation, and repair/alternative generation capabilities; (d) User interaction and audit timeline illustrating the plan-compile-execute-audit cycle with session logging.

duce fungicide use by 10% on low-risk fields while keeping disease index below 0.35."

Compile. Natural language objectives are compiled into *Field Management Cards*, machine-readable documents encoding:

- Multi-objective reward functions (yield, resource use, environmental impact)
- Hard constraints (regulatory limits, safety thresholds)
- Soft preferences (risk tolerance, sustainability goals)

The compilation process leverages both physics-based models and learned preferences from historical decisions.

Execute. The workflow engine orchestrates farm operations through three modes:

1. **Autonomous with checkpoints**: Agent executes approved actions, pausing at predefined gates
2. **Human-in-the-loop**: Each action requires explicit farmer approval
3. **Advisory only**: Agent provides recommendations without execution capability

Audit. Each decision generates an immutable audit trail including:

- Original user query and compiled Management Card
- Twin state at decision time
- All simulation runs and constraint checks
- Final action plan with rationales
- Actual execution results and deviations

This provenance enables both real-time monitoring and post-season analysis for continuous improvement.

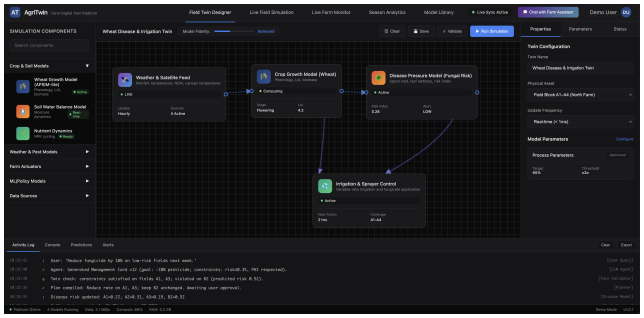

Figure 2: Field Twin Designer interface with workflow editor, connected data sources, and natural language interaction log.

User Interface and Interaction Design

We developed a web-based interface that makes the digital twin framework accessible to farmers and agronomists through four integrated views. Figure 2 shows the Field Twin Designer view, which provides the primary interface for configuring workflows and interacting with the digital twin through natural language.

Field Twin Designer

This visual workflow editor provides farmers and agronomists with an intuitive interface for configuring how data flows through models to generate decisions. Users can connect diverse data sources including real-time weather feeds, soil sensor networks, and satellite imagery streams to build comprehensive situational awareness. The interface supports configuration of model pipelines where outputs cascade through processing stages, such as crop growth predictions feeding into disease risk assessments that ultimately inform spray timing decisions. Farmers

define Management Cards that encode their specific objectives and constraints, balancing factors like yield targets, resource budgets, and sustainability goals. The system allows flexible configuration of autonomy levels and approval gates, enabling farmers to specify which decisions require human review based on their risk tolerance and regulatory requirements.

Live Simulation and What-If Analysis

This view enables rapid exploration of intervention scenarios through interactive simulation capabilities. When a farmer poses questions like "What if I reduce irrigation by 20% next week?", the system initiates a comprehensive analysis pipeline. The natural language query is first parsed into a concrete action plan with specific parameters and timing. The digital twin then runs parallel simulations across all affected fields, accounting for spatial variability in soil types and current moisture levels. The interface displays predicted impacts on critical variables including soil moisture trajectories, crop stress indices, and projected yield impacts over multiple time horizons. Any constraint violations, such as moisture dropping below permanent wilting point or exceeding stress thresholds during critical growth stages, are immediately highlighted with visual warnings. If the original plan proves infeasible, the system automatically generates and simulates alternative scenarios that achieve similar objectives while respecting all constraints. Real-time visualization overlays irrigation schedules with weather forecasts and risk indices, enabling farmers to see how proposed changes interact with expected environmental conditions.

