Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

From Prediction to Action: A Structured Scoping Review and Framework Synthesis of Integrative AI Decision Systems

Dominika Zak

dzak@fordham.edu

Fordham University https://orcid.org/0009-0008-2224-0730

Research Article

Keywords: artificial intelligence, machine learning, big data analytics, decision systems, healthcare analytics, value-based care, model validation, MLOps, dataset shift, high-stakes AI

Posted Date: May 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9684818/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

From Prediction to Action: A Structured Scoping Review and Framework Synthesis of Integrative AI Decision Systems

Dominika Zak

Fordham University; New York, USA

@Corresponding Author: Dominika Zak

Email: dzak@fordham.edu

ABSTRACT:

Artificial intelligence, machine learning, and big data analytics are increasingly used in settings where predictions influence clinical, financial, operational, and public decisions. In these settings, the central issue is not whether a model can produce a score, but whether the score can be translated into reliable action under real-world constraints. A structured scoping technical review and framework synthesis was conducted in 2026. Searches covered PubMed/MEDLINE, PubMed Central, IEEE Xplore, ACM Digital Library, Google Scholar, and targeted publisher websites for literature published from 2018 through 2026. Search terms combined artificial intelligence, machine learning, clinical decision support, model validation, calibration, dataset shift, fairness, explainability, MLOps, workflow integration, value-based care, fraud detection, traffic prediction, smart infrastructure, predictive maintenance, and operational decision support. Records were screened for deployment relevance, methodological detail, and applicability to high-stakes decision systems. The search and citation process identified 112 candidate records. After duplicate removal, 78 records were screened by title and abstract, 34 full texts were assessed, and 23 studies or guidance documents were included in the final synthesis. The evidence mapped consistently to six linked components of AI deployment: data integration, engineering pipelines, learning, validation, serving and workflow, and value capture. Governance was identified as a cross-cutting requirement because privacy, interpretability, fairness, accountability, documentation, monitoring, and human oversight affected every stage of the system. The review develops the Integrative AI Decision Systems (IADS) framework as the main research output. Healthcare is used as the primary application domain because AI outputs may affect diagnosis, triage, utilization management, care coordination, patient safety, and value-based care performance. Finance, smart infrastructure, and industry are used as comparative sectors because they reveal similar deployment problems involving drift, class imbalance, adversarial behavior, real-time reliability, transferability, and cost-sensitive decisions. AI creates durable value only when predictive performance is connected to validated workflows, monitoring, and accountable decision processes.

Keywords: artificial intelligence, machine learning, big data analytics, decision systems, healthcare analytics, value-based care, model validation, MLOps, dataset shift, high-stakes AI

Introduction

Artificial intelligence (AI), machine learning, and big data analytics have moved beyond the stage of isolated prediction tools. In healthcare and other high-stakes sectors, these technologies are increasingly embedded in software, data pipelines, dashboards, triage processes, operational workflows, and governance routines [1-5]. A model may begin as a statistical object, but once it influences clinical action, resource allocation, fraud review, routing, maintenance, or policy planning, it becomes part of a decision system. That distinction is important because the reliability of the final decision depends on more than algorithmic performance.

Healthcare provides a useful primary setting for studying this shift. AI systems are now discussed in relation to diagnosis, imaging interpretation, population risk stratification, utilization management, patient monitoring, care coordination, and value-based care [1-6]. These applications are technically promising, but they also expose a gap

between retrospective model performance and operational readiness. A model that performs well on a benchmark dataset may still fail after deployment if clinical practice changes, if the patient population differs from the development cohort, if subgroup performance is uneven, or if the output is not understandable to the people expected to act on it [7-11].

