

Review

From Decision-Support Tools to Digital Twins: A Review of 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 Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA; gong44@wisc.edu

² Department of Informatic, Instituto Federal de Goiás, Goiânia 74055-110, Brazil; noronha@ifg.edu.br

* Correspondence: vcabrera@wisc.edu

Abstract

Sustainability targets for livestock require decision support that is both scientifically credible and operationally usable on farms. This integrative narrative review synthesizes the broader peer-reviewed literature on digital farming, artificial intelligence and machine learning, simulation modeling, optimization, and digital-twin concepts as applied to sustainable dairy systems, and uses selected peer-reviewed dairy studies from one integrated research program as illustrative worked examples that show how these elements can be connected end-to-end. We organize the synthesis around a data-to-decision pipeline that links data foundations and interoperability, governance and trust, analytics, decision engines, and deployment, comparing model classes by data needs, temporal resolution, interpretability, and deployment maturity. Recurring barriers to impact—weak ground truth, data drift, fragmented identifiers, and misaligned incentives—are highlighted alongside the design principles that address them. The contribution of the review is the transferable pipeline framework, demonstrated through worked examples drawn from one integrated research program; the program's studies appear repeatedly because they together trace decisions across all pipeline layers, not because they constitute the field. We conclude with a practical roadmap and implementation checklist for designing and scaling decision-intelligence systems with 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



Academic Editors: Tiziano Gomiero, Roberto Mancinelli, Steve W. Lyon, Ana Isabel Roca-Fernández and Magdalena Arévalo Turrubiarte

Received: 18 March 2026

Revised: 30 June 2026

Accepted: 3 July 2026

Published: 7 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

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. These expectations align with broader livestock and climate-change mitigation literature, which emphasizes that sustainability strategies must be evaluated in whole-system context rather than as isolated interventions [1–4]. In dairy systems, 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 [5,6]. Therefore, the central challenge is not only to measure

sustainability outcomes, but to connect measurement with decisions that are feasible, timely, and meaningful at the farm level.

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 [7–10]. Recent bibliometric and synthesis work documents the global expansion of PLF research and its convergence with biometric sensing, big data, and connected platforms in digital livestock systems [11–14]. These developments sit within a broader push in digital agriculture toward integrated farm-level data systems, including the cluster of sensor, cloud, and analytics technologies often labeled Agriculture 4.0 and the more recent extension that adds on-farm artificial intelligence (AI) and autonomy, sometimes called Agriculture 5.0 [15–17]. Digital-agriculture evidence and adoption are more established in row-crop production than in livestock systems, where adoption data remain more limited and many technologies are still emerging. Dairy systems also pose distinct integration challenges, including high-frequency individual-animal data, fragmented sensor and management records, cow-level biological variability, bioeconomic decision tradeoffs, and animal-welfare considerations that require near-real-time interpretation [18–21].

However, in practice, the bottleneck to PLF is often not simply data availability but interoperability, data quality, and the sociotechnical conditions required to make multi-source data reliable and reusable for decision support [19,22–24]. This is consistent with the broader FAIR data literature, which emphasizes that data and related digital objects must be findable, accessible, interoperable, and reusable to support transparent, reproducible, and reusable science [25]. Important progress has been made in translating these principles for agricultural data systems [26], linking FAIR sub-indicators to sustainable agricultural performance [27], mapping interoperable data standards for precision livestock farming [19], and outlining recommendations for farmer data privacy and confidentiality [28]. Even so, much of the FAIR and interoperability literature remains focused on data stewardship, standards, and governance, with fewer studies demonstrating the full pipeline from interoperable data foundations to analytics, decision engines, and on-farm action.

When interoperable and quality-controlled data streams are available, the next layer of value comes from analytics. AI and machine learning (ML) methods can help convert raw data streams into decision-relevant signals, including risk prediction, anomaly detection, phenotyping, and performance assessment, and have become increasingly important in precision animal agriculture [29–32]. Recent systematic and scoping reviews document this trajectory in dairy specifically: ML has been applied across disease detection, milk-yield prediction, and milk-quality assessment [33]; deep-learning architectures have been used for cattle identification, health monitoring, and livestock computer vision [34–36]; and the broader AI-in-dairy landscape has been mapped across 151 papers in a recent scoping review [37]. 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” [38], and natural-language interfaces to reduce friction when interacting with complex numerical databases, while raising new needs for verification and safe deployment [39].

Beyond prediction and monitoring, sustainability decisions also require prescriptive and prospective capabilities that quantify tradeoffs and identify feasible strategies under constraints. That capability increasingly comes from decision engines that combine optimization, simulation, and emerging digital-twin approaches. Recent reviews articulate conceptual frameworks, applications, implementation barriers, and research agendas

for digital twins in agriculture and livestock [40–43], the operations-research and bioeconomic modeling literature for beef and dairy farming [44–46], and the optimization toolkit for dairy and livestock systems [47–49]. In dairy, platform-scale modeling efforts [50], sustainability-oriented decision tools [51], and optimization frameworks for implementable mitigation strategies [52,53] 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 [23,54,55], and on socio-ethical considerations documented in the social-science literature on digital agriculture [56–63].

Previous literature has made important contributions to specific parts of this problem. Reviews of PLF, smart farming, and digital agriculture have clarified the roles of sensing, connectivity, advisory systems, and data-intensive management [7,8,13,56]. Recent reviews of AI/ML in dairy have mapped the predictive opportunities created by large and complex farm datasets [29,33,37], and reviews of decision engines in agricultural and livestock systems have cataloged modeling, simulation, digital-twin, and optimization tools for farm-management decisions [42,44,47]. However, these streams of literature often treat data generation, analytics, decision modeling, and adoption as separate topics. In contrast, the contribution of this review is to connect these elements into a single data-to-decision pipeline 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 synthesize how data foundations, governance, analytics, decision engines, and deployment practices can be combined to support scientifically credible, user-centered, and scalable sustainability decisions.

The remainder of the manuscript is organized as follows. Section 2 defines the review approach, scope, and the two evidence roles (external peer-reviewed literature and program-authored worked examples). Section 3 presents the data-to-decision pipeline and synthesizes evidence on data foundations, governance, and analytics. Section 4 examines decision engines, with emphasis on optimization, simulation, and digital-twin concepts. Section 5 focuses on deployment requirements, including validation, monitoring, adoption, and user-centered design. Finally, Sections 6 and 7 translate the synthesis into a practical roadmap, an implementation checklist, and a concluding outlook for sustainable dairy systems.

2. Review Approach and Scope

This article is an integrative narrative review of digital farming, AI, and decision-intelligence research for sustainable dairy systems. Two distinct bodies of evidence inform the synthesis. First, an external peer-reviewed literature base—reviews and methodological papers on precision livestock farming [7,9–12,36,64,65], smart farming and Agriculture 5.0 [8,15,60], ML in animal agriculture [14,29–35,37], data governance and FAIR principles [19,24–28,61–63], simulation modeling and whole-farm bioeconomic models [44–46,66], optimization in dairy and livestock systems [47–49], digital twins in agriculture [40–43,67], ML validation and deployment [68–74], and user-centered design and adoption [56–59,75–78]—establishes the broader landscape and the methodological standards against which approaches are compared. Second, a curated set of peer-reviewed dairy studies from one integrated research program provides worked examples that illustrate how the heterogeneous elements of the pipeline can be operationalized end-to-end. The program studies are used to illustrate integration, not to represent the field; field-level claims rest on the external literature. This approach is appropriate because no single research stream provides end-to-end pipeline traceability in the published dairy literature, and conceptual integration across diverse evidence types requires both broad external

grounding and concrete worked examples. Narrative and scoping approaches are commonly used when the purpose is conceptual integration across diverse evidence types rather than statistical aggregation or complete enumeration of all studies [79–81].

2.1. Two Evidence Roles: External Literature and Worked Examples

The synthesis draws on two complementary bodies of evidence. External peer-reviewed reviews and methodological papers establish the broader landscape and the methodological standards used throughout the manuscript. They are cataloged by pipeline layer in Appendix A Table A1. A curated set of 37 peer-reviewed dairy studies from one integrated research program is used as illustrative case studies that show how the pipeline layers can be connected end-to-end. These worked examples are cataloged in Appendix A Table A2. External literature defines what is established for each pipeline layer in the field. The worked examples show how the layers can be integrated within a single coordinated research program.

External peer-reviewed reviews and methodological papers were selected through a purposive, layer-by-layer search rather than a single systematic database protocol, consistent with the design of an integrative narrative review. For each pipeline layer (data foundations, governance, analytics, decision engines, deployment), we sought peer-reviewed sources that either established methodological standards, surveyed the state of the literature through systematic or scoping reviews, or articulated a conceptual framework relevant to the layer. We prioritized recent (post-2020) reviews and methodological papers, supplementing them with widely cited pre-2020 foundational references where no recent equivalent of comparable authority was available. We did not include primary empirical studies unless they were canonical references for a specific method or principle. Institutional and gray-literature reports were not used to support methodological claims and appear in the Introduction only as landscape context.