Live Farm Monitor

The monitoring dashboard provides real-time situational awareness through interactive maps displaying soil moisture, disease risk heatmaps, and active intervention tracking. Sensor data from IoT devices deployed across fields is continuously ingested and processed, with the digital twin automatically reconciling measurements with model predictions to detect discrepancies that may indicate sensor drift or unexpected field conditions. The system performs continuous anomaly detection using statistical process control methods, alerting farmers to unusual conditions like dry patches developing faster than predicted or pest pressure hotspots emerging in unexpected locations. Each alert includes contextual information such as historical patterns for that field location, nearby sensor readings, and recommended investigation or response actions. The interface visualizes ongoing interventions with real-time status updates, showing which equipment is currently operating, progress toward completion, and any deviations from planned application rates or coverage patterns. Complete activity logs capture all system actions, user decisions, and override events for operational transparency and regulatory compliance, with timestamps, GPS coordinates, and associated sensor readings automatically recorded for each field operation.

Season Analytics

Post-season analysis tools aggregate performance metrics including yield, resource use, and input costs across all fields and interventions, enabling comparative evaluation of AI-guided decisions against baseline approaches and historical benchmarks. The analytics dashboard provides multi-dimensional performance visualization, breaking down outcomes by field characteristics, intervention types, and environmental conditions to identify which management strategies were most effective under different circumstances. Resource efficiency metrics quantify water use per unit yield, chemical application rates versus disease pressure outcomes, and energy consumption for mechanical operations, with automated identification of opportunities for optimization in future seasons. The system tracks decision quality through multiple indicators including constraint violation rates that measure safety compliance, prediction accuracy that compares forecasted outcomes against actual measurements, and user override patterns that reveal where human judgment disagreed with AI recommendations. Machine learning models analyze these override patterns to identify systematic biases or knowledge gaps, feeding insights back into model refinement for continuous improvement. Economic analysis tools calculate return on investment for different intervention strategies, accounting for input costs, labor requirements, yield impacts, and market prices to support data-driven budget planning for subsequent growing seasons.

Table 2: Natural Language Intent Grounding Performance

Query Category	Cases	Match Rate	Avg Score
Complex Query	6	100.0%	1.00
Decision Support	6	100.0%	1.00
Multi-field Query	2	100.0%	0.50
Resource Adjustment	12	100.0%	1.00
Safety Compliance	8	87.5%	0.88
Status Query	14	100.0%	0.82
What-if Simulation	12	75.0%	0.75
Overall	60	93.3%	0.88

Evaluation

We evaluate our framework through 85 test cases assessing natural language intent grounding (60 cases) and safety constraint enforcement (25 cases). The virtual farm digital twin evaluation suite includes 13 tool capabilities spanning query, simulation, and control functions across realistic agricultural scenarios.

Metrics

For intent grounding, we measure partial match rate (binary success for correct tool selection) and average match score (continuous quality metric from 0.0 to 1.0 reflecting both tool selection and parameter accuracy). For safety enforcement, we track detection rate, repair rate, and overall success rate in identifying and correcting violations.

Results

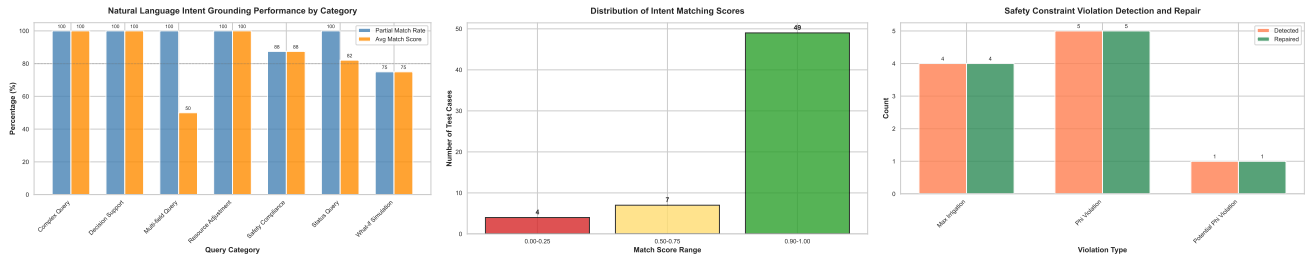

Figure 3: Evaluation results: (a) Intent grounding performance by category; (b) Intent matching score distribution; (c) Safety violation detection and repair by type.

