Review

Not peer-reviewed version

From Decision-Support Tools to Digital Twins: Digital Farming, Data Platforms, and AI for Sustainable Dairy Systems

Yijing Gong , Eduardo Noronha de Andrade Freitas , Victor E. Cabrera *

Posted Date: 20 March 2026

doi: 10.20944/preprints202603.1613.v1

Keywords: digital farming; precision livestock farming; data platforms; decision-support tools; artificial intelligence; machine learning; optimization; digital twins; sustainability; dairy

Review

From Decision-Support Tools to Digital Twins: Digital Farming, Data Platforms, and AI for Sustainable Dairy Systems

Yijing Gong [1], Eduardo Noronha de Andrade Freitas [2] and Victor E. Cabrera [1,*]

[1] University of Wisconsin-Madison (USA)
[2] Instituto Federal de Goiás (Brazil)
* Correspondence: vcabrera@wisc.edu

Abstract

Livestock sustainability targets increasingly require decision support that is both scientifically credible and operationally usable on farms. This narrative review synthesizes advances in data-enabled decision-support tools and whole-farm "digital twin" modeling for dairy systems, focusing on how heterogeneous streams (herd management records, sensors, milking systems, nutrition, and environment) can be converted into actionable sustainability and profitability decisions. We organize the evidence as a data-to-decision pipeline that links (i) data foundations and interoperability (interoperability, metadata, and continuous quality control), (ii) governance, trust, privacy, and standards (ownership, privacy, and standards that enable scalable sharing and auditing), (iii) analytics and Artificial Intelligence/Machine Learning (AI/ML) models (monitoring/benchmarking, AI/ML for prediction, anomaly detection, phenotyping, and natural-language interaction), and (iv) decision engines (optimization and simulation/digital twins) for prospective and prescriptive "what-if" evaluation of management, nutrition, manure, and welfare strategies under real constraints. Using deployed decision-support ecosystems and peer-reviewed worked examples, we compare model classes by data needs, temporal resolution, interpretability, and deployment maturity, and we identify recurring pitfalls that limit impact, including weak ground truth, dataset shift and drift, fragmented identifiers, and misaligned incentives. We conclude with a practical roadmap and checklist for designing, validating, monitoring, and scaling decision-intelligence systems that produce transparent tradeoffs and measurable sustainability outcomes.

Keywords: digital farming; precision livestock farming; data platforms; decision-support tools; artificial intelligence; machine learning; optimization; digital twins; sustainability; dairy

---

1. Introduction

Sustainable dairy production is increasingly expected to deliver multiple outcomes simultaneously: profitable and resilient farm businesses, reduced greenhouse gas (GHG) emissions and nutrient losses, and credible reporting for supply chains and policy. Yet these outcomes are shaped by interacting biological processes, management constraints, and external variability, so interventions that look attractive in isolation can perform differently once whole-farm feedbacks and real constraints are considered [1,2].

At the same time, dairy farms now generate high-frequency data through herd management software, automated milking and feeding systems, sensors, and cloud services. Precision livestock farming (PLF) and "smart farming" concepts have clarified how continuous data can enable earlier detection of problems, tighter control of processes, and more timely decision-making [3,4]. However, the practical bottleneck is often not data availability, but interoperability, data quality, and the sociotechnical conditions required to make multi-source data reliable and reusable for decision support [5,6].

AI/ML methods can convert raw streams into decision-relevant signals—risk prediction, anomaly detection, phenotyping, and benchmarking—and have matured rapidly in agriculture and animal science [7]. Still, analytics only add value when they are robust to data drift, auditable, and aligned with the timing and controllable levers of farm decisions. Recent dairy-focused examples illustrate both the opportunity and the requirements: unsupervised detection of outliers in daily milk yield streams to strengthen "analytics readiness" [8], and natural-language interfaces to reduce friction when interacting with complex numerical databases—while raising new needs for verification and safe deployment [9].

Sustainability decisions also require prescriptive and prospective capability, quantifying tradeoffs and identifying feasible strategies under constraints. That capability increasingly comes from decision engines that combine optimization, simulation, and emerging "digital twin" approaches. In dairy, platform-scale modeling efforts [10], sustainability-oriented decision tools [11], and optimization frameworks for implementable mitigation strategies [12,13] show how integrated pipelines can move from retrospective assessment toward actionable "what-if" evaluation. Adoption at scale further depends on governance and trust, including data ownership, privacy, and standardization [14–16].

In this review, we synthesize advances across data foundations and interoperability, governance and standards, AI/ML analytics, and decision engines for sustainable dairy systems. Using dairy as a case study and drawing from a coherent set of worked examples from an integrated research and tool-development program, we organize the evidence as a data-to-decision pipeline and conclude with practical guidance for designing and evaluating decision-intelligence systems that are scientifically credible, user-centered, and scalable.

2. Review Approach and Scope

This article is a narrative, concept-driven review organized around a practical data-to-decision pipeline for sustainable dairy systems. We intentionally do not pursue an exhaustive systematic review across the full digital-livestock literature; instead, we use a program-anchored set of worked examples to synthesize design principles that are directly actionable for tool and platform development. This approach aligns with established typologies of review methods, where narrative reviews are appropriate for integrating evidence across heterogeneous methods, data types, and decision contexts when the objective is conceptual synthesis rather than complete enumeration of all studies [17].

2.1. Corpus Assembly and Inclusion Criteria

The 37-source corpus was assembled from peer-reviewed journal articles and review papers (Appendix Table A1). Each source was screened for: (a) a clear decision context (e.g., management recommendations, scenario evaluation, optimization planning, benchmarking, or operational decision support); (b) explicit data inputs and assumptions (e.g., herd records, milking/sensor streams, nutrition/cropping information, economics, or environmental parameters); and (c) evidence of evaluation (e.g., validation, case study demonstration, tool deployment description, or documented performance assessment). We excluded work that was purely descriptive without a decision output, lacked sufficient methodological detail to assess transferability, or focused on component technologies without potential integration into a decision workflow.

2.2. Scope Boundaries

Dairy systems are used as the primary case study because they concentrate many of the challenges facing sustainable livestock production: high-frequency data generation, complex bioeconomic feedbacks, and multi-objective tradeoffs among profitability, emissions, nutrient losses, and animal outcomes. While the worked examples are dairy-focused, the synthesis emphasizes generalizable principles—interoperability, governance, validation, monitoring, and usability—that can transfer to other livestock sectors when adapted to species biology and data ecosystems.

doi:10.20944/preprints202603.1613.v1

2.3. *Evidence Organization and Synthesis Method*

Evidence is synthesized by mapping each source to one or more pipeline layers: data foundations and interoperability; governance, trust, privacy, and standards; analytics (AI/ML); decision engines (optimization and simulation/digital twins); and deployment (validation, monitoring, adoption). Within each layer, we summarize common design patterns, requirements for credibility and reproducibility, and implementation failure modes that limit scaling (e.g., fragmented identifiers, weak ground truth, data drift, misaligned incentives). Although this is not a PRISMA-guided systematic review, we adopt reporting practices consistent with modern transparency expectations—clear scope boundaries, explicit inclusion criteria, and a complete inventory of included sources via Appendix Table A1 [18,19]. We also use a narrative-review quality lens (e.g., explicit rationale, structured synthesis, and balanced limitations) to support readability and rigor [20].

2.4. *Generative AI Disclosure*

Generative AI tools were used to assist with drafting and language editing of portions of this manuscript. All content was reviewed and revised by the authors, who take full responsibility for accuracy, interpretation, and the final text; no generative AI was used to generate original data or to replace verification of cited sources.

3. Evidence Synthesis: The Data-to-Decision Pipeline for Sustainable Dairy Systems

We frame sustainability improvement as decision intelligence: an operational workflow that converts heterogeneous farm data into timely, credible, and actionable choices. In dairy, this means linking high-frequency observations (e.g., herd records, milking-system outputs, and sensor streams) with analytics, bioeconomic reasoning, and practical constraints so recommendations match real management levers and decision cadence [21,22]. Value is realized when multi-source streams are engineered into interoperable, analytics-ready inputs that tools can consume reliably, rather than being repeatedly rebuilt in ad hoc pipelines [23,24].

Accordingly, we organize the evidence into a five-layer data-to-decision pipeline that mirrors how farm-ready systems are built and deployed (Figure 1): (1) data foundations and interoperability (definition, integration, and quality assurance/control, QA/QC) [22,24]; (2) governance, trust, privacy, and standards that enable sharing, auditing, and reuse at scale [14,15]; (3) analytics and AI/ML models that transform raw data into decision-relevant signals while accounting for imperfect ground truth and drift [25,26]; (4) decision engines—optimization and whole-farm simulation/digital-twin concepts—that translate signals and constraints into implementable strategies and explicit tradeoffs [10,12]; and (5) deployment practices (validation, lifecycle monitoring, and user-centered design) that determine adoption and sustainability impact [21,22]. Each layer is necessary; value is realized only when outputs are decision-relevant and maintained over time.

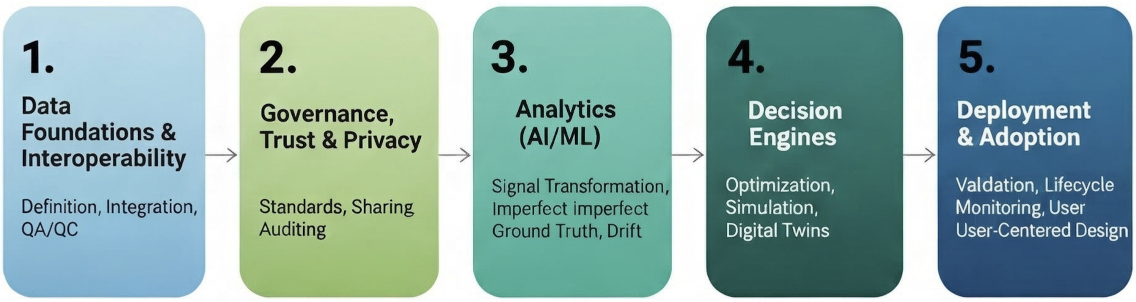

Figure 1. Data-to-decision pipeline architecture for sustainable dairy systems.

 doi:10.20944/preprints202603.1613.v1

Throughout Section 3, we emphasize dependencies across layers—for example, how standardization enables valid benchmarking, how data quality and drift management protect credibility, and how coupling analytics with decision engines improves actionability. This organization also clarifies where implementation commonly fails (e.g., fragmented identifiers, weak QA/QC, unclear governance, poor calibration, or outputs that do not map to controllable levers), motivating the deployment roadmap and checklist presented later in the manuscript [10,22].