The same pattern appears outside healthcare. Financial institutions use machine learning for fraud detection, compliance monitoring, credit assessment, and transaction screening, where drift, class imbalance, adversarial behavior, and regulatory expectations affect deployment [13]. Smart infrastructure systems use mobility, sensor, weather, and geospatial data to support congestion management, resource allocation, and control tasks [14,16]. Industrial systems use telemetry and maintenance logs to predict equipment failure, but useful deployment depends on whether alerts are timely, trusted, and tied to an action that prevents downtime [15]. Across these sectors, predictive accuracy is necessary, but it is not sufficient. This article reports a structured scoping technical review and framework synthesis. The research question was: what technical, operational, and governance requirements recur across high-stakes AI systems when predictive outputs must be translated into action? The objective was not to estimate a pooled effect size. The objective was to derive a practical framework that can be used to evaluate AI systems as deployed decision systems rather than isolated algorithms. The resulting framework is the Integrative AI Decision Systems (IADS) framework.

Methods

This study used a structured scoping technical review design with framework synthesis. Scoping review methods were selected because the research question spans multiple domains, data types, model classes, and outcome definitions rather than a single intervention or effect estimate. The review was informed by PRISMA-ScR reporting principles for scoping reviews, JBI guidance on scoping review conduct, and synthesis-without-meta-analysis principles for transparent narrative synthesis [24-26]. The primary sources were PubMed/MEDLINE, PubMed Central, IEEE Xplore, ACM Digital Library, Google Scholar, and targeted publisher websites. The publication window was 2018 through 2026, with older foundational literature retained only when it directly supported high-stakes interpretability or clinical AI deployment. Searches were limited to English-language sources because the objective was to synthesize technical and reporting concepts with sufficient methodological detail for direct comparison.

The search strategy combined terms for AI technology, deployment context, validation, governance, and sector-specific decision use. Healthcare searches focused on artificial intelligence, machine learning, clinical decision support, model validation, calibration, dataset shift, fairness, explainability, MLOps, workflow integration, and value-based care. Comparative searches used fraud detection, smart infrastructure, traffic prediction, predictive maintenance, operational decision support, concept drift, and cost-sensitive decisions. Targeted citation searching was used to identify reporting guidelines and highly cited deployment-relevant sources not captured by the initial database searches.

Eligibility was defined before full-text review. Articles were included if they addressed AI or machine learning in relation to decision support, deployment, validation, governance, monitoring, workflow integration, or operational use. Sources were also included if they described sector-specific deployment risks relevant to decision-system design, such as drift in fraud detection, real-time reliability in infrastructure, or false alarms in predictive maintenance. Articles were excluded if they focused only on benchmark accuracy without discussing validation, implementation, interpretation, governance, or decision consequences. Non-technical editorials, unsupported commentary, inaccessible full texts, and sources outside the scope of AI decision support were excluded. For each included source, the following extraction fields were recorded: domain, data source, model or system type, decision target, validation concept, operational risk, governance issue, workflow or actionability implication, and relevance to framework development. Extraction was organized around a directed framework matrix. Sources were first coded inductively for recurring deployment requirements and then grouped into six system components: data integration, engineering pipelines, learning, validation, serving and workflow, and value capture. Governance was coded across all components because it was not limited to one point in the AI lifecycle.

The synthesis was qualitative and technical rather than statistical. Meta-analysis was not appropriate because the included sources varied by domain, task, data type, evaluation design, and outcome definition. Instead, the review used framework synthesis to derive a reusable structure for evaluating AI decision systems. The final framework was retained only if each component was supported by recurring evidence across more than one domain or by reporting guidance directly relevant to deployment readiness.