The program-authored worked examples were selected deliberately because they allow the review to follow how assumptions, data structures, validation needs, decision outputs, and deployment considerations propagate across connected layers within a single coordinated research program. This matters because isolated papers on sensors, models, optimization, or decision tools do not by themselves show how these components can be linked into operational decision-intelligence systems for sustainable dairy production. Many dairy studies of comparable scientific quality exist in the broader literature, but they tend to address individual pipeline layers rather than the end-to-end pipeline. The worked examples are therefore used as illustrative dairy applications, not as a claim that the authors' work represents the entire field; the field-level evidence base is provided by the external peer-reviewed reviews and methodological papers cited throughout and cataloged in Table A1.

Each worked example 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.

The resulting set of worked examples combines foundational studies with recent contributions spanning governance and standardization, AI-enabled analytics and interfaces, anomaly detection, and whole-farm digital-twin and optimization frameworks [38,39,50,53–55].

This structure supports the manuscript's goal of identifying generalizable design requirements for sustainable dairy decision-intelligence systems while maintaining transparency about the scope and limits of the evidence base.

2.2. Scope Boundaries

Dairy systems, including confined and grazing operations, 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 broader design principles—interoperability, governance, validation, monitoring, and usability—that may also inform work in other livestock sectors. However, this manuscript does not claim direct validation beyond dairy systems, and any transfer to other species would require adaptation to species biology, management conditions, and data ecosystems. The worked examples are not intended to represent the full digital-livestock literature exhaustively; the broader landscape is established through the external peer-reviewed reviews and methodological papers cataloged in Appendix A Table A1, while the worked examples illustrate end-to-end pipeline operationalization in dairy.

2.3. Evidence Organization and Synthesis Method

Consistent with the integrative narrative-review design, sources are organized into two complementary roles. External peer-reviewed reviews and methodological papers establish the landscape and the methodological reference standards used in the synthesis; they are cataloged by pipeline layer in Appendix A, Table A1. Studies from the integrated research program are organized in Appendix A, Table A2, as worked examples that illustrate how the pipeline layers can be operationalized in dairy systems. 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 external literature and worked examples via Appendix A Tables A1 and A2 [80,82]—and use a narrative-review quality lens (e.g., explicit rationale, structured synthesis, and balanced limitations) to support readability and rigor [81]. A formal risk-of-bias scoring instrument was not applied because the synthesis aim is conceptual integration rather than effect-size estimation; methodological transparency is provided through explicit reporting of source roles, the two-table inventory in the appendix, and Figure 1 as the synthesis diagram showing how the evidence was organized across the data-to-decision pipeline.

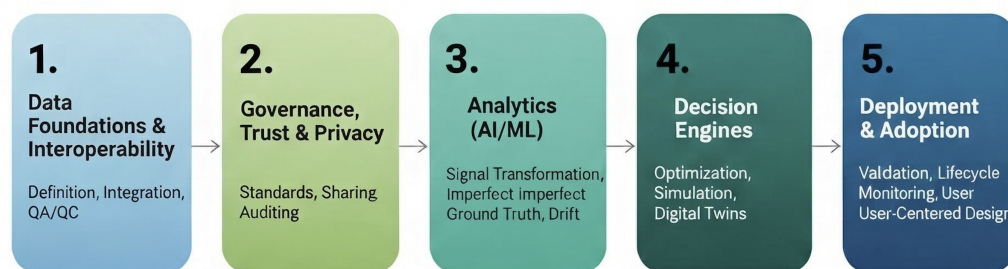


Figure 1. Data-to-decision pipeline architecture for sustainable dairy systems.

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 such as herd records, milking-system outputs, and sensor streams with analytics, bioeconomic reasoning, and practical constraints. The goal is to ensure that recommendations match real management levers and decision cadence [83,84]. Value is realized when multi-source streams are engineered into interoperable, analytics-ready inputs that tools can consume reliably. Without that integration, the same information is repeatedly rebuilt in ad hoc pipelines [22,85].

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/quality control (QA/QC)) [22,84]; (2) governance, trust, privacy, and standards that enable sharing, auditing, and reuse at scale [54,55]; (3) analytics and AI/ML models that transform raw data into decision-relevant signals while accounting for imperfect ground truth and drift [86,87]; (4) decision engines—optimization and whole-farm simulation/digital-twin concepts—that translate signals and constraints into implementable strategies and explicit tradeoffs [50,52]; and (5) deployment practices (validation, lifecycle monitoring, and user-centered design) that determine adoption and sustainability impact [83,84]. Each layer is necessary; value is realized only when outputs are decision-relevant and maintained over time.

Figure 1 summarizes the central conceptual contribution of this review: sustainability-oriented digital farming should be understood as an end-to-end pipeline rather than as a collection of isolated tools. The figure shows that value is created sequentially and cumulatively across layers. Data foundations and interoperability provide the conditions for consistent, analytics-ready inputs; governance and standards determine whether those data can be shared, audited, and reused with trust; analytics transform raw data into decision-relevant signals; and decision engines translate those signals into implementable strategies and explicit tradeoffs. This relationship is central to the framework: AI/ML analytics generate decision-relevant signals, decision engines convert those signals into candidate actions and quantified tradeoffs, and deployment determines whether those outputs can be validated, monitored, and sustained in routine farm use. The final deployment layer is not an afterthought but the mechanism through which validation, monitoring, and user-centered design determine whether the system delivers durable value in practice. In this way, Figure 1 is intended to make explicit why fragmented solutions often fail: weaknesses in any one layer can prevent scientifically sound models from becoming reliable, scalable, farm-facing decision systems.

3.1. Data Foundations and Interoperability

Digital sustainability decision-making is constrained by the quality, completeness, and interoperability of farm data. Data foundations for dairy decision-intelligence sit within a broader peer-reviewed literature on interoperable data standards in PLF [19], FAIR principles for scientific data management [25], FAIR principles applied to agri-food data ecosystems [26,27], and applied data engineering for dairy data streams [22,23]. Recent reviews emphasize that data findability, accessibility, interoperability, and reusability are increasingly practical prerequisites for farm-scale decision-support systems, rather than abstract data-management ideals [25–27]. 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 [22,84]. 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 [23,85].

Data Streams, Integration, and Quality Control

Across the pipeline, the most decision-relevant inputs usually 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 such as prices, weather, and regulations [83,84,88]. These integrated inputs are essential because environmental and economic outcomes are co-determined at the 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 [5,6]. 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 [89].

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 cadences (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 [22]. 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 [23,55], and these bottlenecks remain a primary obstacle to scalable analytics and deployment [90].

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 a minimum, this typically includes range and consistency checks, temporal-logic validation (e.g., event ordering), duplicate detection, completeness/latency monitoring, and auditable transformation

logs [22,84]. 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 [38]. Similar principles apply to sensor 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 [91]. 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 [22,55].

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 [22,85]. Benchmarking platforms illustrate the point: their value depends on consistent data definitions and repeatable pipelines that update indicators without repeated manual wrangling [89]. 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 [23,84].

3.2. Governance, Trust, Privacy, and Standards

Interoperable, quality-controlled data are necessary but not sufficient for scalable decision intelligence. Agricultural data governance has emerged as a distinct field with its own peer-reviewed literature, articulating the privacy, confidentiality, ownership, access, and control constraints that bear directly on dairy decision intelligence [28,92], and complementing the long-standing social-science literature on adoption of digital agriculture that links governance, trust, inclusion, and responsible innovation to technology adoption [56–58]. 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 [54,85].

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 [54]. 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 [84,85]. 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 [23].

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 [54]. 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 [39,86].

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 [22,55]. 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 [23,89].

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 [83,89]. Conversely, misaligned incentives—where data value accrues primarily to third parties—can erode participation even when technical integration is feasible [54]. 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 [86,87].

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 [54,55,83].

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

Analytics and AI/ML models convert integrated farm data into decision-relevant signals, including predictions, risk scores, phenotypes, benchmarks, and anomaly flags, that can be acted upon within routine management cycles. ML for dairy systems has expanded rapidly in recent years; recent systematic reviews provide important external scaffolding. Slob et al. [33] synthesize 38 ML studies covering disease detection, milk-yield prediction, and milk-quality assessment, and Mahmud et al. [34] review deep-learning applications for precision cattle farming, including cattle identification and health monitoring. A 2025 scoping review by Palma et al. [37] maps 151 AI and analytics studies in dairy systems, while Oliveira et al. [35] provide complementary depth on deep-learning-based computer-vision architectures for livestock phenotyping and precision management. Together, these reviews provide the methodological context against which the worked examples in this section are positioned. In sustainable dairy systems, analytics create 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 [83,84]. They should therefore be treated as sustainability-enabling signals rather than endpoints. Their practical value emerges when predictions are translated into constrained decisions that explicitly quantify environmental gains and economic consequences [52,53,93]. 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).

Not all decision-relevant analytics require AI/ML. Conventional statistical or time-series models remain useful when relationships are relatively structured, sample sizes are limited, or the main objective is interpretable effect estimation and trend characterization. AI/ML becomes especially useful when dairy data are high-dimensional, nonlinear, multi-modal, or weakly structured, as in sensor streams, anomaly detection, and flexible pattern recognition [29]. In practice, these approaches are complementary rather than competing: statistical models often provide transparent baselines and interpretable summaries, whereas AI/ML can improve prediction and feature extraction when data complexity justifies the added flexibility [83,84].