3.1. Data Foundations and Interoperability

Digital sustainability decision-making is constrained by the quality, completeness, and interoperability of farm data. Dairy systems generate rich, high-frequency information, yet these data are often fragmented across vendors, captured at different temporal resolutions, and stored in formats that are not directly usable for analytics or whole-farm modeling. The value of "more data" is realized only when ingestion, harmonization, and QA/QC are engineered as a reliable pipeline that produces consistent, analytics-ready inputs for downstream decision tools [22,24]. From a data-platform perspective, interoperability is an ecosystem requirement rather than a single integration step: shared identifiers, minimum metadata, and scalable integration services are prerequisites for robust analytics and decision engines [16,23].

3.1.1. Data Streams, Integration, and Quality Control

Across the pipeline, the most decision-relevant inputs typically combine herd management and performance records (production, reproduction, health events, treatments, culling), milk and milking-system outputs (yield, frequency, flow patterns, conductivity proxies), sensor streams linked to behavior and physiology (activity, rumination, mobility, environment), and nutrition/feed-management information (rations, ingredients, intake or intake proxies, grouping and feeding events). When decisions involve explicit tradeoffs among economics, emissions, and nutrient flows, whole-farm assessment also requires manure-handling information, cropping and feed sourcing, and key external drivers (prices, weather, regulations) [21,22,27]. From a sustainability perspective, these integrated inputs are essential because environmental and economic outcomes are co-determined at whole-farm scale. Early integrated dairy-farm modeling showed that profitability and nitrogen outcomes must be evaluated jointly, and that climate variability can materially change optimal management choices [28,29]. Practical implementations show that integrating these streams enables decision contexts—such as benchmarking and performance comparisons—that are not achievable from any single data system and depend on consistent definitions and comparable data products [30].

Integration failures are frequently driven less by "missing data" than by mismatches across systems: inconsistent cow identifiers and event definitions, timestamp and time-zone errors, unit inconsistencies, and incompatible sampling cadence (e.g., event-based health records versus minute-level sensors versus daily milk yields). Work on dairy data engineering emphasizes that successful integration requires explicit mapping of entities (cow, group/pen, ration, ingredient/lot, device), time-alignment rules, and standardized units before data can be meaningfully analyzed [24]. Standardization efforts further underscore that semantic alignment—agreement on what a data element means (event codes, treatment categories, measurement methods)—is essential for comparability across farms and tools [15,16], and these bottlenecks remain a primary obstacle to scalable analytics and deployment [31].

Because high-frequency farm data are prone to missingness, duplicates, device artifacts, and biologically implausible values, QA and QC must be embedded in ingestion and continuous monitoring rather than treated as ad hoc cleanup. At minimum, this typically includes range and consistency checks, temporal-logic validation (e.g., event ordering), duplicate detection, completeness/latency monitoring, and auditable transformation logs [22,24]. Automated anomaly handling can materially stabilize downstream analyses; for example, unsupervised outlier detection in daily milk yield helps distinguish true biological variation from recording artifacts [8]. Similar principles apply to sensor

 doi:10.20944/preprints202603.1613.v1

streams, where baseline drift, calibration changes, and device replacement can shift distributions even when biology and management are unchanged.

Interoperability also depends on documenting how data were generated. ML-enabled measurement devices can improve capture, but they introduce dependencies on calibration, operating conditions, and versioning, making metadata and validation essential for trustworthy reuse [32]. In practice, minimum metadata should include measurement method, device and software versions, units, sampling frequency, and transformations applied at the source; without provenance, downstream users cannot distinguish real biological/management change from shifts in measurement or preprocessing [15,24].

Finally, scalable decision tools require that integrated pipelines produce standardized, reusable "data products"—for example, curated cow-day performance tables, event histories, ration/group summaries, and sensor-derived features—made accessible via controlled interfaces (e.g., application programming interfaces, APIs) and aligned with governance controls [23,24]. Benchmarking platforms illustrate the point: their value depends on consistent data definitions and repeatable pipelines that update indicators without repeated manual wrangling [30]. These stable data products become the inputs that enable reproducible analytics (Section 3.3) and decision engines (Section 4) for evaluating sustainability strategies across time and farms [16,22].

3.2. Governance, Trust, Privacy, and Standards

Interoperable, quality-controlled data are necessary but not sufficient for scalable decision intelligence. In practice, governance and trust determine whether farms and partners will share data, whether integrated pipelines can be sustained, and whether tool outputs will be accepted and used in routine decision cycles. Dairy data ecosystems are inherently multi-stakeholder—producers, advisors, veterinarians, nutritionists, cooperatives/processors, technology vendors, and researchers—often operating across multiple software systems and cloud services. As data flows expand in volume and sensitivity, uncertainty about ownership, permissible use, and privacy can become a binding constraint on participation and adoption [14,23].

A recurring challenge is that many producers remain unsure who controls farm data and what happens once data are shared with third parties. Governance frameworks that explicitly define ownership, authorized users, consent and withdrawal mechanisms, and auditability can reduce friction and support durable data-sharing relationships [14]. From a pipeline perspective, governance is operational rather than aspirational: it determines which roles can access which data products, under what conditions, and how downstream use—including secondary use and AI model training—is monitored and documented [22,23]. This becomes even more consequential as systems evolve from single-purpose tools toward integrated platforms in which multiple services rely on the same data substrate [16].

Privacy and security concerns directly affect adoption, especially when data are used for benchmarking, sustainability reporting, or third-party analytics, where farms may perceive risk from disclosure of performance, management practices, or business-sensitive information. Effective governance therefore needs to be paired with technical controls—role-based access, secure ingestion, encryption, and auditable logs—to demonstrate that the system behaves as promised [14]. These requirements intensify as AI services are introduced, particularly conversational interfaces or "assistant" layers, because such systems can increase the risk of unintended disclosure or unsupported recommendations if verification and access controls are not designed in from the outset [9,25].

Standards are the practical mechanism that enables interoperability at scale and reduces repeated, costly data wrangling. Standardization efforts emphasize that meaningful data exchange requires more than common file formats: it requires shared definitions (what an event or metric means), minimum metadata (how and when data were generated), and consistent units and identifiers so that multiple systems interpret data consistently [15,24]. Without semantic alignment, integrated pipelines can yield misleading comparisons and unstable models due to hidden differences in recording practices across farms, devices, or software versions. Recent syntheses of data integration in the dairy industry similarly highlight that harmonized definitions and governance processes are prerequisites for reliable analytics and fair, reproducible benchmarking across herds [16,30].

Finally, participation persists when incentives are aligned and the farm-facing value proposition is clear. Benchmarking platforms and decision-support tools illustrate that adoption improves when farms receive actionable outputs (not only reports) and when the burden of data preparation is minimized through automated pipelines and standard interfaces [21,30]. Conversely, misaligned incentives—where data value accrues primarily to third parties—can erode participation even when technical integration is feasible [14]. As AI and integrated knowledge systems mature, incentive alignment also extends to model governance, including transparency about training-data use, opt-in/opt-out mechanisms, and documented benefits returned to participants [25,26].

Taken together, the evidence supports a simple logic: governance should be defined before integration is scaled; privacy and security should be embedded as architecture, not added after deployment; minimum metadata and semantic standards are needed to keep integrated data products comparable and auditable; and sustained participation depends on returning credible, low-friction, decision-relevant value to farms and advisors [14,15,21].

3.3. Analytics and AI/ML Models for Sustainable Dairy Systems

Analytics and AI/ML models convert integrated farm data into decision-relevant signals—predictions, risk scores, phenotypes, benchmarks, and anomaly flags—that can be acted upon within routine management cycles. In sustainable dairy systems, analytics deliver value when they improve the timing and targeting of interventions, communicate uncertainty and limitations transparently, and generate outputs that map to controllable levers and operational constraints rather than emphasizing "model performance" in isolation [21,22]. Accordingly, analytics should be treated as sustainability-enabling signals, not endpoints. Their practical value emerges when predictions are translated into constrained decisions that explicitly quantify environmental gains and economic consequences [12,13,33]. This section synthesizes AI/ML applications most relevant to sustainability decision-making and highlights the validation, monitoring, and reporting practices required for credible deployment (Table 1).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2026 doi:10.20944/preprints202603.1613.v1

Table 1. Comparison of model and decision-support classes (AI/ML, simulation/digital twins, and optimization): data needs, temporal resolution, interpretability, and deployment maturity. Worked examples and citations are drawn from the curated 37-source program corpus (Section 2).