Search block	Example terms	Primary sources	Purpose
Healthcare AI	artificial intelligence, machine learning, clinical decision support, healthcare, medicine	PubMed/MEDLINE, PMC	Identify clinical and health-system AI deployment literature
Validation and monitoring	validation, calibration, external validation, dataset shift, model monitoring, MLOps	PubMed/MEDLINE, PMC, IEEE Xplore, ACM	Identify methods for lifecycle reliability and post-deployment oversight
Fairness and explainability	fairness, bias, subgroup performance, explainability, interpretability, accountability	PubMed/MEDLINE, PMC, publisher sites	Identify equity and interpretability requirements
Value-based care	value-based care, episode attribution, utilization management, care coordination, accountable care	PubMed/MEDLINE, Google Scholar, publisher sites	Connect AI outputs to payment, coordination, and resource decisions
Comparative sectors	fraud detection, smart infrastructure, traffic prediction, predictive maintenance, concept drift	IEEE Xplore, ACM, Google Scholar, publisher sites	Compare deployment risks outside healthcare

Table 1. Search strategy used for the structured scoping technical review.

Category	Included	Excluded
Topic	AI, machine learning, deep learning, big data analytics, clinical decision support, decision systems	General digital health articles without AI or decision-system relevance
Domain	Healthcare, finance, smart infrastructure, industrial operations	Domains without high-stakes operational decisions
Technical focus	Validation, calibration, dataset shift, fairness, workflow integration, MLOps, governance	Accuracy-only studies with no deployment or validation discussion
Article type	Peer-reviewed studies, technical reviews, methodological papers, reporting guidelines	Editorials, opinion pieces, unsupported commentary
Language	English-language sources	Non-English sources
Data relevance	Studies with clear deployment, validation, or decision-support implications	Studies without sufficient methodological detail

Table 2. Inclusion and exclusion criteria used for the structured technical review.

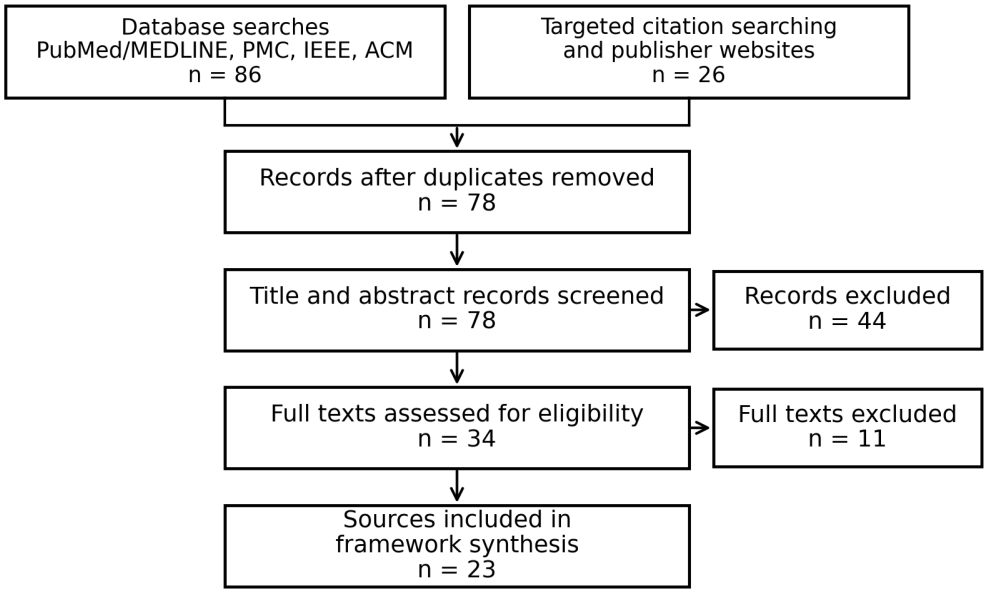

Figure 1. Screening flow used to identify studies and guidance documents for the IADS framework synthesis.