Table 1. Comparison of model and decision-support classes (AI/ML, simulation/digital twins, and optimization): data needs, temporal resolution, interpretability, and deployment maturity. This table is an integrative synthesis combining external peer-reviewed reviews and methodological papers (Appendix A Table A1) with the curated set of 37 program-authored worked examples (Appendix A Table A2). Representative citations are included where most directly relevant; corresponding worked examples are expanded in Table 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. External: [33–35,37]. Worked examples: [38,39,94].	Prospective “what-if” outcomes over time: economics, emissions, nutrients, herd dynamics; supports strategy evaluation and policy/scenario analysis. External: [40–46]. Worked examples: [50,88,89,93].	Implementable plans under constraints: group/ration/crop-diet allocations, mitigation strategies; supports “what to do given limits” decisions. External: [44,47–49]. Worked examples: [50,52,53,95].
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	From real-time to daily/weekly updates; performance is sensitive to data latency and drift.	Weekly–annual (or longer) horizons; can ingest daily summaries but outputs are often strategic/seasonal.	Decision-interval dependent (daily to seasonal); typically run on a scenario basis rather than continuously.
Interpretability and 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.

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 [83]; “real-time continuous decision-making” vision [84]
Ingestion and 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 [22]; Dairy Brain integration perspective [85]
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 [38]; “big-data” continuous decision context [84]
Governance and 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 [23,54,55,85]
Privacy and 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 [85]; LLM (large language models)/NLP (natural language processing models) interfaces motivate stronger controls [39]
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 [94]; tailored lactation-curve estimation for simulation impacts [96]
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 [52]; dual-objective optimization of nutritional grouping under ingredient constraints [53]
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 [23,88,93]; RuFaS platform concept and scaling [50]
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 [38]; “continuous decision-making” implies lifecycle monitoring [84]

Table 2. *Cont.*

Domain (Pipeline Element)	Minimum Requirement	Practical Checks/Evidence to Document	Worked Examples
Reporting and 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 [85]; tool-oriented reporting in DairyPrint [51]
UX and 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 [89]; decision-support tool design principles [83]; DairyPrint tool deployment [51]
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 [39]; AI integration perspective [86,97]

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 [94]. In that study, the best continuous model correctly classified 85% of clinical mastitis cases with 72% overall accuracy, providing a concrete example of how integrated farm data can support earlier and more targeted intervention [94]. 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 [98]. 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.

A concrete worked example is early mastitis risk prediction from routinely collected farm data. In pipeline terms, the case begins with integrated production and management records, applies supervised learning to identify cows at elevated risk, and returns a prioritized monitoring signal that can be used before clinical detection [94]. This example illustrates how the analytics layer becomes practically relevant only when model output is translated into a manageable action, namely, targeted observation and earlier intervention for cows most likely to benefit.

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 [38]. 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 [22,84].

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 [99]. Such phenotypes become inputs to higher-level decision engines by improving the 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 [96]. For example, when evaluated on 10 commercial dairy farms, this calibration approach reduced the root mean square percentage error for estimating annual herd milk production from 40.6% to 2.22%, illustrating how improved phenotyping can materially increase simulation fidelity [96].

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 [100]. 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 [89]. Interpretation of benchmarking results should also remain explicit about the 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 [101]. 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 [54,55].

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 [102]. 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 [86]. 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 [39]. Conceptually, this aligns with emerging views that animal-agriculture AI is moving toward integrated knowledge systems that combine data access, analytics, and decision workflows [87].

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 and welfare outcomes; dataset shift across herds and seasons; and model drift as technology, genetics, and practices evolve. These issues make validation, drift monitoring, and adoption central requirements rather than optional concerns. This emphasis is consistent with broader ML deployment literature: Paleyes et al. [68] survey the case-study literature on deploying machine learning at scale; Bayram et al. [71] formalize a taxonomy of performance-aware drift detectors; Sahiner et al. [69] articulate the implications of data drift in deployed ML and propose remediation strategies originally developed for medical ML but transferable to agricultural ML, and Kapoor and Narayanan [70] document how data leakage contributes to reproducibility failures in ML-based science. Therefore, analytics intended for decision support should be evaluated using validation designs that reflect deployment reality, including external herds when feasible, temporal holdouts, and sensitivity to missingness and noise, with calibration and thresholds tied to action capacity [83,84]. Ongoing monitoring is also essential: data drift, such as changing sensor baselines or recording practices, and concept drift, such as 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 [83,89].

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 [50,52]. 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?). Optimization is strongest when the feasible action space, objectives, and constraints can be stated explicitly and the goal is to identify an implementable best course of action. Simulation is strongest when biological feedback, time dynamics, or uncertainty must be propagated to evaluate the likely consequences of candidate strategies over time. Emerging digital-twin concepts extend simulation by emphasizing continuous updates from farm data streams and repeated re-evaluation as conditions change [50,84,103].

4.1. Prescriptive Optimization for Implementable Strategies

Optimization for dairy and livestock systems sits within a broader operations-research literature that has been synthesized in several recent reviews. Bang et al. [44] survey 110 decision-support papers in beef and dairy farming (2016–2022) and articulate the modeling choices that distinguish optimization from simulation; Palma et al. [47] catalog the mathematical methods applied to dairy cow replacement decisions, including dynamic programming, stochastic simulation, and machine learning; Mijić et al. [48] present a web-based linear programming (LP) ration-formulation tool; and Das et al. [49] review classical and metaheuristic optimization techniques across livestock, poultry, and aquaculture feed

formulation. Optimization is particularly useful when farms must allocate limited resources (feed, land, labor, replacements) across competing objectives. A central advantage is that optimization can support 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 [104]. 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 [93]. 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 [52,53,93].

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 [52]. This kind of formulation is particularly relevant for 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 [105], 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 [95]. A synthesis of grouping considerations further underscores that “optimal” grouping depends on objectives, constraints, and the data available to define groups reliably [106]. 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 [53]. Specifically, optimization without nutritional grouping reduced feed cost by USD 2.52/cow per day (32%) under the cost-minimization objective and reduced enteric methane by 57 g/cow per day (12%) under the methane-minimization objective. These quantitative results further support framing nutritional grouping and diet optimization as an operational decision engine for achieving multiple objectives under practical farm constraints [53].

4.2. Prospective Simulation for Tradeoff-Aware “What-If” Evaluation

Whole-farm simulation is 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 such as nutrition, manure handling, reproductive strategy, and replacement policy, and to quantify downstream impacts on both economic and environmental metrics. This role sits within a broader bioeconomic and decision-support modeling literature: Menendez et al. [45] describe a

five-step mathematical-modeling pipeline for confined and extensive precision livestock production, Kearney et al. [46] integrate bioeconomic, GHG, and feed-food competition outcomes in pasture-based dairy-beef systems, and Bang et al. [44] situate these contributions within the broader decision-support modeling landscape. These external models provide methodological reference points for the program-authored simulation worked examples that follow.

4.2.1. Whole-Farm Bioeconomic Simulation as a Foundation

Integrated dairy-farm models have long demonstrated the value of combining biological and economic processes to evaluate strategies that affect profitability and nutrient outcomes [5]. Extensions that incorporate seasonal climate variability further show why prospective modeling matters for sustainability: expected outcomes can change substantially as external conditions shift, with consequences for both resource use and environmental performance [5]. Likewise, simulation-based evaluations of greenhouse gas mitigation indicate that cost-effectiveness depends on interactions among feeding strategies, manure management, and farm-specific baselines [88]. Taken together, these studies show that sustainability recommendations should be evaluated prospectively at the whole-farm level rather than through isolated component analyses [5,6,88].

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

Several sustainability-relevant decisions operate through herd dynamics and long-term 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 [107]. 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 [96]. 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 management processes that enables more consistent representation of production and management across scenarios [108], supporting whole-farm evaluation and linking management interventions to outcomes [50].

4.2.3. Operational Sustainability Tools as Simulation “Products”

Decision-support models developed for stakeholder use show how simulation outputs can be packaged into forms that are operationally interpretable. For example, the DairyPrint model demonstrates how integrated modeling can support sustainability assessment for farms and stakeholders by translating scenario outcomes into actionable metrics and comparisons [51]. Platform-level developments further suggest that these tool-oriented outputs can be embedded within broader digital-twin ecosystems, thereby enabling repeated scenario testing and continuous sustainability improvement at both farm and network levels [50,51]. A consistent implication is that simulation outputs must be presented in a way that aligns with decisions, namely what changes, by how much, and under which assumptions, rather than as technical model outputs alone.

A worked example is DairyPrint, which packages whole-farm simulation into a stakeholder-facing decision tool. Rather than returning only technical model states, the tool compares management scenarios using sustainability indicators that users can interpret directly, allowing farmers and advisors to examine how candidate changes alter environ-

mental and economic outcomes under a common modeling framework [51]. This case illustrates the transition from simulation as a research engine to simulation as a practical product for farm and advisor decision-making.