Table 3: Safety Constraint Enforcement Performance

Violation Type	Cases	Detected	Repaired	Success
Low Moisture	1	0/1	0/0	100%
Max Irrigation	4	4/4	4/4	100%
Max Spray Rate	2	0/2	0/0	0%
Null	11	0/11	0/0	100%
PHI Violation	6	5/6	5/5	83%
Pot. PHI Violation	1	1/1	1/1	100%
Overall	25	10	10	88%

Natural Language Intent Grounding Our system achieved 87.5% average intent matching score with 93.3% partial match rate across 60 queries. As shown in Table 1 and Figure 3(a), performance was strongest on resource adjustments and decision support (100%), representing common farmer interactions. Lower scores on what-if simulations (75%) and multi-field queries (50%) primarily stemmed from parameter naming inconsistencies rather than intent misunderstanding.

Figure 3(b) shows a bimodal distribution where the system either performs nearly perfectly or fails completely, with 81.7% of cases achieving scores of 0.90–1.00.

Safety Constraint Enforcement The digital twin safety system achieved an 88% overall success rate across 25 test scenarios designed to trigger various constraint violations. As detailed in Table 2 and Figure 3(c), the system demonstrated particularly strong performance on irrigation limit enforcement, detecting and repairing all four test cases where excessive water application was requested. This perfect performance on irrigation constraints is critical for preventing equipment damage and water waste.

Pre-harvest interval violations, which represent one of the most important regulatory constraints in agricultural chemical application, were handled with 83% success rate. The system detected five out of six attempted PHI violations and successfully repaired all detected cases by either delaying the application or suggesting alternative treatments. The single missed detection occurred in an edge case where the query was ambiguous about timing, highlighting the importance of clear communication interfaces in safety-critical systems.

A crucial finding is that the system achieved a 100% repair rate for all detected violations, meaning that once a safety issue was identified, the digital twin always successfully generated a safe alternative action. This demonstrates the effectiveness of the twin-based approach where the system can simulate alternatives and verify their safety before proposing them to the farmer.

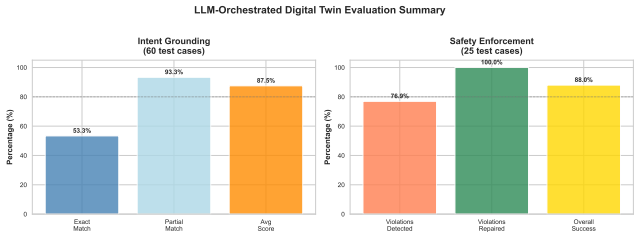

Figure 4: Overall evaluation summary showing intent grounding (93.3% partial match rate, 87.5% average quality) and safety enforcement (76.9% violation detection, 100% repair rate).

System Performance Summary Figure 4 provides a comprehensive view of system performance across both evaluation dimensions. The intent grounding results show that while only 53.3% of cases achieved exact parameter matches, 93.3% achieved at least partial matches with correct tool selection, and the overall average quality score of 87.5% indicates strong performance. The gap between partial match rate and average score is relatively small at 5.8%, suggesting that even partial matches tend to be of high quality rather than barely passing the threshold.

The safety enforcement results reveal a critical system property where detection is the primary challenge while repair is highly reliable. The 76.9% detection rate indicates room for improvement in identifying all potential violations, particularly those expressed through ambiguous natural language. However, the 100% repair rate for detected violations provides confidence that the system fails safely when issues are identified. This architecture ensures that safety violations that make it past the detection layer are extremely unlikely, as they would require both a detection failure and the absence of downstream validation.

Qualitative Analysis

To better understand system behavior, we examine representative examples from each evaluation category. For successful intent grounding, consider the query "Check moisture on field A1, then simulate adding 10mm irrigation." The system correctly decomposed this into two sequential operations, first invoking query_moisture with field_id="A1", then simulate_irrigation with field_id="A1" and amount_mm=10. This demonstrates the system's ability to maintain context across multi-step operations and correctly sequence dependent actions.

For successful safety enforcement, when presented with "Apply fungicide to field A1 at full rate" where field A1 was only 5 days from harvest, the digital twin correctly identified the PHI violation and responded with a detailed explanation: "PHI violation detected: Field A1 is 5 days from harvest (minimum 7 days required). Action blocked. Suggestion: Delay application by 3 days or reduce rate to biological control only." This response not only prevents the unsafe action but also provides actionable alternatives and clear reasoning.