Dimension	AI/ML analytics	Simulation and digital twins	Optimization (prescriptive)
Primary output	Predictions, risk scores, anomaly flags, phenotypes, benchmarks; supports "who/when to act" decisions [8,9,31].	Prospective "what-if" outcomes over time: economics, emissions, nutrients, herd dynamics; supports strategy evaluation and policy/scenario analysis [10,27,30,33].	Implementable plans under constraints: group/ration/crop-diet allocations, mitigation strategies; supports "what to do given limits" decisions [10,12,13,34].
Typical data	High-frequency, labeled or proxy-labeled data; consistent definitions and ground truth for target outcomes.	Structured farm system inputs (herd, nutrition, manure/crops, prices, parameters) with documented assumptions and uncertainty ranges.	Constraint sets plus resource inventories (land, feed, grouping rules) plus objective weights; requires interpretable inputs and decision levels.
Temporal resolution	Real-time to daily/weekly updates; performance sensitive to data latency and drift.	Weekly–annual (or longer) horizons; can ingest daily summaries but outputs often strategic/seasonal.	Decision-interval dependent (daily to seasonal); typically run on scenario basis rather than continuously.
Interpretability & auditability	Varies by model class; requires calibration, explainability for decisions, and transparent uncertainty.	High transparency when assumptions and modules are documented; traceable inputs/outputs enable audit.	Typically interpretable because objectives/constraints are explicit; sensitivity analysis is critical.
Validation by intended use	Temporal holdouts, external-herd tests where feasible; calibration of risk scores and subgroup robustness.	Compare outputs to observed ranges; plausibility checks; scenario validation with expert review.	Feasibility checks; sensitivity analysis; reproducible runs; validate that solutions map to real controllable actions.
Common failure modes	Biased training data; weak/shifted ground truth; unhandled missingness; model drift; outputs not actionable.	Uncertain or undocumented parameters; poor calibration; unrealistic assumptions; inadequate uncertainty propagation.	Over-simplified constraints; objective mis-specification; solutions infeasible in practice; missing stakeholder preferences.
Best-fit sustainability questions	Early warning (health, performance), anomaly detection, benchmarking readiness, phenotyping for downstream models.	Tradeoffs among profitability–GHG–nutrients; long-run system impacts of management and technology.	Mitigation and resource-allocation planning under constraints; prescriptive "next best action" policies.
Deployment maturity	Often deployed as decision-support services when monitoring + governance are in place; requires lifecycle management.	Often used for research/extension and strategic planning; digital-twin capability requires continuous data integration.	Deployed when constraints and objectives reflect farm reality and solutions integrate into workflows.

3.3.1. Predictive Modeling for Health and Performance Decisions

Predictive models matter most when they change decisions—for example, prioritizing animals for observation, triggering preventive actions, or tailoring management before losses occur. Early prediction of clinical mastitis from routinely collected farm data illustrates how supervised machine learning can shift management from reactive treatment to proactive allocation of attention and resources [31]. Similar approaches for milk-yield dynamics highlight a key deployment challenge: models must remain robust across herds, parities, and environments, not only within a single dataset. Frameworks that explicitly address heterogeneity (including genotype-by-environment effects and within-herd relatedness) reinforce the need for designs that generalize under realistic conditions [35]. From a sustainability perspective, the purpose of prediction is to reduce avoidable losses (health events, premature culling) and improve resource-use efficiency (feed, labor, treatments) while maintaining animal outcomes.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2026 doi:10.20944/preprints202603.1613.v1

3.3.2. Data Quality, Anomaly Detection, and Analytics Readiness

High-frequency dairy data are frequently affected by missingness, sensor artifacts, and implausible values, and these issues can bias models and undermine trust if not managed upstream. Unsupervised machine learning approaches for detecting outliers in daily milk yield provide a concrete example of embedding anomaly detection into the pipeline to separate true biological variation from data artifacts and to stabilize downstream analyses [8]. This type of quality-focused analytics is especially valuable because it strengthens the entire pipeline: improved data reliability benefits forecasting, phenotyping, benchmarking, and system modeling. In practice, anomaly detection should be integrated with ingestion QA/QC and paired with review/correction workflows so that automated flags become actionable rather than generating alert fatigue [22,24].

3.3.3. Phenotyping and Lactation-Curve Analytics as Sustainability-Relevant Signals

Phenotyping—deriving biologically meaningful traits from observational data—creates new opportunities to measure and manage sustainability-relevant outcomes at scale. Lactation-curve parameters are a prime example: large-scale analyses of temporal, geographic, and management drivers of lactation-curve parameter variation demonstrate how phenotypes extracted from routine records can capture systematic differences linked to management and environment [36]. Such phenotypes become inputs to higher-level decision engines by improving representation of production dynamics and enabling more realistic scenario evaluation. Recent work also shows that tailoring lactation-curve estimation to the target context can have measurable systemic effects when used in farm simulation pipelines, reinforcing that "better phenotypes" can improve whole-farm decision fidelity rather than merely improving descriptive fit [37].

3.3.4. Trend Analytics, Monitoring, and Benchmarking

Beyond individual-animal prediction, analytics can support sustainability through system monitoring, benchmarking, and detection of structural changes over time. Time-series analysis of milk productivity at larger scales illustrates how temporal methods can identify trends and structural shifts that inform both management strategies and policy/industry interpretation [38]. At the farm and advisor level, benchmarking platforms emphasize operational usability: analytics must produce comparable metrics, explain differences transparently, and update reliably as new data arrive [30]. Interpretation of benchmarking results should also remain explicit about functional unit and system boundary, because environmental conclusions can shift when impacts are expressed on different output bases. This boundary discipline is critical for credible sustainability interpretation across farms and products [39]. These applications depend heavily on standardization and governance—without consistent definitions and auditable data pipelines, benchmarking can lead to misleading comparisons and loss of confidence [14,15].

3.3.5. Behavioral Analytics and Emerging AI Interfaces

Some analytics applications focus on behavior and interaction patterns rather than physiology alone. For example, social-network analysis applied to automated milking systems demonstrates how connectivity and behavioral structure can be quantified from routine system logs, opening additional pathways for management interpretation and targeted interventions [31]. In parallel, AI interfaces are evolving from models that output scores to systems that improve access to data and decision logic. Narrative reviews of computer vision systems and large language models in livestock production highlight both opportunity and risk: while these approaches can expand measurement and reduce friction, they also heighten requirements for validation, bias assessment, and safe deployment [25]. Natural-language interaction with numerical databases (e.g., Dairy GPT) exemplifies this direction by enabling farmers to query structured datasets conversationally; however, such systems must be designed to return verifiable results, preserve privacy, and maintain auditability—particularly when outputs influence management actions [9]. Conceptually, this aligns with emerging views that animal-agriculture AI is moving toward integrated knowledge systems that combine data access, analytics, and decision workflows [26].

3.3.6. Cross-Cutting Requirements: Validation, Drift, and Adoption

Across these use cases, several recurring issues limit scalability: imperfect ground truth labels (especially for health/welfare outcomes), dataset shift across herds and seasons, and model drift as technology, genetics, and practices evolve. Therefore, analytics intended for decision support should be evaluated using validation designs that reflect deployment reality (external herds when feasible, temporal holdouts, sensitivity to missingness/noise), with calibration and thresholds tied to action capacity [21,22]. Ongoing monitoring is essential: data drift (sensor baselines, recording practices) and concept drift (changing relationships between predictors and outcomes) can degrade field performance and erode trust if not detected and managed. Finally, analytics create sustainability value only when delivered through user-centered outputs—prioritized lists, interpretable drivers, uncertainty cues, and workflow integration—so that farmers and advisors can act with confidence [21,30].

In summary, AI/ML analytics in dairy systems are most impactful when they (i) strengthen data readiness and phenotyping, (ii) support targeted, timely decisions, and (iii) connect to prescriptive decision engines that quantify tradeoffs under constraints [10,12]. This linkage is the bridge from "insight" to implementable sustainability strategies, motivating the decision-engine synthesis that follows.

4. Decision Engines: Optimization, Simulation, and Digital Twins

Analytics become sustainability action only when they are coupled to decision engines that can (i) translate signals into implementable choices under constraints and (ii) quantify tradeoffs among outcomes (profitability, emissions, nutrients, and animal performance). In dairy systems, the most mature decision engines fall into two complementary classes: prescriptive optimization (what should we do, given objectives and constraints?) and prospective simulation (what will likely happen, given biology, management rules, and uncertainty?). Emerging "digital twin" concepts extend simulation by emphasizing continuous updates from farm data streams and repeated re-evaluation as conditions change [10,22,40].

4.1. Prescriptive Optimization for Implementable Strategies

Optimization is particularly useful when farms must allocate limited resources (feed, land, labor, replacements) across competing objectives. A central advantage is transparency: objectives and constraints can be stated explicitly, allowing users to see why a recommended strategy is feasible and which tradeoffs are being made.

4.1.1. Whole-Farm and Herd-Level Optimization

Early work using large-scale linear programming frameworks demonstrated how optimization can be used to jointly improve net income while accounting for diet decisions and environmental outputs such as nitrogen excretion [41]. More recent approaches extend this idea to multi-objective farm performance by jointly optimizing productivity, herd structure, environmental performance, and profitability—explicitly framing sustainability as a constrained decision problem rather than a reporting exercise [33]. In practice, optimization outputs should report not only the recommended strategy but also the marginal tradeoff between environmental improvement and profitability. Presenting these tradeoff curves supports implementable choice among feasible sustainability pathways rather than a single opaque optimum [12,13,33].

Mixed-integer linear programming (MILP) is well suited to dairy decision problems where discrete choices matter (crop rotations, ingredient inclusion/exclusion, manure or mitigation options). A clear example is MILP-based crop–diet planning to mitigate enteric methane while maintaining implementability under farm constraints [12]. This kind of formulation is particularly relevant for

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2026 doi:10.20944/preprints202603.1613.v1

sustainability because it can expose efficient frontiers (e.g., how much methane reduction is achievable at what cost) while respecting feed inventories, cropping capacity, and ration structure.

4.1.2. Ration Grouping as an Operational Decision Engine

Nutritional grouping is a practical lever that farms can implement quickly, but it has complex economic implications (feed costs, sorting, within-group variability). Economic evaluations show that grouping decisions can materially affect profitability [42], and empirical/simulation-based work demonstrates that multiple ration-grouping strategies can improve nutritional accuracy and economics when designed to match herd structure and management realities [34]. A synthesis of grouping considerations further underscores that "optimal" grouping depends on objectives, constraints, and the data available to define groups reliably [43]. These studies illustrate a broader principle: optimization is most valuable when the decision variables correspond to levers farms can change and when constraints reflect on-farm feasibility.