Results

The search and citation process identified 112 candidate records. Eighty-six records were identified through database searches and 26 additional records were identified through targeted citation searching and publisher websites. After removal of 34 duplicates or clearly overlapping records, 78 records remained for title and abstract screening. Forty-four records were excluded at this stage because they were accuracy-only papers, general digital health papers without AI deployment relevance, opinion pieces without methodological detail, or records outside the intended high-stakes decision-system scope. Thirty-four full texts were assessed for eligibility. Eleven full texts were excluded after detailed review. The most common exclusion reasons were lack of deployment or validation detail, absence of a decision-system implication, limited technical transparency, or narrow focus on algorithmic performance without workflow, governance, or monitoring discussion. Twenty-three sources were retained for the final synthesis, including clinical AI reviews, multimodal biomedical AI studies, dataset shift papers, fairness studies, MLOps literature, value-based care sources, cross-sector AI studies, and reporting or evaluation guidelines.

The included evidence did not support a single pooled quantitative estimate because the sources differed substantially in domain, task, model class, dataset, and outcome. The synthesis therefore focused on recurring system requirements. Across the included sources, six requirements appeared repeatedly: the need to integrate heterogeneous data, maintain engineering pipelines, select models appropriate to the decision context, validate performance beyond retrospective discrimination, embed outputs into workflow, and measure whether the system improves decisions or resource use. Healthcare sources contributed most strongly to the validation, fairness, workflow, and governance components. Finance, infrastructure, and industrial sources strengthened the cross-sector interpretation by showing that the same prediction-to-action problem appears when fraud patterns adapt, traffic and mobility behavior changes, or equipment conditions shift over time. Reporting guidelines contributed structure for transparent reporting, risk-of-bias assessment, human-AI interaction, and early clinical evaluation.

Evidence group	Included sources	Main deployment issue	Framework contribution	Representative citations
Clinical and biomedical AI	Reviews and technical studies of AI in medicine, multimodal learning, big data, and clinical impact	Retrospective performance may not translate into clinical usefulness	Supported the need to treat AI as a deployed decision system rather than an isolated model	[1-5,10,23]
Dataset shift and fairness	Papers on dataset shift, real-world generalization, and fair machine learning	Performance may degrade across sites, time periods, or subgroups	Supported validation, monitoring, subgroup evaluation, and governance requirements	[7-9]
MLOps and lifecycle management	Healthcare MLOps and model monitoring literature	Models require versioning, monitoring, retraining, documentation, and ownership	Supported engineering pipelines, serving, monitoring, and lifecycle governance	[11]
Value-based care	Sources on AI, value-based health insurance, and episode-level transparency	Prediction must connect to care coordination, utilization, attribution, and accountability	Supported value capture and actionability as distinct from risk ranking	[6,12]
Cross-sector AI	Fraud detection, traffic prediction, smart infrastructure, and predictive maintenance studies	Operational AI faces drift, class imbalance, false alarms, transferability limits, and real-time constraints	Supported the generalizability of the IADS framework beyond healthcare	[13-16]
Reporting and evaluation guidance	TRIPOD+AI, PROBAST+AI, CONSORT-AI, SPIRIT-AI, DECIDE-AI, PRISMA-ScR, JBI, SWiM	Deployment studies require transparent reporting, applicability assessment, and synthesis structure	Supported reporting discipline, extraction logic, and evaluation of AI as an intervention in context	[17-21,24-26]

Table 3. Evidence map showing how included sources contributed to framework development.

Dataset and Technical System Description

The literature reviewed in this study describes several broad classes of data used in AI decision systems. In healthcare, common sources include electronic health records, claims, laboratory values, medication histories, imaging, pathology data, clinical notes, device signals, and administrative records [1-5,10]. In value-based care, claims and provider data are especially important because payment models depend on cost, utilization, episode attribution, quality measures, and care coordination [6,12]. Healthcare data are rarely clean decision inputs. The same patient may be represented across multiple systems, coding practices may differ across institutions, and clinical notes may contain information that is difficult to standardize. Claims data are useful for longitudinal utilization analysis, but they are produced for billing and administration rather than direct clinical measurement. Imaging and pathology data can contain rich signals, but they require careful annotation, external validation, and site-level review before generalization can be assumed [4,10].