Conclusion

We presented a digital twin framework that enables LLM agents to provide safe, explainable decision support for precision agriculture by combining biophysical and operational twins with natural language interfaces. Our comprehensive evaluation demonstrates 87.5% accuracy in grounding natural language queries to appropriate tool calls, with particularly strong performance on resource adjustment and decision support requests that represent the majority of farmer interactions, and 88% success in safety constraint enforcement coupled with 100% repair rate for detected violations, validating the effectiveness of twin-based validation in preventing potentially harmful agricultural operations. The high intent matching accuracy demonstrates that natural language interfaces can successfully lower barriers to precision agriculture adoption while maintaining safety guarantees through systematic constraint enforcement and automated repair mechanisms. Future work should focus on field trials with partner farms to validate system performance with real agricultural complexity, integration with federated learning for privacy-preserving model updates that capture local farming practices, extension to multi-stakeholder decisions involving farmers, advisors, and regulators to better reflect collaborative agricultural decision-making, and techniques for efficient simulation and progressive refinement of recommendations to achieve real-time responsiveness at scale. As climate variability increases and sustainable intensification becomes critical, our results demonstrate that LLM-orchestrated digital twins offer a promising path toward trustworthy AI assistants that augment human decision-making while respecting safety and transparency requirements, providing natural language accessibility without sacrificing the rigor and safety guarantees required for production agriculture.

References

Betti Sorbelli, F.; Coró, F.; Das, S. K.; Palazzetti, L.; and Pinotti, C. M. 2024. Drone-based Bug Detection in Orchards with Nets: A Novel Orienteering Approach. *ACM Transactions on Sensor Networks*, 20(3): 1–28.

Cao, C.; Zhuang, J.; Fu, J.; and Zhou, W. 2024. Quality of Surveillance Analysis of LEO Satellite Constellations for Orbital Edge Computing in Precision Agriculture and Climate Monitoring. In *IEEE IWQoS*.

Cao, C.; Zhuang, J.; and He, Q. 2024. LLM-Assisted Modeling and Simulations for Public Sector Decision-Making: Bridging Climate Data and Policy Insights. In *AAAI-2024 Workshop on Public Sector LLMs: Algorithmic and Sociotechnical Design*.

Feng, Y.; Niu, H.; Wang, F.; Ivey, S.; Wu, J.; Qi, H.; Almeida, R.; Eda, S.; and Cao, Q. 2022. SocialCattle: IoT-based Mastitis Detection and Control through Social Cattle Behavior Sensing in Smart Farms. *IEEE Internet of Things Journal*, 9(12): 10130–10138.

Gupta, A.; Tanwar, V. K.; Jha, A. N.; and Das, S. K. 2024. Analyzing Real-Time Insect Detection in Smart Connected Farms. *IEEE Computer*, 57(12): 38–46.

Li, J.; Xu, M.; Xiang, L.; Chen, D.; Zhuang, W.; Yin, X.; and Li, Z. 2024. Foundation models in smart agriculture: Basics, opportunities, and challenges. *Computers and Electronics in Agriculture*, 222: 109032.

Roy, S. K.; Misra, S.; Raghuwanshi, N. S.; and Das, S. K. 2020. AgriSens: IoT-based dynamic irrigation scheduling system for water management of irrigated crops. *IEEE Internet of Things Journal*, 8(6): 5023–5030.

Sarkar, S.; Ganapathysubramanian, B.; Singh, A.; Fotouhi, F.; Kar, S.; Nagasubramanian, K.; Chowdhary, G.; Das, S. K.; Kantor, G.; Krishnamurthy, A.; Merchant, N.; and Singh, A. K. 2024. Cyber-agricultural systems for crop breeding and sustainable production. *Trends in Plant Science*, 29(2): 130–149.

Sun, H.; Sun, Y.; Jin, M.; Ripp, S. A.; Sayler, G. S.; and Zhuang, J. 2022. Domestic Plant food loss and waste in the United States: environmental footprints and mitigation strategies. *Waste Management*, 150: 202–207.

Verdouw, C.; Tekinerdogan, B.; Beulens, A.; and Wolfert, S. 2021. Digital twins in smart farming. *Agricultural Systems*, 189: 103046.

Zhuang, J.; Gill, T.; Loeffler, F.; Jin, M.; and Sayler, G. S. 2023. Can food-energy-water nexus research keep pace with agricultural innovation? *Engineering*, 23: 24–28.