Recent work suggests that nutritional grouping can be evaluated as a joint economic and environmental decision, rather than solely as a feeding-management choice [13]. Specifically, this study examined whether grouping could reduce feed costs and enteric methane when diets were optimized using existing farm ingredients with only modest adjustments. The findings further support framing nutritional grouping as an operational decision engine for achieving multiple objectives under practical farm constraints.

4.2. Prospective Simulation for Tradeoff-Aware "What-If" Evaluation

Simulation engines are essential when system dynamics, biological feedbacks, and uncertainty make closed-form decision rules unrealistic. In sustainable dairy systems, simulation provides a credible way to evaluate scenarios (nutrition, manure handling, reproductive strategy, replacement policy) and quantify downstream impacts on both economic and environmental metrics.

4.2.1. Whole-Farm Bioeconomic Simulation as a Foundation

Integrated dairy farm models have long demonstrated the value of combining biology and economics to evaluate strategies that influence profitability and nutrient outcomes [28]. Extensions that incorporate seasonal climate variability highlight why prospective modeling matters for sustainability: expected outcomes can change meaningfully as external conditions vary, affecting both resource use and environmental performance [28]. Similarly, simulation-based evaluations of greenhouse gas mitigation strategies show that cost-effectiveness depends on interactions among feeding strategies, manure management, and farm-specific baselines [27]. This pattern is consistent with earlier integrated simulation evidence showing that management impacts on environment and profit are strongly context-dependent and climate-sensitive. Therefore, robust sustainability recommendations require prospective whole-farm evaluation rather than isolated component analysis [27–29].

4.2.2. Stochastic Life-Cycle Simulation for Reproductive and Herd-Structure Decisions

Several sustainability-relevant decisions operate through herd dynamics and longtime horizons (replacement, reproduction programs, survival). Stochastic animal life-cycle simulation provides a principled way to evaluate such strategies under uncertainty and to quantify expected economic returns [36]. This is important because sustainability outcomes (emissions intensity, nutrient flows, resource use) are often mediated by herd structure and longevity, which cannot be inferred reliably from short-term data alone.

Improving simulation fidelity with better biological submodels is also essential. Simulation credibility depends on whether key biological processes are represented with sufficient fidelity. Work on tailored lactation-curve estimation illustrates that improving how production dynamics are parameterized can have systemic impacts when embedded in farm simulation workflows [37]. At a larger scale, platform-level efforts emphasize modular biophysical representations of animals and management. For example, the RuFaS animal module provides a detailed description of animal

 doi:10.20944/preprints202603.1613.v1

management processes that enables more consistent representation of production and management across scenarios [44], supporting whole-farm evaluation and linking management interventions to outcomes [10].

4.2.3. Operational Sustainability Tools as Simulation "Products"

Decision-support models designed for stakeholder use demonstrate how simulation outputs can be packaged into tools that are operationally interpretable. For example, the DairyPrint model illustrates how integrated modeling can support sustainability assessment for farms and stakeholders by translating scenario outcomes into actionable metrics and comparisons [11]. At the same time, platform-level development indicates that these tool-oriented outputs can be embedded in broader digital-twin ecosystems, enabling repeated scenario testing and continuous sustainability improvement cycles at farm and network levels [10,11]. This highlights a recurring design requirement: simulation outputs must be presented in ways that align with decisions (what changes, by how much, and under what assumptions), rather than only providing technical model outputs.

4.3. Toward Digital Twins and Hybrid Decision Intelligence

Digital twin concepts build on simulation but emphasize continuous updating from live data streams, repeated recalibration, and ongoing scenario evaluation as farm conditions evolve [22,40]. In practice, the most actionable pathway is often hybrid decision intelligence: analytics provide timely signals (risk scores, phenotypes, anomaly flags), while optimization and/or simulation translate those signals into implementable strategies and quantified tradeoffs. Platform approaches such as RuFaS further motivate this direction by explicitly positioning the system as a foundation for future research and action in sustainable dairy farming, where consistent data pipelines and modular components enable repeated, comparable "what-if" evaluation [10,44]. Across the evidence, three principles emerge for decision-engine design in sustainability contexts: (i) make objectives, constraints, and assumptions explicit so tradeoffs are transparent; (ii) prioritize outputs that map to controllable levers and operational cadence; and (iii) report uncertainty and sensitivity because sustainability decisions often operate under biological, climate, and market variability [21,27,28]. These principles set the stage for Section 5, where we focus on deployment requirements (validation, monitoring, and user-centered design) that determine whether decision engines achieve real-world sustainability impact.

5. Deployment: Validation, Monitoring, Adoption, and User-Centered Design

Even strong analytics and decision engines can fail to generate sustainability impact if they are not validated in realistic conditions, monitored for degradation, and delivered through user-centered workflows that fit farm decision cycles. Deployment is therefore not a "final step," but a continuous lifecycle that links model credibility, governance, data operations, and adoption. Evidence across dairy decision-support and digital-farming systems consistently shows that farmers and advisors adopt tools when they are trustworthy, actionable, and low-friction—and disengage when outputs are unstable, opaque, or misaligned with decision cadence [21,22,29].

5.1. Validation Aligned to Intended Use

Validation should match the decision context, not just maximize generic predictive accuracy. For example, a health-risk model is only useful if it improves decisions under real constraints (labor capacity, treatment protocols, pen logistics) and if the cost of false positives/negatives is understood. For supervised learning applications such as early mastitis prediction, validation should include temporal holdouts (forward-in-time testing), external herd evaluation when feasible, and calibration checks so predicted risks correspond to observed event rates [22,31]. For unsupervised methods used in QA/QC (e.g., milk-yield outlier detection), validation should emphasize stability across seasons and device conditions and confirm that flagged anomalies reflect data artifacts or biologically meaningful exceptions rather than routine variation [8].

 doi:10.20944/preprints202603.1613.v1

For optimization and whole-farm simulation, validation focuses on different but equally practical criteria: (i) input fidelity (do constraints, inventories, and herd structure reflect farm reality?), (ii) output realism (are predicted trajectories within plausible ranges?), and (iii) scenario credibility (do "what-if" comparisons preserve internal consistency?). Demonstrations of mitigation planning under practical farm constraints and platform-based whole-farm modeling underscore that credibility depends on transparent assumptions, sensitivity analysis, and alignment of outputs with implementable actions [10,11,13]. For sustainability-focused engines, validation should also include directional consistency checks across environmental accounting perspectives, ensuring that predicted mitigation effects remain coherent under realistic whole-farm assumptions and boundary definitions [12,27,39].

5.2. Reporting, Transparency, and Reproducibility

To sustain trust and support adoption, deployment requires a minimum reporting package that enables users and reviewers to understand what the system does, where it applies, and why outputs change. At a minimum, field-ready systems should document:

- Data provenance and definitions: sources, time span, sampling frequency, event/label definitions, and inclusion/exclusion rules [15,24].
- Preprocessing and feature construction: unit conversions, aggregation windows, missing-data handling, and QA/QC rules [8,22].
- Model or engine specification: algorithms/parameters (for AI/ML), objective functions/constraints (for optimization), module assumptions (for simulation), and software versions [12,13,21].
- Performance and uncertainty: metrics appropriate to the use case, calibration where relevant, and evidence of robustness across time/farms [22,31].
- Versioning and change logs: what changed, when, and the expected effect on outputs—critical when tools are used for benchmarking or sustainability reporting [23,30].

These elements are also essential for governance and auditing, especially as platforms integrate multiple services and stakeholders [14,16].

5.3. Monitoring for Data Drift and Model Drift

In continuous data environments, degradation is expected unless drift is actively managed. Drift may arise from (i) data drift (sensor replacement, recording practice changes, seasonal shifts), (ii) concept drift (changing relationships between predictors and outcomes), or (iii) label drift (health event coding changes). Effective deployment therefore requires automated monitoring at two levels:

- Data monitoring: completeness, latency, distribution shifts, device baselines, and key plausibility ratios [22,24].
- Model monitoring: tracking calibration and performance on recent data, comparing "champion–challenger" models, and defining triggers for recalibration or retraining [21,22].

Unsupervised anomaly detection (e.g., in daily milk yield) can serve both as a QA/QC layer and as an early warning system for sensor or pipeline changes that would otherwise silently degrade downstream models [8]. For natural-language interfaces to databases, monitoring must also include guardrails: ensuring outputs are verifiable against underlying records and that access controls prevent unintended disclosure [9,14].

5.4. Adoption by Farmers and Advisors: User-Centered Design Features

Adoption depends on whether the system fits real workflows for farmers and advisors. Evidence from technology uptake in agriculture emphasizes that perceived usefulness, trust, and ease-of-use are decisive—especially when tools add data-entry burden or produce unstable recommendations [21,29]. Across dairy decision-support and digital platform work, several design features consistently support adoption:

 doi:10.20944/preprints202603.1613.v1

1. Actionability: outputs should map directly to decisions (e.g., prioritized cow lists, group/ration recommendations, scenario comparisons) and specify what action is feasible now [13,21,30]. Adoption is further strengthened when recommendations are presented as integrated management bundles (e.g., feed, manure, and crop adjustments) with side-by-side reporting of expected environmental and economic effects. This framing supports practical implementation under real farm constraints [12,27,33].
2. Workflow integration: reduce manual entry; rely on interoperable pipelines and standardized data products so outputs update automatically [23,24].
3. Interpretability and "why": provide key drivers of predictions and binding constraints in optimization; avoid black-box outputs that cannot be interrogated [12,21].
4. Uncertainty communication: show confidence/uncertainty and provide safe defaults when data are insufficient; avoid false precision [22].
5. Trust and governance by design: transparent data-use statements, role-based access, and auditable logs; essential for sustained participation and benchmarking credibility [14,15].