4.3. Toward Digital Twins and Hybrid Decision Intelligence

Digital twins in agriculture have advanced from conceptual proposals toward operational frameworks and applied use cases. Verdouw et al. [40] provide a foundational conceptual framework with applications across five smart-farming use cases, including dairy; Pylaniadis et al. [41] document evidence on digital-twin added value and provide an adoption roadmap; Purcell and Neubauer [42] catalog the state of the art across agricultural production and supply chain applications; and Arulmozhi et al. [43] emphasize the potential of digital twins in agrifood systems while highlighting implementation barriers. The decision-intelligence framework articulated in this review is consistent with the integration patterns documented in these reviews. Digital-twin concepts build on simulation by emphasizing continuous updating from live data streams, repeated recalibration, and ongoing scenario evaluation as farm conditions evolve [84,103]. In practice, the most actionable pathway is often a hybrid form of decision intelligence in which analytics provide timely signals, such as risk scores, phenotypes, and anomaly flags, while optimization and simulation translate those signals into implementable strategies and explicit tradeoffs. Platform approaches such as RuFaS reinforce this direction by positioning the system as a foundation for future research and action in sustainable dairy farming, where consistent data pipelines and modular components enable repeated and comparable “what-if” evaluation [50,108]. Across the evidence reviewed here, three principles consistently emerge for decision-engine design in sustainability contexts: objectives, constraints, and assumptions should be explicit so tradeoffs remain transparent; outputs should map to controllable levers and operational cadence; and uncertainty should be reported because sustainability decisions are made under biological, climatic, and market variability [5,6,83,88]. These principles motivate the deployment requirements discussed in Section 5.

A more explicit comparison across recent studies further clarifies the strengths and limitations of the main model classes. In predictive analytics, meaningful animal-level performance is possible, as illustrated by early mastitis prediction that correctly classified 85% of cases with 72% overall accuracy; however, such models still face challenges in transportability across herds and conditions [94,98]. In simulation-oriented modeling, improving biological parameterization can materially strengthen system fidelity, as shown by lactation-curve calibration that reduced root-mean-square percentage error in annual herd milk prediction from 40.6% to 2.22% [96]. In optimization, recent work demonstrated practical multi-objective value by reducing feed cost by USD 2.52/cow per day under a cost-minimization objective and enteric methane by 57 g/cow per day under a methane-minimization objective [53]. Taken together, these studies show that the field now offers credible quantitative examples within individual model classes, but still contains relatively few empirically validated studies that connect prediction, whole-farm decision engines, and real deployment in one integrated farm-facing workflow.

Viewed comparatively, the reviewed evidence suggests a consistent tradeoff across model classes. AI/ML studies are strongest for extracting short-horizon signals from large and heterogeneous data streams, but many remain limited by weak interpretability, external validation, and uncertain transportability across farms and management contexts [29,38,94]. Optimization studies provide the clearest action recommendations because objectives and constraints are explicit, yet their practical value depends heavily on whether those constraints adequately represent real farm conditions and stakeholder preferences [52,53]. Simulation and digital-twin approaches offer the broadest capacity to represent whole-farm

feedbacks and long-run sustainability tradeoffs, but they are more demanding in terms of calibration, assumptions, and continuous data integration [50,51]. Across the worked examples, the main gap is therefore not the absence of promising individual methods, but the limited number of studies that connect validated analytics, decision engines, and deployment evidence within one integrated farm-facing workflow.

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 [83,84,109].

5.1. Validation Aligned to Intended Use

A growing body of peer-reviewed literature documents the challenges of validating and deploying machine learning systems in real-world settings, including issues such as data leakage and the gap between performance on held-out splits and performance after deployment [68,70]. 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. The validation practices discussed below for dairy worked examples address several of these concerns, including temporal holdouts, external-herd evaluation, and calibration to observed event rates. 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 [84,94]. 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 [38].

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 [50,51,53]. 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 [52,88,101]. Taken together, these examples illustrate two practical validation scenarios for the framework: animal-level deployment, where predictive analytics support targeted monitoring and earlier intervention [94], and farm-level deployment, where simulation-based tools support stakeholder-facing comparison of management scenarios and sustainability outcomes [50,51].

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 the following:

- Data provenance and definitions: Sources, time span, sampling frequency, event/label definitions, and inclusion/exclusion rules [22,55].
- Preprocessing and feature construction: Unit conversions, aggregation windows, missing-data handling, and QA/QC rules [38,84].
- Model or engine specification: Algorithms/parameters (for AI/ML), objective functions/constraints (for optimization), module assumptions (for simulation), and software versions [52,53,83].
- Performance and uncertainty: Metrics appropriate to the use case, calibration where relevant, and evidence of robustness across time/farms [84,94].
- Versioning and change logs: What changed, when, and the expected effect on outputs—critical when tools are used for benchmarking or sustainability reporting [85,89].

These elements are also essential for governance and auditing, especially as platforms integrate multiple services and stakeholders [23,54].

5.3. Monitoring for Data Drift and Model Drift

Data and concept drift are well-characterized in the broader machine-learning literature: Sahiner et al. [69] review implications and remediation strategies for data drift in deployed ML, and Bayram et al. [71] provide a taxonomy of performance-aware drift detectors. The monitoring approaches recommended here for dairy decision-intelligence draw directly on this external methodological foundation. 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 [22,84].
- Model monitoring: Tracking calibration and performance on recent data, comparing “champion–challenger” models, and defining triggers for recalibration or retraining [83,84].

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 [38]. 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 [39,54].

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

Adoption of digital agriculture is shaped by socio-ethical and institutional factors that are now documented in a substantial social-science literature. Klerkx et al. [56] provide the canonical review of social science on digital agriculture; Klerkx and Rose [57] frame Agriculture 4.0 in terms of responsible innovation and food-system transition pathways; Eastwood et al. [58] apply responsible-innovation principles to smart-farming development; and Werkheiser [59] articulates under-examined risks of precision livestock farming, including privacy, deskilling, and animal-welfare implications. The user-centered design

features below are positioned against this external literature. Adoption depends on whether the system fits real workflows for farmers and advisors. Evidence from dairy decision-support tools suggests that practical usefulness, usability, and fit with management needs are important for uptake, especially when tools add data-entry burden or produce unstable recommendations [83,109]. Across dairy decision-support and digital platform work, several design features consistently support adoption:

- **Actionability:** Outputs should map directly to decisions (e.g., prioritized cow lists, group/ration recommendations, scenario comparisons) and specify what action is feasible now [53,83,89]. 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 [52,88,93].
- **Workflow integration:** Reduce manual entry; rely on interoperable pipelines and standardized data products so outputs update automatically [22,85].
- **Interpretability and “why”:** Provide key drivers of predictions and binding constraints in optimization; avoid black-box outputs that cannot be interrogated [52,83].
- **Uncertainty communication:** Show confidence/uncertainty and provide safe defaults when data are insufficient; avoid false precision [84].
- **Trust and governance by design:** Transparent data-use statements, role-based access, and auditable logs; essential for sustained participation and benchmarking credibility [54,55].

Several limitations recur across model classes and have direct practical implications. AI/ML models are limited by the quality of ground-truth labels, dataset shift, and reduced interpretability when deployed beyond the conditions represented in the training data; in practice, this means they should usually support prioritization and early warning rather than operate as unchecked automation [84,94]. Optimization models can appear precise yet remain fragile when constraints, objective weights, or stakeholder preferences are incompletely specified; practically, candidate solutions should be treated as decision support only after feasibility review and sensitivity testing [52,53]. Simulation and digital-twin approaches are limited by uncertain parameters, calibration burden, and system-boundary assumptions, so their outputs are often most defensible for scenario comparison and learning rather than point prediction of exact farm outcomes [50,51,101]. Across all classes, the practical implication is the same: credible deployment depends on transparent assumptions, ongoing monitoring, and human interpretation within farm workflows rather than reliance on model outputs alone [54,83].

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 [50,83,84]. 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 → analytics → decision engines → deployment—rather than isolated models or one-off reports (Figure 2a). Figure 2a expresses this idea as a maturity model, showing that progress is not defined by algorithmic sophistication alone but by the extent to which systems become interoperable, validated, monitored, and decision-relevant. Figure 2b extends that logic into an implementation pathway by organizing the review’s recom-

recommendations into staged milestones, from establishing data readiness and governance, to validating analytics in workflows, to scaling optimization, simulation, and digital-twin capabilities. Together, the two panels translate the manuscript's main contribution from a conceptual synthesis into a practical roadmap for building decision-intelligence systems that are credible, usable, and scalable in sustainable dairy production [50,54,55,84].

The roadmap is intended to make the sustainability contribution of the review operational. Rather than treating greenhouse gas emissions, nutrient losses, profitability, animal outcomes, and resilience as separate endpoints, the framework encourages researchers and developers to connect these outcomes to the data streams, model assumptions, validation practices, and decision points that determine whether sustainability recommendations can be implemented on farms. In this sense, the roadmap provides a practical bridge between sustainability assessment and sustainability action in dairy systems.