Finance, infrastructure, and industry rely on different data sources, but they face similar engineering problems. Fraud detection systems use transactions, documents, network relationships, device behavior, and historical outcomes [13]. Infrastructure systems use traffic counts, mobility traces, geospatial data, weather, and sensor streams [14,16]. Industrial predictive maintenance systems use telemetry, logs, control data, operating conditions, and maintenance history [15]. In each domain, the useful dataset is not simply the largest dataset. It is the dataset that is appropriate for the decision, stable enough for deployment, and governed well enough to support accountability.

The technical stack of a decision system can be summarized in five layers. The first layer is data acquisition and integration. The second layer is data engineering, including preprocessing, feature construction, linkage, quality checks, and leakage prevention. The third layer is learning, where models estimate risks, classify events, detect

anomalies, forecast demand, or recommend actions. The fourth layer is validation, where discrimination, calibration, subgroup performance, external validity, and operational utility are assessed. The fifth layer is serving and monitoring, where outputs are delivered to users and observed over time.

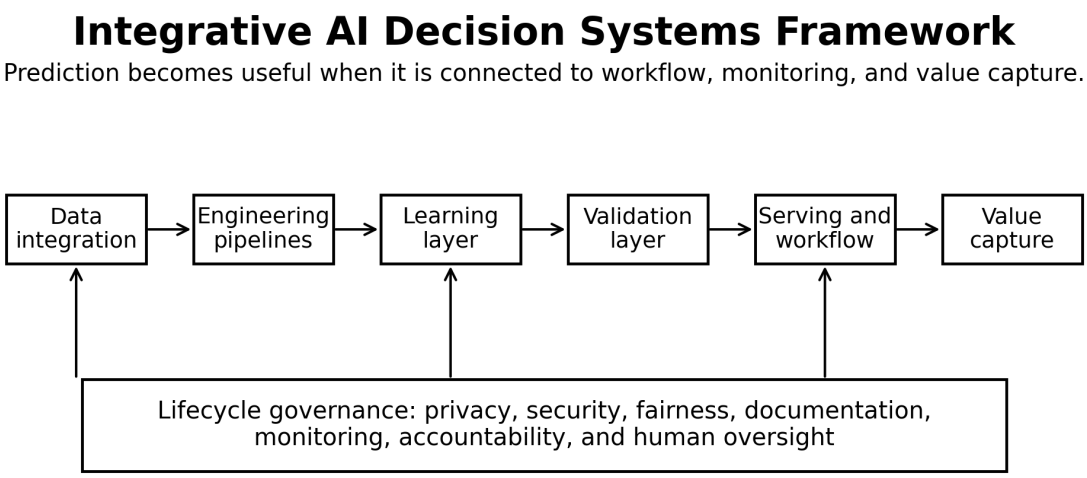

Figure 2. Integrative AI Decision Systems framework showing six linked components and lifecycle governance.

Framework on Integrative AI Decision Systems

The IADS framework defines an AI decision system as the combined set of data, models, software interfaces, monitoring routines, governance processes, and human decisions that translate information into action. This definition deliberately places the model inside a larger system. In high-stakes settings, the model is only one part of the chain that determines whether a decision is useful, safe, and accountable. The first component is data integration. A decision system needs data sources that represent the decision problem with enough completeness and reliability to support use. In healthcare, this may require linking claims, EHRs, laboratory values, medications, notes, and provider information. In finance, it may require linking transactions, account histories, documents, and network relationships. In infrastructure and industry, it may require joining sensor streams with weather, maintenance, or control-system data. The central issue is whether the system captures the right signals without introducing leakage, bias, or unmeasured context.