In summary, deployment success is achieved when validated analytics and decision engines are embedded within governed data pipelines, continuously monitored for drift, and delivered through user-centered designs that support farm and advisor decision cycles [10,21,22]. These practices determine whether digital decision intelligence delivers durable sustainability gains beyond retrospective reporting and toward sustained improvement in management outcomes.

6. Roadmap and Implementation Checklist

The evidence synthesized in Sections 3-5 supports a consistent conclusion: sustainability impact from digital farming depends on an end-to-end, governed pipeline—data foundations $\rightarrow$ analytics $\rightarrow$ decision engines $\rightarrow$ deployment—rather than isolated models or one-off reports (Figure 2a). A useful maturity criterion is whether the system can repeatedly connect boundary-aware environmental assessment, constrained optimization, and stakeholder-facing scenario interpretation in one workflow. This integration is central to turning innovation into sustained, auditable sustainability gains [11,12,39]. Accordingly, scalable implementation should prioritize readiness and credibility before expansion, while maintaining lifecycle practices (monitoring, versioning, and governance) that preserve trust over time [21,22,24]. This section provides a staged roadmap (Figure 2b) and a checklist that can be used by farms, advisors, technology providers, and researchers to plan, evaluate, and maintain decision-intelligence systems for sustainable dairy production [10,14,15].

 doi:10.20944/preprints202603.1613.v1

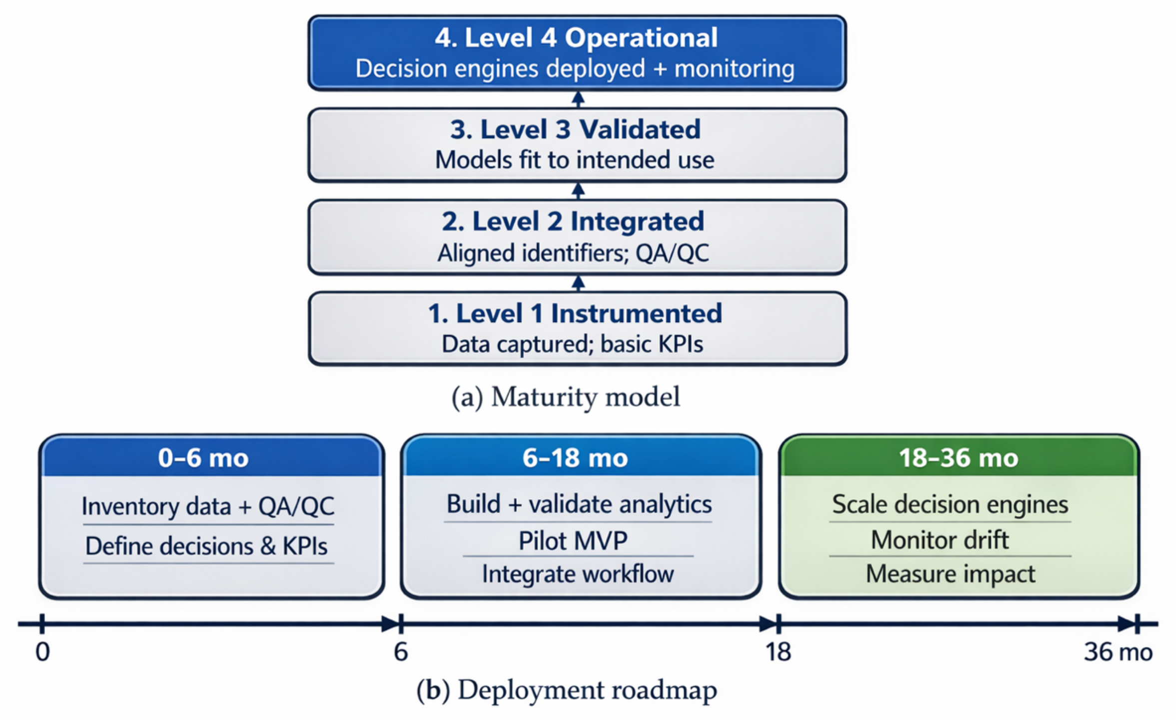

Figure 2. Maturity model (a) and deployment roadmap (b): from data readiness to validated decision engines at farm and sector scale.

6.1. Implementation Roadmap (Staged Milestones)

6.1.1. Stage 1 (0–6 Months): Establish Decision Targets and Data Readiness

Define the decision use cases and success metrics (economic + environmental + animal outcomes), specify decision cadence and controllable levers, and inventory required data streams. Implement foundational ingestion, harmonization, and QA/QC with documented provenance (identifiers, units, timestamps, metadata). In parallel, establish governance (ownership, permissible use, roles/authorization, consent/withdrawal, auditability) and produce a baseline snapshot of current KPIs and data completeness [14,15,24]. Early value can be created by delivering stable, comparable "data products" (e.g., cow-day tables, event histories, ration/group summaries) that reduce repeated manual wrangling and enable reliable benchmarking [30].

6.1.2. Stage 2 (6–18 Months): Validate Analytics and Integrate into Workflows

Develop or adopt analytics aligned to the decision context and validate using temporal holdouts and external herds where feasible, including calibration for risk scores. Define decision thresholds tied to action capacity, integrate outputs into farm/advisor workflows (reports, prioritized lists, scenario summaries), and implement monitoring for data/model drift with explicit response protocols [8,21,22]. Pilot deployments should evaluate decision impact (behavior change, avoided losses, time saved, improved efficiency), not only predictive performance. Where relevant, begin coupling analytics outputs to scenario evaluation so users can test "what-if" alternatives rather than receiving isolated scores [11].

6.1.3. Stage 3 (18–36 Months): Scale Decision Engines and Digital-Twin Capability

Couple analytics with decision engines (optimization and/or whole-farm simulation) to support tradeoff-aware decisions under realistic constraints and provide implementable strategies [10,12,13]. At this stage, modular platform architecture becomes critical so animal, feed/crop, and manure domains can be evaluated consistently under shared assumptions across scenarios and farms. This is a key requirement for scalable, system-level sustainability intelligence [10]. Expand interoperability using shared definitions and secure APIs, operationalize reproducibility (preprocessing and model version

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2026 doi:10.20944/preprints202603.1613.v1

control, change logs, rollback), and scale to multi-farm cohorts with privacy-preserving benchmarking and consistent metric definitions [14,15,22]. As systems mature, adopt digital-twin practices where appropriate: continuous updating, periodic recalibration, and repeated scenario evaluation as farm conditions evolve [10].

6.2. Implementation Checklist

To make the recommendations operational for sustainable dairy farm systems, we summarize them in Table 2 as an implementation checklist organized by pipeline layer (data foundations, governance, analytics/validation, monitoring, and user experience). Each checklist item is paired with worked dairy examples from the reviewed corpus to illustrate what "good" looks like in practice—supporting traceability from evidence to implementation—and to help teams build decision-intelligence systems that are both environmentally credible and operationally usable on farms.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 March 2026 doi:10.20944/preprints202603.1613.v1

Table 2. Implementation checklist for scalable decision-intelligence systems (data, governance, validation, monitoring, UX, reporting) with worked dairy examples.

Domain (pipeline element)	Minimum requirement	Practical checks / evidence to document	Worked examples
Data readiness	Clear decision use-case(s), KPIs, cadence, and controllable levers	Decision statement; KPI definitions (economic + environmental + animal); decision timing (daily/weekly/seasonal); action capacity	Decision-support framing and tool deployment focus [21]; "real-time continuous decision-making" vision [22]
Ingestion & integration	Multi-source ingestion with time alignment and unit harmonization	Source inventory; mapping of identifiers (cow, group, ration, field, device); unit/time zone standards; ETL (extract transform load/ELT (extract load transform) logs	Data engineering pipelines for dairy streams [24]; Dairy Brain integration perspective [23]
Quality control (QA/QC)	Continuous automated QA/QC at ingestion + feedback loop for corrections	Completeness/latency dashboards; range/consistency checks; duplicate detection; missingness policy; audit trail of corrections	Outlier/anomaly handling in daily milk yield streams [8]; "big-data" continuous decision context [22]
Governance & trust	Ownership, permissible use, roles, consent/withdrawal, auditability	Data-use agreement; role-based access control; consent management; access logs; secondary-use policy	Data integration pathways and governance needs [14–16,23]
Privacy & security	Security-by-design for data and services	Encryption in transit/at rest; least-privilege access; secure API gateway; incident plan; vendor risk review	Farmer-facing trust barriers and adoption considerations in digital systems [23]; LLM (large language models)/NLP (natural language processing models) interfaces motivate stronger controls [9]
Analytics validity (AI/ML)	Validation aligned to the decision context (not only "accuracy")	Temporal holdout tests; external herd validation when feasible; calibration for risk scores; sensitivity to missingness/noise; subgroup checks	Early prediction of clinical mastitis [31]; tailored lactation-curve estimation for simulation impacts [37]
Optimization credibility	Transparent objectives + constraints reflecting farm reality	Constraint list (land, feed inventory, ration structure, discrete options); feasibility checks; sensitivity analysis; reproducible runs	MILP-based joint crop rotation and diet planning under environmental regulations [12]; dual-objective optimization of nutritional grouping under ingredient constraints [13]
Simulation / digital-twin credibility	Inputs/assumptions traceable; outputs plausible; uncertainty explored	Module input tables; parameter sources; plausibility bounds; scenario definitions; stochastic/sensitivity analyses; comparison to observed ranges	Integrated farm system modeling for economics–GHG tradeoffs [16,27,33]; RuFaS platform concept and scaling [10]
Monitoring for drift	Continuous monitoring for data drift + model drift with response protocol	Data drift: distribution shifts, sensor baselines, latency; model drift: calibration tracking; retraining triggers; rollback/versioning	Outlier/drift-relevant patterns in high-frequency milk data [8]; "continuous decision-making" implies lifecycle monitoring [22]
Reporting & reproducibility	Sufficient documentation to reproduce and audit outputs	Data definitions; preprocessing; model specs; validation design; uncertainty reporting; version/change logs; model registry	Data-platform and analytics documentation needs emphasized in Dairy Brain perspective [23]; tool-oriented reporting in DairyPrint [11]
UX & adoption (farm + advisor)	Actionable outputs integrated into workflows and decision cadence	"Next best action" design; thresholds linked to action capacity; shareable advisor reports; training/support; ROI tracking	Benchmarking platform approach [30]; decision-support tool design principles [21]; DairyPrint tool deployment [11]
Safe deployment of LLM/NLP interfaces	Verifiable outputs, guardrails, and auditability	Retrieval/citation of source records; no "unsupported" recommendations; PII handling; logging of prompts/outputs; fallback behaviors	Natural-language access to numerical databases [9]; AI integration perspective [25,45]

 doi:10.20944/preprints202603.1613.v1

Together, the staged roadmap and checklist provide a practical basis for building, evaluating, and sustaining decision-intelligence systems that are scientifically credible, operationally usable, and scalable—conditions required for durable sustainability impact in dairy production.