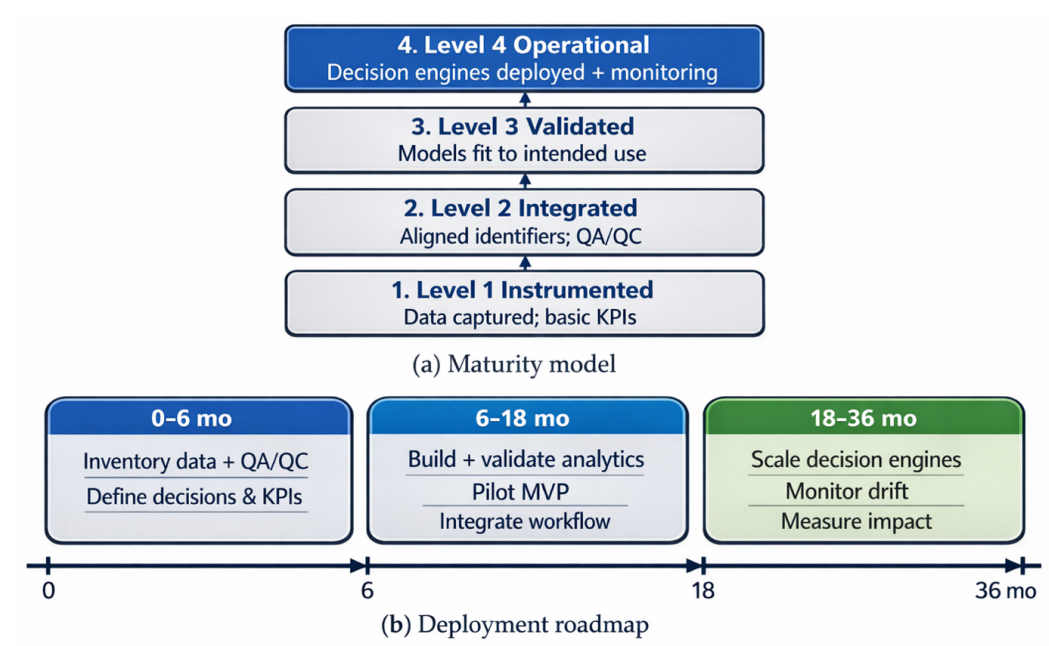


Figure 2. Maturity model (a) and deployment roadmap (b): from data readiness to validated decision engines at farm and sector scale.

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 [22,54,55]. 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 [89].

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 [38,83,84]. Pilot deployments should evaluate deci-

sion 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 [51].

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 [50,52,53]. 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 [50]. Expand interoperability using shared definitions and secure APIs, operationalize reproducibility (preprocessing and model version control, change logs, rollback), and scale to multi-farm cohorts with privacy-preserving benchmarking and consistent metric definitions [54,55,84]. As systems mature, adopt digital-twin practices where appropriate: continuous updating, periodic recalibration, and repeated scenario evaluation as farm conditions evolve [50].

6.2. Implementation Checklist

To translate the roadmap into practice, Table 2 presents an implementation checklist for scalable decision-intelligence systems that support sustainable dairy farm management. The checklist is organized by pipeline layer and links technical requirements, governance needs, validation practices, lifecycle monitoring, user experience, and reporting to worked dairy examples. Each checklist item is paired with worked dairy examples and corresponding references in the final column to illustrate what “good” looks like in practice, support traceability from evidence to implementation, and help teams build decision-intelligence systems that are both environmentally credible and operationally usable on farms.

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. The central contribution of this review is an integrated, deployment-oriented synthesis of that challenge through a single data-to-decision pipeline for sustainable dairy systems. Its novelty lies in bringing together five elements that are often treated separately in the literature—data foundations and interoperability, governance and trust, analytics, decision engines, and deployment—and showing that their value depends on how well they function as one operational workflow. 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.

The value of this synthesis for other researchers is not limited to the specific worked examples included in this review. Rather, the contribution is the transferable structure used to organize digital dairy sustainability research into connected pipeline layers. Researchers can use this framework to classify their own work by decision layer, identify missing components between data collection and farm-level action, compare validation and reporting practices, and design studies that connect sustainability metrics to operational decisions. The implementation checklist and roadmap also provide practical guidance for developing future decision-support systems that are transparent, reproducible, and adaptable to other datasets, regions, and livestock production contexts.

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.d.A.F. and Y.G.; methodology, V.E.C.; validation, V.E.C., E.N.d.A.F. and Y.G.; formal analysis, V.E.C.; investigation, V.E.C., E.N.d.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.d.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:

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

Appendix A.1

Table A1. External peer-reviewed literature foundation for the data-to-decision pipeline, organized by pipeline layer and methodological focus. References are cited throughout the main text.

Reference	Pipeline Layer(s)	Methodological Focus	Relevance to Dairy Synthesis
[7]	Cross-cutting/context	General introduction to precision livestock farming	Foundational PLF reference
[11]	Cross-cutting/context	Scientometric review of PLF (1973–2023)	Establishes global PLF research landscape
[15]	Data foundations; Cross-cutting	Smart farming → Agriculture 5.0 review	Frames transition to integrated digital pipelines
[12]	Data foundations; Analytics	Digital livestock: biometric sensing, big data, blockchain	Anchors PLF/sensors discussion
[8]	Data foundations; Cross-cutting	Big Data in Smart Farming review	Foundational smart-farming reference
[29]	Analytics (AI/ML)	ML and data mining in precision animal agriculture	Foundational ML-in-livestock reference
[33]	Analytics (AI/ML)	Systematic literature review of ML for dairy management	Primary external benchmark for ML in dairy
[34]	Analytics (AI/ML)	Systematic review of deep learning for precision cattle farming	Computer-vision and DL applications in cattle
[35]	Analytics (AI/ML)	Deep-learning algorithms for computer vision in livestock	Object detection for livestock phenotyping
[37]	Analytics (AI/ML)	Scoping review of AI and analytics in dairy (151 papers)	Most recent external AI-in-dairy synthesis
[25]	Data foundations; Governance	FAIR Guiding Principles	Foundational data-stewardship reference
[26]	Data foundations; Governance	FAIR principles for agri-food data ecosystems	Operationalizing FAIR in dairy data
[27]	Data foundations; Sustainability	FAIR data ↔ sustainable agricultural performance	Links FAIR to sustainability outcomes
[28]	Governance	Farmer data privacy and confidentiality	Privacy/encryption/de-identification practice
[19]	Data foundations; Governance	Interoperable data standards in PLF	Canonical interoperability reference
[44]	Decision engines (sim + opt)	Decision support for beef and dairy (110 papers 2016–2022)	Anchor for whole-farm modeling literature
[45]	Decision engines (sim)	ASAS–NANP symposium on mathematical modeling	Five-step modeling pipeline
[47]	Decision engines (sim + opt)	Mathematical methods for dairy cow replacements	External benchmark for herd modeling
[46]	Decision engines (sim)	Bioeconomic + GHG + feed-food modeling (pasture-based)	External integrated whole-farm model
[48]	Decision engines (opt)	LP-based dairy ration-optimization web tool	External operational LP example
[49]	Decision engines (opt)	Review of classical/metaheuristic feed optimization	Cross-species optimization review
[40]	Decision engines (digital twin)	Conceptual framework for digital twins in smart farming	Foundational DT framework w/ dairy use case
[41]	Decision engines (digital twin)	Empirical evidence on DT value + adoption roadmap	Operational DT evidence base
[42]	Decision engines (digital twin)	State-of-the-art DT review across agriculture	DT typology for livestock + supply chain
[43]	Decision engines (digital twin)	Livestock-specific DT review	Most recent livestock-DT synthesis
[68]	Deployment (validation)	Challenges in deploying ML: case-studies survey	ML-deployment methodological reference
[69]	Deployment (drift)	Data drift in medical ML: implications and remedies	Transferable drift framework
[70]	Deployment (validation)	Data leakage and reproducibility in ML-based science	Validation/reproducibility standard

Table A1. Cont.