The second component is engineering. Engineering pipelines determine how raw information is transformed into model-ready data and how those transformations are maintained. This includes missing data handling, feature construction, time-window definitions, normalization, version control, and data quality checks. Weak engineering can make a model appear strong during development while making the deployed system unstable. A decision system should therefore document the data pipeline with the same seriousness as the model architecture. The third component is learning. This layer includes the model classes used to estimate risk, detect anomalies, forecast events, or support recommendations. Tree-based ensembles, survival models, deep learning systems, natural language processing models, graph models, and multimodal models can all be appropriate depending on the task [1-4,10]. The IADS framework does not assume that the most complex model is the best choice. In a high-stakes setting, the appropriate model is the one that supports the required decision with acceptable validity, interpretability, stability, and operational cost.

The fourth component is validation. Validation determines whether the system is ready for the setting in which it will be used. Discrimination metrics such as AUROC or AUPRC can show whether the model separates cases, but they do not show whether predicted probabilities are reliable. Calibration matters when outputs are used to set thresholds, allocate resources, or trigger interventions [6,17]. External validation matters because a model tested

only in its development environment may not perform reliably elsewhere [7,17,18]. Subgroup evaluation matters because average performance can hide clinically important failures [8,9]. The fifth component is serving and workflow embedding. An AI output has practical value only if it reaches the right user at the right time in a form that supports action. A prediction that arrives after the decision point has passed is not useful. A prediction that increases alert burden without changing care or operations may create cost without benefit. A prediction that clinicians, analysts, or operators cannot audit may not be trusted even if it is statistically accurate.

The sixth component is value capture. Value capture asks whether the system improves the decision it was designed to support. In healthcare, this may include better care coordination, reduced avoidable utilization, improved patient safety, improved access, more transparent episode attribution, or better alignment with value-based care goals [6,12]. In finance, it may include reduced fraud loss and fewer unnecessary reviews. In infrastructure, it may include improved routing, reduced congestion, or better resource allocation. In industry, it may include less downtime and more reliable maintenance timing.

Component	Technical function	Deployment question
Data integration	Combines heterogeneous sources into a usable representation	Do the inputs represent the decision problem without avoidable bias, leakage, or missing context?
Engineering pipelines	Transforms, validates, and maintains data for model use	Are preprocessing, features, versions, and quality checks stable across time?
Learning layer	Produces predictions, classifications, rankings, or recommendations	Is the selected model appropriate for the decision and the users who must act on it?
Validation layer	Evaluates discrimination, calibration, generalization, fairness, and utility	Is performance reliable in the intended population and workflow?
Serving and workflow	Delivers outputs into operational settings	Does the output arrive in time and in a form that supports action?
Value capture	Measures whether the system improves decisions or resource use	Does the system create measurable clinical, operational, or financial value without unacceptable harm?

Table 4. Core components of the Integrative AI Decision Systems framework.

Validation, Monitoring, and Governance Parameters

A common weakness in applied AI is the tendency to treat retrospective accuracy as deployment readiness. This is especially risky in healthcare because models may influence decisions about triage, diagnosis, utilization, care coordination, and resource allocation. A model can achieve strong discrimination in a retrospective dataset while still being poorly calibrated, poorly timed, biased across subgroups, or unable to support a feasible intervention [7-11].

Calibration is central to deployment because many operational decisions depend on thresholds. A poorly calibrated model can overstate or understate risk, which may lead to unnecessary intervention, missed care opportunities, or distorted resource allocation. Calibration should therefore be evaluated alongside discrimination. Reliability plots, calibration-in-the-large, calibration slope, Brier score, and threshold-based assessment can help determine whether the output is suitable for the intended decision [17,18]. External validation is equally important. A model that performs well within one hospital, payer, city, market, or industrial site may not perform the same way in another. Differences in documentation, patient mix, coding, treatment patterns, device conditions, or behavioral patterns can alter performance [7]. Reporting standards such as TRIPOD+AI and PROBAST+AI make this issue explicit by emphasizing transparent reporting, risk of bias, and applicability [17,18].