7. Conclusions and Outlook

Digital farming has made dairy one of the most data-rich livestock sectors, yet sustainability impact depends on whether those data are converted into trusted, implementable decisions that fit farm management cycles. Four takeaways stand out. First, data foundations and interoperability remain the binding constraint for scalable applications; without stable identifiers, harmonized definitions, and continuous quality control, sustainability metrics are difficult to validate, compare, and operationalize across farms and over time. Second, governance and trust are enabling infrastructure, not optional additions; clear rules for ownership, privacy, permissible use, and auditability are prerequisites for durable data flows and adoption as platforms expand across stakeholders. Third, AI/ML contributes most when it strengthens decision readiness, improving data reliability, phenotyping, and timely targeting of interventions, while being deployed with validation designs that reflect real operating conditions and include ongoing monitoring for drift. Fourth, sustainability decisions require more than prediction; the most actionable pathway is hybrid decision intelligence, where analytics are coupled with optimization and whole-farm simulation to generate implementable strategies and transparent tradeoffs among economic, environmental, and animal outcomes under real constraints.

Looking ahead, progress is likely to be driven less by marginal algorithmic gains and more by the engineering, governance, and deployment disciplines that make decision-intelligence systems reliable at scale. Priorities include interoperable, machine-readable standards and minimum metadata to enable reproducible metrics and comparable benchmarking; privacy-preserving architectures and governance processes that preserve farmer agency while enabling learning across farms; routine lifecycle management (drift detection, revalidation, versioning, and transparent change logs) so systems remain credible as technologies and management evolve; and evaluation designs that measure decision impact and sustainability outcomes, not only predictive accuracy. Methodologically, high-impact opportunities include tighter coupling of phenotyping with system modeling, expanded use of constrained optimization for mitigation planning, and more explicit treatment of uncertainty to support resilient decisions under climate and market variability.

Author Contributions: Conceptualization, V.E.C, E.N. de A.F., and Y.G.; methodology, V.E.C; validation, V.E.C, E.N. de A.F., and Y.G.; formal analysis, V.E.C; investigation, V.E.C, E.N. de A.F., and Y.G.; resources, V.E.C; data curation, V.E.C; writing—original draft preparation, V.E.C; writing—review and editing, E.N. de A.F. and Y.G.; visualization, V.E.C; supervision, V.E.C; project administration, V.E.C; funding acquisition, V.E.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

 doi:10.20944/preprints202603.1613.v1

AI	Artificial intelligence
API	Application programming interface
DSS	Decision support system
ELT	Extract, load, transform
ETL	Extract, transform, load
GHG	Greenhouse gas
KPI	Key performance indicator
LLM	Large language model
MILP	Mixed-integer linear programming
ML	Machine learning
NLP	Natural language processing
PII	Personally identifiable information
PLF	Precision livestock farming
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QA/QC	Quality assurance / quality control
ROI	Return on investment
RuFaS	Ruminant Farm Systems

Appendix A

Table A1. Core corpus mapping to the data-to-decision pipeline including pipeline layers, data inputs, methods, decisions, and validation

Paper	Pipeline layer(s)	Data inputs	Method type	Decision output	Validation / deployment evidence
[39]	Decision engines (assessment)	Wisconsin dairy-farm production + management data; feed / manure / energy use; emission factors; cheese yield assumptions	Partial life cycle assessment (GHG accounting) with specialty cheese functional unit	GHG emissions intensity per functional unit; hotspot identification; comparative scenarios	Case-study parameterization; internal consistency checks and sensitivity/assumption analysis (no on-farm deployment reported)
[15]	Data foundations & interoperability; Governance / trust / standards	Stakeholder and industry context; existing data standards and data-collection practices	Standards synthesis / framework paper	Recommended common data elements, definitions, and collection guidance; path-forward roadmap	Expert / consensus synthesis; intended for industry adoption (no empirical validation required)
[34]	Decision engines (management strategy); Analytics (supporting metrics)	Cow/group nutrient requirements; ration formulations; feed costs; herd production / management constraints	Scenario-based nutritional grouping and economic evaluation	Improved nutritional accuracy and economics via multiple ration-grouping strategy; guidance on grouping design	Model-based comparisons across grouping strategies using realistic farm inputs; outputs reported as economic/nutritional gains
[43]	Decision engines (management guidance); Deployment/adoption	Published literature on grouping strategies and implementation considerations	Narrative literature review	Practical considerations and evidence synthesis for nutritional grouping decisions	Literature-based synthesis; supports implementation decisions (no new empirical validation)
[14]	Governance / trust / privacy / standards; Deployment/adoption	U.S. and global perspectives on farm data ownership/privacy; stakeholder experiences and policy context	Perspective / narrative synthesis of governance issues	Principles and options for ownership, consent, access control, and privacy-preserving sharing	Evidence grounded in stakeholder and policy context; provides actionable governance guidance (not an algorithmic validation study)

Table A1. *Cont.*

Paper	Pipeline layer(s)	Data inputs	Method type	Decision output	Validation / deployment evidence
[29]	Deployment (adoption & decision behavior); Governance/trust (perceptions)	Farmer survey data; attitudes and perceived usefulness of seasonal climate forecasts	Empirical survey / adoption-behavior analysis	Factors associated with willingness to use climate forecast technology; adoption barriers / enablers	Field survey evidence with statistical analysis; directly informs deployment strategy
[41]	Decision engines (optimization); Analytics (derived indicators)	Herd state structure; culling/replacement parameters; diet options; prices; N excretion relationships	Large Markovian linear programming optimization	Optimal replacement policy maximizing net income; associated N excretion outcomes	Computational verification and scenario evaluation; outputs reported as optimized policies and tradeoffs
[21]	Roadmap / checklist (synthesis); Decision engines (assessment framing); Deployment	Cross-disciplinary sustainability and GHG literature	Review / perspective synthesis	State-of-the-field framing for dairy sustainability; priority needs and research directions	Conceptual synthesis; supports positioning and roadmap development (no primary validation)
[22]	Pipeline overview (Data→Decision); Data foundations; Analytics; Decision engines; Deployment	Precision dairy technology streams and farm records; examples of big-data workflows	Symposium review / framework	Real-time continuous decision-making workflow; requirements across data, analytics, and action layers	Framework supported by worked examples; guidance for scalable implementation
[45]	Analytics (AI/ML); Deployment (translation)	Dairy Brain initiative examples; farm data streams and analytics concepts	Perspective / commentary	Strategic framing of AI applications and translational pathways in dairy science	Conceptual synthesis with practical examples; not an empirical validation paper
[16]	Data foundations & interoperability; Analytics; Governance/trust (enablers)	Industry data streams, integration challenges, and analytics requirements	Narrative synthesis / pathways paper	Challenges and recommended pathways for data integration and analytics at scale	Expert synthesis grounded in sector context; supports architecture and roadmap
[28]	Decision engines (whole-farm simulation); Deployment (user-facing tool)	Farm management data; herd and crop/manure parameters; weather / climate inputs; environmental loss functions (N leaching)	Dynamic whole-farm simulation model with user-friendly software implementation	Profit and environmental outcome projections under alternative management/climate scenarios	Model evaluation/ parameterization described; delivered as a computerized decision-support tool
[29]	Decision engines (integrated simulation); Deployment (scenario analysis)	Farm system parameters; seasonal climate variability inputs; environmental impact accounting	Integrated dairy farm model / scenario simulation	Management strategies to reduce environmental impacts under climate variability; tradeoffs with production / profit	Scenario-based validation and sensitivity analyses; supports decision planning
[23]	Data foundations & interoperability; Analytics readiness	Precision dairy technology data streams; farm information systems context	Data management framework / pipeline design	Operational framework for ingesting, structuring, and analyzing precision dairy data	Framework supported by applied examples; intended to guide implementation
[11]	Decision engines (sustainability DSS); Deployment (stakeholder-facing tool)	Farm inputs describing herd, feeding, manure/crops (as applicable), economics and sustainability indicators	Decision support model integrating sustainability metrics (DairyPrint)	Scenario based sustainability assessment outputs for farms and stakeholders; decision guidance	Model demonstration with case examples; tool-oriented presentation supporting deployment