Reference	Pipeline Layer(s)	Methodological Focus	Relevance to Dairy Synthesis
[71]	Deployment (drift monitoring)	Taxonomy of performance-aware drift detectors	MLOps drift-monitoring reference
[56]	Deployment (adoption)	Social science of digital agriculture (canonical review)	Adoption/socio-ethical reference
[57]	Deployment (adoption)	Responsible innovation for Agriculture 4.0	Responsible-innovation framing
[58]	Deployment (adoption)	Responsible research and innovation in smart farming	Smart-dairy case (New Zealand)
[59]	Deployment (adoption)	Underexamined risks and concerns in PLF	Ethics/risks of PLF adoption
[9]	Cross-cutting/context	Foundational critique of PLF (engineer/animal/farmer views)	Early framing of PLF promise and limits
[10]	Data foundations; Analytics	Real-time sensors for animal well-being and production	Sensor-to-decision data context
[14]	Data foundations; Analytics	Sensors, big data, and ML in modern animal farming	anchors sensor/big-data discussion
[64]	Cross-cutting; Sustainability	Environmental impact of livestock farming and PLF as mitigation	Links PLF to environmental sustainability
[65]	Cross-cutting; Sustainability	PLF as environmental/economic/social compromise in dairy	Dairy-specific PLF sustainability review
[36]	Analytics (AI/ML)	Systematic review of ML in precision livestock farming	External ML-in-PLF benchmark
[30]	Analytics (AI/ML)	Review of ML across agricultural production	Foundational ML-in-agriculture reference
[31]	Analytics (AI/ML)	Survey of deep learning in agriculture	Foundational DL-in-agriculture reference
[32]	Analytics (AI/ML)	Comprehensive updated review of ML in agriculture	Recent ML synthesis including livestock
[24]	Data foundations; Governance	Reference architecture for farm software ecosystems	Interoperability across multi-vendor systems
[61]	Governance; Deployment (adoption)	Trust, transparency, and benefit-sharing in smart farming	Data-trust and adoption reference
[62]	Governance	Ethics of smart farming (ownership, power, society)	Data-ethics reference
[63]	Deployment (adoption)	Review of digitalisation of agricultural advice networks	Advisory/extension adoption reference
[60]	Deployment (adoption)	Responsible innovation in smart farming (Agriculture 4.0)	Responsible-innovation framing
[75]	Deployment (adoption)	Cross-regional EU study of precision-ag adoption	Empirical adoption-driver evidence
[76]	Deployment (adoption); Cross-cutting	PLF technologies and farm animal welfare	Welfare dimension of PLF adoption
[77]	Deployment (value of information)	Conceptual framework for value of information in PLF	Links data to decision value
[78]	Analytics; Deployment (validation)	Systematic review of validated dairy welfare sensors	Sensor-validation evidence base
[66]	Decision engines (opt + sim)	Multi-objective optimization/design of farming systems (FarmDESIGN)	External multi-objective whole-farm model
[67]	Decision engines (digital twin)	Digital-twin paradigm integrated with ML for agriculture	Recent DT framework reference
[2]	Cross-cutting; Sustainability	Enteric methane mitigation opportunities in dairy	Quantifies GHG mitigation potential
[3]	Decision engines (simulation); Sustainability	Whole-farm modeling of dairy GHG emissions	Whole-farm GHG accounting reference
[4]	Sustainability (context)	Rumen methanogenesis and CH ₄ mitigation review	Biological basis for mitigation strategies
[72]	Deployment (validation/MLOps)	Software-engineering workflow for ML systems	ML lifecycle/engineering reference
[73]	Deployment (validation/MLOps)	ML production-readiness test rubric	Production-readiness checklist reference
[74]	Deployment (MLOps)	Hidden technical debt in ML systems	Motivates lifecycle/maintenance discipline

Appendix A.2

Table A2. Illustrative dairy worked examples from one integrated research program, mapped to the data-to-decision pipeline. Each entry shows the pipeline layers addressed, data inputs, method type, decision output, and validation/deployment evidence. The external peer-reviewed literature foundation is presented in Table A1.

Paper	Pipeline Layer(s)	Data Inputs	Method Type	Decision Output	Validation/Deployment Evidence
[101]	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)
[55]	Data foundations and 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)
[95]	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
[106]	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)
[54]	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)
[109]	Deployment (adoption and 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
[104]	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
[83]	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)
[84]	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

Table A2. Cont.

Paper	Pipeline Layer(s)	Data Inputs	Method Type	Decision Output	Validation/Deployment Evidence
[97]	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
[23]	Data foundations and 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
[5]	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
[6]	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
[85]	Data foundations and 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
[51]	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
[88]	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
[90]	Data foundations and 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
[94]	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
[89]	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)
[96]	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

Table A2. Cont.

Paper	Pipeline Layer(s)	Data Inputs	Method Type	Decision Output	Validation/Deployment Evidence
[52]	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
[53]	Decision engines (optimization)	Cow/group nutrient requirement; group number; grouping criteria; feed ingredient composition; prices/costs	Duo-objective linear optimization	Optimal nutritional grouping strategy and ration formulations for separate groups that consider both GHG emissions and feed cost	Case-study optimization results
[39]	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
[108]	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
[38]	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
[105]	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
[103]	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
[99]	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
[107]	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
[100]	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
[93]	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

Table A2. Cont.

Paper	Pipeline Layer(s)	Data Inputs	Method Type	Decision Output	Validation/Deployment Evidence
[86]	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
[87]	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
[91]	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
[50]	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
[22]	Data foundations and 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
[98]	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

- 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*, 1st ed.; Cambridge University Press: Cambridge, UK, 2023. [\[CrossRef\]](#)
- Knapp, J.R.; Laur, G.L.; Vadas, P.A.; Weiss, W.P.; Tricarico, J.M. Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. *J. Dairy Sci.* **2014**, *97*, 3231–3261. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rotz, C.A. Modeling greenhouse gas emissions from dairy farms. *J. Dairy Sci.* **2018**, *101*, 6675–6690. [\[CrossRef\]](#) [\[PubMed\]](#)
- Beauchemin, K.A.; Ungerfeld, E.M.; Eckard, R.J.; Wang, M. Review: Fifty years of research on rumen methanogenesis: Lessons learned and future challenges for mitigation. *Animal* **2020**, *14*, s2–s16. [\[CrossRef\]](#) [\[PubMed\]](#)
- 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. *Comput. Electron. Agric.* **2005**, *49*, 286–308. [\[CrossRef\]](#)
- Cabrera, V.E.; Hildebrand, P.E.; Jones, J.W.; Letson, D.; De Vries, A. An integrated North Florida dairy farm model to reduce environmental impacts under seasonal climate variability. *Agric. Ecosyst. Environ.* **2006**, *113*, 82–97. [\[CrossRef\]](#)
- Berckmans, D. General introduction to precision livestock farming. *Anim. Front.* **2017**, *7*, 6–11. [\[CrossRef\]](#)
- Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big Data in Smart Farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [\[CrossRef\]](#)
- Wathes, C.M.; Kristensen, H.H.; Aerts, J.M.; Berckmans, D. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? *Comput. Electron. Agric.* **2008**, *64*, 2–10. [\[CrossRef\]](#)
- Halachmi, I.; Guarino, M.; Bewley, J.; Pastell, M. Smart Animal Agriculture: Application of Real-Time Sensors to Improve Animal Well-Being and Production. *Annu. Rev. Anim. Biosci.* **2019**, *7*, 403–425. [\[CrossRef\]](#) [\[PubMed\]](#)

11. Jiang, B.; Tang, W.; Cui, L.; Deng, X. Precision Livestock Farming Research: A Global Scientometric Review. *Animals* **2023**, *13*, 2096. [[CrossRef](#)] [[PubMed](#)]
12. Neethirajan, S.; Kemp, B. Digital Livestock Farming. *Sens. Bio-Sens. Res.* **2021**, *32*, 100408. [[CrossRef](#)]
13. Morrone, S.; Dimauro, C.; Gambella, F.; Cappai, M.G. Industry 4.0 and Precision Livestock Farming (PLF): An up to Date Overview across Animal Productions. *Sensors* **2022**, *22*, 4319. [[CrossRef](#)] [[PubMed](#)]
14. Neethirajan, S. The role of sensors, big data and machine learning in modern animal farming. *Sens. Bio-Sens. Res.* **2020**, *29*, 100367. [[CrossRef](#)]
15. Saiz-Rubio, V.; Rovira-Más, F. From Smart Farming towards Agriculture 5.0: A Review on Crop Data Management. *Agronomy* **2020**, *10*, 207. [[CrossRef](#)]
16. da Silveira, F.; Lermen, F.H.; Amaral, F.G. An overview of agriculture 4.0 development: Systematic review of descriptions, technologies, barriers, advantages, and disadvantages. *Comput. Electron. Agric.* **2021**, *189*, 106405. [[CrossRef](#)]
17. Bissadu, K.D.; Sonko, S.; Hossain, G. Society 5.0 enabled agriculture: Drivers, enabling technologies, architectures, opportunities, and challenges. *Inf. Process. Agric.* **2025**, *12*, 112–124. [[CrossRef](#)]
18. McFadden, J.; Casalini, F.; Griffin, T.; Antón, J. *The Digitalisation of Agriculture: A Literature Review and Emerging Policy Issues*; Technical Report 176; OECD Publishing: Paris, France, 2022. [[CrossRef](#)]
19. Bahlo, C.; Dahlhaus, P.; Thompson, H.; Trotter, M.G. The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review. *Comput. Electron. Agric.* **2019**, *156*, 459–466. [[CrossRef](#)]
20. Kaur, U.; Malacco, V.M.R.; Bai, H.; Price, T.P.; Datta, A.; Xin, L.; Sen, S.; Nawrocki, R.A.; Chiu, G.; Sundaram, S.; et al. Invited review: Integration of technologies and systems for precision animal agriculture—A case study on precision dairy farming. *J. Anim. Sci.* **2023**, *101*, skad206. [[CrossRef](#)] [[PubMed](#)]
21. Brito, L.F.; Heringstad, B.; Klaas, I.C.; Schodl, K.; Cabrera, V.E.; Stygar, A.; Iwersen, M.; Haskell, M.J.; Stock, K.F.; Gengler, N.; et al. Invited review: Using data from sensors and other precision farming technologies to enhance the sustainability of dairy cattle breeding programs. *J. Dairy Sci.* **2025**, *108*, 10447–10474. [[CrossRef](#)] [[PubMed](#)]
22. 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. *Livest. Sci.* **2021**, *250*, 104602. [[CrossRef](#)]
23. 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. [[CrossRef](#)] [[PubMed](#)]
24. Kruize, J.W.; Wolfert, S.; Scholten, H.; Verdouw, C.N.; Kassahun, A.; Beulens, A.J.M. A reference architecture for Farm Software Ecosystems. *Comput. Electron. Agric.* **2016**, *125*, 12–28. [[CrossRef](#)]
25. 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. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)] [[PubMed](#)]
26. Top, J.; Janssen, S.; Boogaard, H.; Knapen, R.; Şimşek-Şenel, G. Cultivating FAIR principles for agri-food data. *Comput. Electron. Agric.* **2022**, *196*, 106909. [[CrossRef](#)]
27. Ali, B.; Dahlhaus, P. The Role of FAIR Data towards Sustainable Agricultural Performance: A Systematic Literature Review. *Agriculture* **2022**, *12*, 309. [[CrossRef](#)]
28. Kaur, J.; Hazrati Fard, S.M.; Amiri-Zarandi, M.; Dara, R. Protecting farmers' data privacy and confidentiality: Recommendations and considerations. *Front. Sustain. Food Syst.* **2022**, *6*, 903230. [[CrossRef](#)]
29. 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 agriculture. *J. Anim. Sci.* **2018**, *96*, 1540–1550. [[CrossRef](#)] [[PubMed](#)]
30. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [[CrossRef](#)] [[PubMed](#)]
31. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
32. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors* **2021**, *21*, 3758. [[CrossRef](#)] [[PubMed](#)]
33. Slob, N.; Catal, C.; Kassahun, A. Application of machine learning to improve dairy farm management: A systematic literature review. *Prev. Vet. Med.* **2021**, *187*, 105237. [[CrossRef](#)] [[PubMed](#)]
34. Mahmud, M.S.; Zahid, A.; Das, A.K.; Muzammil, M.; Khan, M.U. A systematic literature review on deep learning applications for precision cattle farming. *Comput. Electron. Agric.* **2021**, *187*, 106313. [[CrossRef](#)]
35. Oliveira, D.A.B.; Pereira, L.G.R.; Bresolin, T.; Ferreira, R.E.P.; Dorea, J.R.R. A review of deep learning algorithms for computer vision systems in livestock. *Livest. Sci.* **2021**, *253*, 104700. [[CrossRef](#)]
36. García, R.; Aguilar, J.; Toro, M.; Pinto, A.; Rodríguez, P. A systematic literature review on the use of machine learning in precision livestock farming. *Comput. Electron. Agric.* **2020**, *179*, 105826. [[CrossRef](#)]