 doi:10.20944/preprints202603.1613.v1

Table A1. *Cont.*

Paper	Pipeline layer(s)	Data inputs	Method type	Decision output	Validation / deployment evidence
[27]	Decision engines (mitigation planning); Deployment (management options)	Wisconsin dairy system parameters; feeding strategies; manure management options; costs; emission factors	Scenario analysis with economic + GHG accounting	Cost-effective mitigation options and management recommendations	Modeled case-study comparisons with sensitivity to assumptions; actionable mitigation guidance
[31]	Data foundations & interoperability; Governance/standards (constraints)	Industry data-flow context; bottleneck characterization across systems and stakeholders	Bottleneck analysis / narrative synthesis	Identification of key data bottlenecks and practical solutions to enable downstream analytics	Evidence-based synthesis using sector examples; intended to inform platform design
[31]	Analytics (prediction); Deployment (decision workflow potential)	Farm production and management records; mastitis event labels; engineered features	Supervised machine learning (algorithm comparison)	Early mastitis risk prediction to support targeted monitoring/intervention	Cross-validation/ holdout evaluation reported; performance compared across algorithms
[30]	Deployment (benchmarking platform); Analytics; Data foundations (data products)	Economic and performance records; benchmarking indicators; standardized farm data products	Software / platform design + benchmarking methodology	Benchmark dashboards/metrics to support economic decision-making	Platform demonstration with defined metrics; deployment readiness described (user-facing platform)
[37]	Analytics (phenotyping); Decision engines (simulation coupling)	Milk yield time series; herd management covariates; simulation model inputs	Two-step lactation curve estimation + integration into farm simulation	Tailored lactation curves; quantified systemic impact on simulation outputs	Quantitative evaluation comparing curve methods and downstream simulation implications
[12]	Decision engines (optimization)	Crop rotation options; yields; feed ingredient composition; nutrition constraints; farm constraints; policy constraints; prices/costs	Mixed-integer linear programming (MILP) + Duo-objective optimization	Optimal crop + diet plans that reduce enteric methane while meeting constraints and maintaining profitability under different environmental regulations	Case-study optimization results; policy scenario/sensitivity analysis reported
[13]	Decision engines (optimization)	Cow / group nutrient requirement; group number; grouping creteria; feed ingredient composition; prices/costs	Duo-objective linear optimization	Optimal nutritional grouping strategy and ration formulations for seperate groups that consider both GHG emissions and feed cost	Case-study optimization results
[9]	Deployment (user interface); Analytics (LLM/NL); Data foundations (queryable datasets)	Numerical farm databases; user natural-language queries; metadata/schema for retrieval	LLM-enabled natural-language interface to structured data (NL querying)	Conversational access to numeric indicators and summaries to reduce decision friction	Prototype/system evaluation with test interactions; deployment considerations discussed
[44]	Decision engines (digital twin module)	Animal biology and management parameters; feeding, reproduction, growth; physiology relationships	Biophysical/mechanistic modeling (RuFaS animal module)	Simulated animal performance and resource flows as part of integrated farm modeling	Module description with verification/consistency checks and illustrative outputs
[8]	Analytics (anomaly detection); Data foundations (QA/QC)	Daily milk yield time series from dairy cows	Unsupervised machine learning for outlier detection	Outlier flags to improve data reliability and downstream analytics	Quantitative evaluation of detection performance; practical examples for integration into QC workflows
[42]	Decision engines (management strategy); Analytics (supporting economics)	Cow nutrient requirements; feed costs; milk production; grouping constraints	Nutritional grouping strategy evaluation (model-based)	Reduced feed costs and improved profitability via improved grouping	Comparative evaluation of grouping scenarios using realistic herd inputs

 doi:10.20944/preprints202603.1613.v1

Table A1. *Cont.*

Paper	Pipeline layer(s)	Data inputs	Method type	Decision output	Validation / deployment evidence
[40]	Decision engines (platform/digital twin); Deployment (platform framing)	Whole-farm system components (animal, manure, crops, economics); integration architecture	Modeling environment description (integrated dairy system management)	Integrated platform enabling system-level scenario evaluation and decision support	Platform concept and illustrative use cases; emphasizes extensibility and integration
[36]	Analytics (phenotyping/statistical inference)	Large-scale lactation records; temporal / geographic metadata; management covariates	Statistical modeling of lactation curve parameters	Quantified effects of temporal / geographic / management factors on curve parameters; phenotypes for downstream models	Model fit and inference on large dataset; supports generalizable phenotyping
[46]	Decision engines (stochastic simulation)	Heifer and lactating cow reproductive program parameters; transition probabilities; costs; herd structure	Stochastic animal life-cycle simulation within whole-farm context	Value assessment of combined reproductive management programs; economic and performance outcomes	Simulation experiments with comparative scenarios; sensitivity analyses reported
[38]	Analytics (time series monitoring); Deployment (industry monitoring)	State-level milk productivity time series data	Time-series analysis / forecasting	Trends and dynamic patterns in milk productivity; potential early signals for planning	Model-fit and forecast evaluation; supports monitoring / benchmarking use cases
[33]	Decision engines (optimization)	Herd demographics/structure; productivity parameters; environmental performance metrics; economic inputs	Optimization modeling of herd structure and performance tradeoffs	Optimal herd structure and strategies balancing productivity, profitability, and environmental performance	Computational case-study results with scenario comparisons
[25]	Analytics (AI/ML); Deployment (implementation considerations)	Published literature on computer vision and large language models for livestock	Narrative review	Synthesis of applications, opportunities, risks, and deployment considerations for AI in livestock systems	Literature-based synthesis; highlights evaluation and deployment needs
[26]	Analytics (AI/ML); Deployment (integrated knowledge systems)	AI applications and systems literature for animal science	Perspective / conceptual synthesis	Framework for evolution from point solutions to integrated knowledge systems; research roadmap	Conceptual synthesis; supports strategic roadmap rather than empirical validation
[32]	Data foundations (data generation); Analytics (embedded ML)	Milk weighing sensor/scale signals; reference measurements for calibration	Machine learning regression/calibration for measurement device	Improved milk-weight measurement via ML-based weighing scale	Device validation against reference measurements; performance reported for measurement accuracy
[10]	Decision engines (digital twin platform); Deployment (open platform)	Integrated whole-farm modules and datasets; extensible model components	Platform description and roadmap for RuFaS	Open platform to support research and action for sustainable dairy farming; scenario evaluation capability	Platform readiness described; integration of modules and use cases; community/open development pathway
[24]	Data foundations & interoperability; Deployment (operational data products)	Contemporary dairy operation data streams; Precision dairy tech outputs; farm management records	Applied data engineering methods for stream management	Architectures and techniques for integrated, decision-ready data streams to support improved farm decisions	Applied/illustrative implementations and engineering case examples; practical guidance for scaling
[35]	Analytics (prediction)	Daily milk yield; within-herd relatives' records; pedigree/relatedness; herd environment data	Predictive modeling incorporating within-herd relatives to capture G×E	Improved prediction of daily milk yield for primiparous cows; potential for management targeting	Model evaluation via cross-validation/holdout; comparative performance reported

References

1. Intergovernmental Panel On Climate Change (Ipcc)., Ed. *Climate Change 2022 - Mitigation of Climate Change: Working Group III Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 1 ed.; Cambridge University Press, 2023. https://doi.org/10.1017/9781009157926.
2. Bellarby, J.; Tirado, R.; Leip, A.; Weiss, F.; Lesschen, J.P.; Smith, P. Livestock greenhouse gas emissions and mitigation potential in Europe. *Global Change Biology* **2013**, *19*, 3–18. https://doi.org/10.1111/j.1365-2486.2012.02786.x.
3. Berckmans, D. General introduction to precision livestock farming. *Animal Frontiers* **2017**, *7*, 6–11. https://doi.org/10.2527/af.2017.0102.
4. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big Data in Smart Farming – A review. *Agricultural Systems* **2017**, *153*, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023.
5. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; Da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **2016**, *3*, 160018. https://doi.org/10.1038/sdata.2016.18.
6. Klerkx, L.; Jakku, E.; Labarthe, P. A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda. *NJAS: Wageningen Journal of Life Sciences* **2019**, *90-91*, 1–16. https://doi.org/10.1016/j.njas.2019.100315.
7. Morota, G.; Ventura, R.V.; Silva, F.F.; Koyama, M.; Fernando, S.C. BIG DATA ANALYTICS AND PRECISION ANIMAL AGRICULTURE SYMPOSIUM: Machine learning and data mining advance predictive big data analysis in precision animal agriculture1. *Journal of Animal Science* **2018**, *96*, 1540–1550. https://doi.org/10.1093/jas/sky014.
8. Higaki, S.; Freitas, E.N.D.A.; Negreiro, A.; Dórea, J.R.; Cabrera, V.E. Leveraging unsupervised machine learning techniques for detecting outliers in the daily milk yield data of dairy cows. *Journal of Dairy Science* **2025**, *108*, 9696–9711. https://doi.org/10.3168/jds.2024-25889.
9. Gontijo, D.; Rolins Santana, D.; De Assis Costa, G.; Cabrera, V.E.; Noronha De Andrade Freitas, E. Dairy GPT: Empowering dairy farmers to interact with numerical databases through natural language conversations. *Smart Agricultural Technology* **2025**, *12*, 101097. https://doi.org/10.1016/j.atech.2025.101097.
10. Reed, K.; Tricarico, J.; HekmatiAthar, S.; Waddell, J.; Andreen, D.; Briggs, K.; Liu, A.; Tomlinson, N.; Adamchick, J.; Cabrera, V.; et al. The Ruminant Farm Systems (RuFaS) model is a platform to support future research and actions for sustainable dairy farming. *JDS Communications* **2025**, *6*, S15–S22. https://doi.org/10.3168/jdsc.2025-0861.
11. Da Silva, T.E.; Cabrera, V.E. The DairyPrint model: A decision support model to help dairy farmers and other stakeholders toward improved sustainability. *Journal of Dairy Science* **2024**, *107*, 10998–11015. https://doi.org/10.3168/jds.2024-24946.
12. Gong, Y.; Bellingeri, A.; Fumagalli, F.; Sechi, G.S.; Atzori, A.S.; Masoero, F.; Gallo, A.; Cabrera, V.E. A mixed integer linear programming framework for mitigating enteric methane emissions on dairy farms through optimized crop and diet planning. *Journal of Cleaner Production* **2025**, *511*, 145636. https://doi.org/10.1016/j.jclepro.2025.145636.
13. Gong, Y.; Hu, H.; Reed, K.; Noronha de Andrade Freitas, E.; Cabrera, V. Can nutritional grouping lower feed costs and enteric methane production in dairy farms? *JDS Communications* **2026**. https://doi.org/10.3168/jdsc.2025-0919.
14. Barton, R.; Burchard, J.; Cabrera, V.E.; Cook, D.; Cooley, W.; Cue, R.; Fadul, L.; Mattison, J.; Saha, A. Data Ownership and Privacy in Dairy Farming: Insights from U.S. and Global Perspectives. *Animals* **2025**, *15*, 524. https://doi.org/10.3390/ani15040524.
15. Baldin, M.; Bewley, J.M.; Cabrera, V.E.; Jones, K.; Loehr, C.; Mazon, G.; Perez, J.D.; Utt, M.; Weyers, J. Standardization for Data Generation and Collection in the Dairy Industry: Addressing Challenges and Charting a Path Forward. *Animals* **2025**, *15*, 250. https://doi.org/10.3390/ani15020250.
16. Cabrera, V.E.; Bewley, J.; Breunig, M.; Breunig, T.; Cooley, W.; De Vries, A.; Fourdraine, R.; Giordano, J.O.; Gong, Y.; Greenfield, R.; et al. Data Integration and Analytics in the Dairy Industry: Challenges and Pathways Forward. *Animals* **2025**, *15*, 329. https://doi.org/10.3390/ani15030329.
17. Grant, M.J.; Booth, A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal* **2009**, *26*, 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x.
18. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* **2021**, p. n71. https://doi.org/10.1136/bmj.n71.