37. Palma, O.; Plà-Aragónés, L.M.; Mac Cawley, A.; Albornoz, V.M. AI and Data Analytics in the Dairy Farms: A Scoping Review. *Animals* **2025**, *15*, 1291. [\[CrossRef\]](#) [\[PubMed\]](#)
38. 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. *J. Dairy Sci.* **2025**, *108*, 9696–9711. [\[CrossRef\]](#) [\[PubMed\]](#)
39. 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 Agric. Technol.* **2025**, *12*, 101097. [\[CrossRef\]](#)
40. Verdouw, C.; Tekinerdogan, B.; Beulens, A.; Wolfert, S. Digital twins in smart farming. *Agric. Syst.* **2021**, *189*, 103046. [\[CrossRef\]](#)
41. Pylianidis, C.; Osinga, S.; Athanasiadis, I.N. Introducing digital twins to agriculture. *Comput. Electron. Agric.* **2021**, *184*, 105942. [\[CrossRef\]](#)
42. Purcell, W.; Neubauer, T. Digital Twins in Agriculture: A State-of-the-art review. *Smart Agric. Technol.* **2023**, *3*, 100094. [\[CrossRef\]](#)
43. Arulmozhi, E.; Deb, N.C.; Tamrakar, N.; Kang, D.Y.; Kang, M.Y.; Kook, J.; Basak, J.K.; Kim, H.T. From Reality to Virtuality: Revolutionizing Livestock Farming Through Digital Twins. *Agriculture* **2024**, *14*, 2231. [\[CrossRef\]](#)
44. Bang, R.N.; Guajardo, M.; Hansen, B.G. Recent advances in decision support for beef and dairy farming: Modeling approaches and opportunities. *Int. Trans. Oper. Res.* **2023**, *30*, 2807–2839. [\[CrossRef\]](#)
45. Menendez, H.M.; Brennan, J.R.; Gaillard, C.; Ehlert, K.; Quintana, J.; Tedeschi, L.O. ASAS-NANP Symposium: Mathematical Modeling in Animal Nutrition: Opportunities and challenges of confined and extensive precision livestock production. *J. Anim. Sci.* **2022**, *100*, skac160. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Kearney, M.; O’Riordan, E.G.; McGee, M.; Breen, J.; Crosson, P. Farm-level modelling of bioeconomic, greenhouse gas emissions and feed-food performance of pasture-based dairy-beef systems. *Agric. Syst.* **2022**, *203*, 103505. [\[CrossRef\]](#)
47. Palma, O.; Plà-Aragónés, L.M.; Mac Cawley, A.; Albornoz, V.M. Mathematical Methods Applied to the Problem of Dairy Cow Replacements: A Scoping Review. *Animals* **2025**, *15*, 970. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Mijić, D.; Vico, G.; Popović, B.; Popović, N.; Ljubojević, M.; Savić, M. OPTIMILK: A Web-Based Tool for Least-Cost Dairy Ration Optimization Using Linear Programming. *Agriculture* **2024**, *14*, 1580. [\[CrossRef\]](#)
49. Das, R.; Patty, S.; Das, D.; Das, K.N. Recent advancements of classical and metaheuristic techniques in ration feed optimization for livestock, poultry, and aquaculture: A review. *Int. J. Syst. Assur. Eng. Manag.* **2025**, *16*, 2646–2669. [\[CrossRef\]](#)
50. 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 Commun.* **2025**, *6*, S15–S22. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Da Silva, T.E.; Cabrera, V.E. The DairyPrint model: A decision support model to help dairy farmers and other stakeholders toward improved sustainability. *J. Dairy Sci.* **2024**, *107*, 10998–11015. [\[CrossRef\]](#) [\[PubMed\]](#)
52. 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. *J. Clean. Prod.* **2025**, *511*, 145636. [\[CrossRef\]](#)
53. Gong, Y.; Hu, H.; Reed, K.; Noronha de Andrade Freitas, E.; Cabrera, V. Can nutritional grouping lower feed costs and enteric methane when diets are optimized with modest ingredient changes? *JDS Commun.* **2026**, *7*, 321–326. [\[CrossRef\]](#) [\[PubMed\]](#)
54. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
55. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
56. 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 Wagening. J. Life Sci.* **2019**, *90–91*, 1–16. [\[CrossRef\]](#)
57. Klerkx, L.; Rose, D. Dealing with the game-changing technologies of Agriculture 4.0: How do we manage diversity and responsibility in food system transition pathways? *Glob. Food Secur.* **2020**, *24*, 100347. [\[CrossRef\]](#)
58. Eastwood, C.; Klerkx, L.; Ayre, M.; Dela Rue, B. Managing Socio-Ethical Challenges in the Development of Smart Farming: From a Fragmented to a Comprehensive Approach for Responsible Research and Innovation. *J. Agric. Environ. Ethics* **2019**, *32*, 741–768. [\[CrossRef\]](#)
59. Werkheiser, I. Technology and responsibility: A discussion of underexamined risks and concerns in Precision Livestock Farming. *Anim. Front.* **2020**, *10*, 51–57. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Rose, D.C.; Chilvers, J. Agriculture 4.0: Broadening Responsible Innovation in an Era of Smart Farming. *Front. Sustain. Food Syst.* **2018**, *2*, 87. [\[CrossRef\]](#)
61. Jakku, E.; Taylor, B.; Fleming, A.; Mason, C.; Fielke, S.; Sounness, C.; Thorburn, P. “If they don’t tell us what they do with it, why would we trust them?” Trust, transparency and benefit-sharing in Smart Farming. *NJAS Wagening. J. Life Sci.* **2019**, *90–91*, 100285. [\[CrossRef\]](#)