 doi:10.20944/preprints202603.1613.v1

19. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine* **2018**, *169*, 467–473. https://doi.org/10.7326/M18-0850.
20. Baethge, C.; Goldbeck-Wood, S.; Mertens, S. SANRA—a scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review* **2019**, *4*, 5. https://doi.org/10.1186/s41073-019-0064-8.
21. Cabrera, V. Invited review: Helping dairy farmers to improve economic performance utilizing data-driving decision support tools. *Animal* **2018**, *12*, 134–144. https://doi.org/10.1017/S1751731117001665.
22. Cabrera, V.E.; Barrientos-Blanco, J.A.; Delgado, H.; Fadul-Pacheco, L. Symposium review: Real-time continuous decision making using big data on dairy farms. *Journal of Dairy Science* **2020**, *103*, 3856–3866. https://doi.org/10.3168/jds.2019-17145.
23. Cabrera, V.E.; Fadul-Pacheco, L. Future of dairy farming from the Dairy Brain perspective: Data integration, analytics, and applications. *International Dairy Journal* **2021**, *121*, 105069. https://doi.org/10.1016/j.idairyj.2021.105069.
24. Wangen, S.R.; Zhang, F.; Fadul-Pacheco, L.; Da Silva, T.E.; Cabrera, V.E. Improving farm decisions: The application of data engineering techniques to manage data streams from contemporary dairy operations. *Livestock Science* **2021**, *250*, 104602. https://doi.org/10.1016/j.livsci.2021.104602.
25. Menezes, G.L.; Mazon, G.; Ferreira, R.E.P.; Cabrera, V.E.; Dorea, J.R.R. Artificial intelligence for livestock: a narrative review of the applications of computer vision systems and large language models for animal farming. *Animal Frontiers* **2024**, *14*, 42–53. https://doi.org/10.1093/af/vfae048.
26. Niu, M.; Guo, C.; Cabrera, V.E. Artificial intelligence for animal science: from applications to integrated knowledge systems. *Animal Frontiers* **2025**, p. vfaf054. https://doi.org/10.1093/af/vfaf054.
27. Dutreuil, M.; Wattiaux, M.; Hardie, C.; Cabrera, V. Feeding strategies and manure management for cost-effective mitigation of greenhouse gas emissions from dairy farms in Wisconsin. *Journal of Dairy Science* **2014**, *97*, 5904–5917. https://doi.org/10.3168/jds.2014-8082.
28. Cabrera, V.E.; Breuer, N.E.; Hildebrand, P.E.; Letson, D. The dynamic North Florida dairy farm model: A user-friendly computerized tool for increasing profits while minimizing N leaching under varying climatic conditions. *Computers and Electronics in Agriculture* **2005**, *49*, 286–308. https://doi.org/10.1016/j.compag.2005.07.001.
29. Cabrera, V.E.; Breuer, N.E.; Hildebrand, P.E. North Florida Dairy Farmer Perceptions Toward the Use of Seasonal Climate Forecast Technology. *Climatic Change* **2006**, *78*, 479–491. https://doi.org/10.1007/s10584-006-9053-2.
30. Freitas, E.; Cabrera, V. Dairy Victory Platform: A novel benchmarking platform to empower economic decisions on dairy farms. *JDS Communications* **2025**, *6*, 69–73. https://doi.org/10.3168/jdsc.2024-0617.
31. Fadul-Pacheco, L.; Delgado, H.; Cabrera, V.E. Exploring machine learning algorithms for early prediction of clinical mastitis. *International Dairy Journal* **2021**, *119*, 105051. https://doi.org/10.1016/j.idairyj.2021.105051.
32. Jorge De Paula Nunes Cassimiro, A.; Da Silva Ramos, E.; Cabrera, V.E.; Noronha De Andrade Freitas, E. Milk weighing scale based on machine learning. *Smart Agricultural Technology* **2024**, *7*, 100417. https://doi.org/10.1016/j.atech.2024.100417.
33. Liang, D.; Cabrera, V. Optimizing productivity, herd structure, environmental performance, and profitability of dairy cattle herds. *Journal of Dairy Science* **2015**, *98*, 2812–2823. https://doi.org/10.3168/jds.2014-8856.
34. Barrientos-Blanco, J.A.; White, H.; Shaver, R.D.; Cabrera, V.E. Improving nutritional accuracy and economics through a multiple ration-grouping strategy. *Journal of Dairy Science* **2020**, *103*, 3774–3785. https://doi.org/10.3168/jds.2019-17608.
35. Zhang, F.; Weigel, K.; Cabrera, V. Predicting daily milk yield for primiparous cows using data of within-herd relatives to capture genotype-by-environment interactions. *Journal of Dairy Science* **2022**, *105*, 6739–6748. https://doi.org/10.3168/jds.2021-21559.
36. Li, M.; Rosa, G.; Reed, K.; Cabrera, V. Investigating the effect of temporal, geographic, and management factors on US Holstein lactation curve parameters. *Journal of Dairy Science* **2022**, *105*, 7525–7538. https://doi.org/10.3168/jds.2022-21882.
37. Gong, Y.; Hu, H.; Reed, K.; Cabrera, V. Advancing dairy farm simulations: A 2-step approach for tailored lactation curve estimation and its systemic impacts. *Journal of Dairy Science* **2025**, *108*, 9681–9695. https://doi.org/10.3168/jds.2025-26334.
38. Li, M.; Reed, K.; Cabrera, V. A time series analysis of milk productivity in US dairy states. *Journal of Dairy Science* **2023**, *106*, 6232–6248. https://doi.org/10.3168/jds.2022-22751.

 doi:10.20944/preprints202603.1613.v1

39. Aguirre-Villegas, H.; Passos-Fonseca, T.; Reinemann, D.; Armentano, L.; Wattiaux, M.; Cabrera, V.; Norman, J.; Larson, R. Green cheese: Partial life cycle assessment of greenhouse gas emissions and energy intensity of integrated dairy production and bioenergy systems. *Journal of Dairy Science* **2015**, *98*, 1571–1592. https://doi.org/10.3168/jds.2014-8850.
40. Kebreab, E.; Reed, K.F.; Cabrera, V.E.; Vadas, P.A.; Thoma, G.; Tricarico, J.M. A new modeling environment for integrated dairy system management. *Animal Frontiers* **2019**, *9*, 25–32. https://doi.org/10.1093/af/vfz004.
41. Cabrera, V. A large Markovian linear program to optimize replacement policies and dairy herd net income for diets and nitrogen excretion. *Journal of Dairy Science* **2010**, *93*, 394–406. https://doi.org/10.3168/jds.2009-2352.
42. Kalantari, A.; Armentano, L.; Shaver, R.; Cabrera, V. Economic impact of nutritional grouping in dairy herds. *Journal of Dairy Science* **2016**, *99*, 1672–1692. https://doi.org/10.3168/jds.2015-9810.
43. Barrientos-Blanco, J.A.; White, H.; Shaver, R.D.; Cabrera, V.E. Graduate Student Literature Review: Considerations for nutritional grouping in dairy farms. *Journal of Dairy Science* **2022**, *105*, 2708–2717. https://doi.org/10.3168/jds.2021-21141.
44. Hansen, T.L.; Li, M.; Li, J.; Vankerhove, C.J.; Sotirova, M.A.; Tricarico, J.M.; Cabrera, V.E.; Kebreab, E.; Reed, K.F. The Ruminant Farm Systems Animal Module: A Biophysical Description of Animal Management. *Animals* **2021**, *11*, 1373. https://doi.org/10.3390/ani11051373.
45. Cabrera, V.E. Artificial intelligence applied to dairy science: insights from the Dairy Brain Initiative. *Animal Frontiers* **2024**, *14*, 60–63. https://doi.org/10.1093/af/vfae040.
46. Li, M.; Reed, K.; Lauber, M.; Fricke, P.; Cabrera, V. A stochastic animal life cycle simulation model for a whole dairy farm system model: Assessing the value of combined heifer and lactating dairy cow reproductive management programs. *Journal of Dairy Science* **2023**, *106*, 3246–3267. https://doi.org/10.3168/jds.2022-22396.