62. van der Burg, S.; Bogaardt, M.J.; Wolfert, S. Ethics of smart farming: Current questions and directions for responsible innovation towards the future. *NJAS Wagening. J. Life Sci.* **2019**, *90–91*, 100289. [\[CrossRef\]](#)
63. Fielke, S.; Taylor, B.; Jakku, E. Digitalisation of agricultural knowledge and advice networks: A state-of-the-art review. *Agric. Syst.* **2020**, *180*, 102763. [\[CrossRef\]](#)
64. Tullo, E.; Finzi, A.; Guarino, M. Review: Environmental impact of livestock farming and Precision Livestock Farming as a mitigation strategy. *Sci. Total Environ.* **2019**, *650*, 2751–2760. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Lovarelli, D.; Bacenetti, J.; Guarino, M. A review on dairy cattle farming: Is precision livestock farming the compromise for an environmental, economic and social sustainable production? *J. Clean. Prod.* **2020**, *262*, 121409. [\[CrossRef\]](#)
66. Groot, J.C.J.; Oomen, G.J.M.; Rossing, W.A.H. Multi-objective optimization and design of farming systems. *Agric. Syst.* **2012**, *110*, 63–77. [\[CrossRef\]](#)
67. Nasirahmadi, A.; Hensel, O. Toward the Next Generation of Digitalization in Agriculture Based on Digital Twin Paradigm. *Sensors* **2022**, *22*, 498. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Paleyes, A.; Urma, R.G.; Lawrence, N.D. Challenges in Deploying Machine Learning: A Survey of Case Studies. *ACM Comput. Surv.* **2022**, *55*, 1–29. [\[CrossRef\]](#)
69. Sahiner, B.; Chen, W.; Samala, R.K.; Petrick, N. Data drift in medical machine learning: Implications and potential remedies. *Br. J. Radiol.* **2023**, *96*, 20220878. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Kapoor, S.; Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* **2023**, *4*, 100804. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Bayram, F.; Ahmed, B.S.; Kassler, A. From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowl.-Based Syst.* **2022**, *245*, 108632. [\[CrossRef\]](#)
72. Amershi, S.; Begel, A.; Bird, C.; DeLine, R.; Gall, H.; Kamar, E.; Nagappan, N.; Nushi, B.; Zimmermann, T. Software Engineering for Machine Learning: A Case Study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*; IEEE: New York, NY, USA, 2019; pp. 291–300. [\[CrossRef\]](#)
73. Breck, E.; Cai, S.; Nielsen, E.; Salib, M.; Sculley, D. The ML Test Score: A rubric for ML production readiness and technical debt reduction. In *2017 IEEE International Conference on Big Data (Big Data)*; IEEE: New York, NY, USA, 2017; pp. 1123–1132. [\[CrossRef\]](#)
74. Sculley, D.; Holt, G.; Golovin, D.; Davydov, E.; Phillips, T.; Ebner, D.; Chaudhary, V.; Young, M.; Crespo, J.F.; Dennison, D. Hidden Technical Debt in Machine Learning Systems. In *Proceedings of the Advances in Neural Information Processing Systems 28*; Curran Associates, Inc.: New York, NY, USA, 2015; pp. 2503–2511.
75. Barnes, A.P.; Soto, I.; Eory, V.; Beck, B.; Balaoutis, A.; Sánchez, B.; Vangeyte, J.; Fountas, S.; van der Wal, T.; Gómez-Barbero, M. Exploring the adoption of precision agricultural technologies: A cross regional study of EU farmers. *Land Use Policy* **2019**, *80*, 163–174. [\[CrossRef\]](#)
76. Schillings, J.; Bennett, R.; Rose, D.C. Exploring the Potential of Precision Livestock Farming Technologies to Help Address Farm Animal Welfare. *Front. Anim. Sci.* **2021**, *2*, 639678. [\[CrossRef\]](#)
77. Rojo-Gimeno, C.; van der Voort, M.; Niemi, J.K.; Lauwers, L.; Kristensen, A.R.; Wauters, E. Assessment of the value of information of precision livestock farming: A conceptual framework. *NJAS Wagening. J. Life Sci.* **2019**, *90–91*, 100311. [\[CrossRef\]](#)
78. Stygar, A.H.; Gómez, Y.; Berteselli, G.V.; Dalla Costa, E.; Canali, E.; Niemi, J.K.; Llonch, P.; Pastell, M. A Systematic Review on Commercially Available and Validated Sensor Technologies for Welfare Assessment of Dairy Cattle. *Front. Vet. Sci.* **2021**, *8*, 634338. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Grant, M.J.; Booth, A. A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Inf. Libr. J.* **2009**, *26*, 91–108. [\[CrossRef\]](#) [\[PubMed\]](#)
80. 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. *Ann. Intern. Med.* **2018**, *169*, 467–473. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Baethge, C.; Goldbeck-Wood, S.; Mertens, S. SANRA—a scale for the quality assessment of narrative review articles. *Res. Integr. Peer Rev.* **2019**, *4*, 5. [\[CrossRef\]](#) [\[PubMed\]](#)
82. 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**, *372*, n71. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Cabrera, V. Invited review: Helping dairy farmers to improve economic performance utilizing data-driving decision support tools. *Animal* **2018**, *12*, 134–144. [\[CrossRef\]](#) [\[PubMed\]](#)
84. 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. *J. Dairy Sci.* **2020**, *103*, 3856–3866. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Cabrera, V.E.; Fadul-Pacheco, L. Future of dairy farming from the Dairy Brain perspective: Data integration, analytics, and applications. *Int. Dairy J.* **2021**, *121*, 105069. [\[CrossRef\]](#)

86. 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. *Anim. Front.* **2024**, *14*, 42–53. [[CrossRef](#)] [[PubMed](#)]
87. Niu, M.; Guo, C.; Cabrera, V.E. Artificial intelligence for animal science: From applications to integrated knowledge systems. *Anim. Front.* **2025**, *16*, 26–33. [[CrossRef](#)] [[PubMed](#)]
88. 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. *J. Dairy Sci.* **2014**, *97*, 5904–5917. [[CrossRef](#)] [[PubMed](#)]
89. Freitas, E.; Cabrera, V. Dairy Victory Platform: A novel benchmarking platform to empower economic decisions on dairy farms. *JDS Commun.* **2025**, *6*, 69–73. [[CrossRef](#)] [[PubMed](#)]
90. Fadul-Pacheco, L.; Wangen, S.R.; Da Silva, T.E.; Cabrera, V.E. Addressing Data Bottlenecks in the Dairy Farm Industry. *Animals* **2022**, *12*, 721. [[CrossRef](#)] [[PubMed](#)]
91. 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 Agric. Technol.* **2024**, *7*, 100417. [[CrossRef](#)]
92. Wiseman, L.; Sanderson, J.; Zhang, A.; Jakku, E. Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. *NJAS Wagening. J. Life Sci.* **2019**, *90–91*, 100301. [[CrossRef](#)]
93. Liang, D.; Cabrera, V. Optimizing productivity, herd structure, environmental performance, and profitability of dairy cattle herds. *J. Dairy Sci.* **2015**, *98*, 2812–2823. [[CrossRef](#)] [[PubMed](#)]
94. Fadul-Pacheco, L.; Delgado, H.; Cabrera, V.E. Exploring machine learning algorithms for early prediction of clinical mastitis. *Int. Dairy J.* **2021**, *119*, 105051. [[CrossRef](#)]
95. Barrientos-Blanco, J.A.; White, H.; Shaver, R.D.; Cabrera, V.E. Improving nutritional accuracy and economics through a multiple ration-grouping strategy. *J. Dairy Sci.* **2020**, *103*, 3774–3785. [[CrossRef](#)] [[PubMed](#)]
96. 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. *J. Dairy Sci.* **2025**, *108*, 9681–9695. [[CrossRef](#)] [[PubMed](#)]
97. Cabrera, V.E. Artificial intelligence applied to dairy science: Insights from the Dairy Brain Initiative. *Anim. Front.* **2024**, *14*, 60–63. [[CrossRef](#)] [[PubMed](#)]
98. 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. *J. Dairy Sci.* **2022**, *105*, 6739–6748. [[CrossRef](#)] [[PubMed](#)]
99. Li, M.; Rosa, G.; Reed, K.; Cabrera, V. Investigating the effect of temporal, geographic, and management factors on US Holstein lactation curve parameters. *J. Dairy Sci.* **2022**, *105*, 7525–7538. [[CrossRef](#)] [[PubMed](#)]
100. Li, M.; Reed, K.; Cabrera, V. A time series analysis of milk productivity in US dairy states. *J. Dairy Sci.* **2023**, *106*, 6232–6248. [[CrossRef](#)] [[PubMed](#)]
101. 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. *J. Dairy Sci.* **2015**, *98*, 1571–1592. [[CrossRef](#)] [[PubMed](#)]
102. Fadul-Pacheco, L.; Delgado, H.; Cabrera, V.E. A Preliminary Investigation of Social Network Analysis Applied to Dairy Cow Behavior in Automatic Milking System Environments. *Animals* **2021**, *11*, 1229. [[CrossRef](#)] [[PubMed](#)]
103. 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. *Anim. Front.* **2019**, *9*, 25–32. [[CrossRef](#)] [[PubMed](#)]
104. Cabrera, V. A large Markovian linear program to optimize replacement policies and dairy herd net income for diets and nitrogen excretion. *J. Dairy Sci.* **2010**, *93*, 394–406. [[CrossRef](#)] [[PubMed](#)]
105. Kalantari, A.; Armentano, L.; Shaver, R.; Cabrera, V. Economic impact of nutritional grouping in dairy herds. *J. Dairy Sci.* **2016**, *99*, 1672–1692. [[CrossRef](#)] [[PubMed](#)]
106. Barrientos-Blanco, J.A.; White, H.; Shaver, R.D.; Cabrera, V.E. Graduate Student Literature Review: Considerations for nutritional grouping in dairy farms. *J. Dairy Sci.* **2022**, *105*, 2708–2717. [[CrossRef](#)] [[PubMed](#)]
107. 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. *J. Dairy Sci.* **2023**, *106*, 3246–3267. [[CrossRef](#)] [[PubMed](#)]
108. Hansen, T.L.; Li, M.; Li, J.; Vankervhove, 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. [[CrossRef](#)] [[PubMed](#)]
109. Cabrera, V.E.; Breuer, N.E.; Hildebrand, P.E. North Florida Dairy Farmer Perceptions Toward the Use of Seasonal Climate Forecast Technology. *Clim. Change* **2006**, *78*, 479–491. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.