Monitoring is the bridge between initial validation and sustained use. A deployed system should be monitored for input drift, output drift, calibration drift, subgroup performance, alert burden, override rates, and unexpected downstream effects. In healthcare, silent performance degradation can be dangerous because clinical teams may continue to trust an output after the data environment has changed. In finance, fraud strategies can adapt to detection systems. In infrastructure and industry, physical and behavioral conditions can shift over time [11,13-

16]. Governance should not be added after model development. It should begin when the decision problem is defined. A governance process should document the intended use, target population, data sources, evaluation results, limitations, responsible owners, monitoring plan, escalation process, and human oversight structure. The level of governance should be proportional to the consequences of the decision. A low-stakes scheduling tool does not require the same review as an AI system influencing patient care, payment decisions, or safety-critical infrastructure.

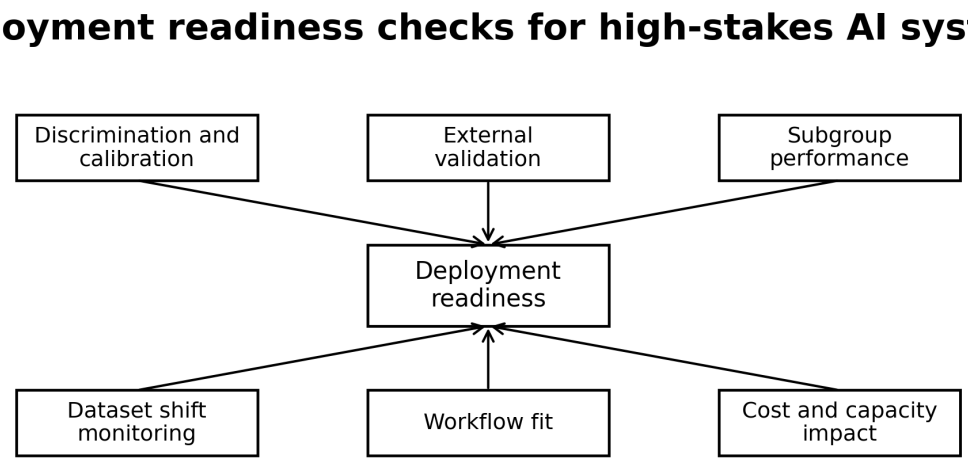

Figure 3. Statistical and operational checks required before and after deployment of high-stakes AI systems.

Parameter	Purpose	Reason for inclusion
Discrimination	Measures ranking or separation between outcome groups	Useful but insufficient alone
Calibration	Assesses whether predicted probabilities match observed outcomes	Needed for thresholds, triage, and resource allocation
External validation	Tests performance outside the development environment	Reduces risk of site-specific overfitting
Subgroup performance	Examines variation across clinically or operationally relevant groups	Detects hidden inequities and generalization failures
Dataset shift monitoring	Tracks changes in inputs, outcomes, and relationships over time	Prevents silent degradation after deployment
Workflow impact	Evaluates timing, alert burden, usability, and actionability	Determines whether predictions can be used in practice
Governance documentation	Records intended use, owners, limitations, and escalation paths	Supports accountability and auditability

Table 5. Validation and governance parameters for deployment readiness.

Cross-Sector Decision Patterns

The cross-sector comparison shows that the IADS framework is not limited to healthcare. The language of each sector differs, but the technical pattern is similar. A decision system must integrate data, produce a prediction, validate its use, embed the result into a workflow, monitor performance, and measure whether the action creates value. In healthcare, the key risks are patient safety, bias, calibration failure, dataset shift, workflow mismatch, and weak accountability. These risks are intensified by the fact that healthcare data are generated through clinical and administrative processes rather than neutral measurement systems. Claims data reflect billing rules, EHR data reflect clinical documentation, and imaging data reflect acquisition protocols and labeling practices. AI systems

used in healthcare must therefore be evaluated not only as technical models, but also as tools operating within a care delivery system [1-12].

In finance, AI systems often operate under class imbalance and adversarial adaptation. Fraud is rare compared with legitimate activity, but missed fraud can be costly and false positives can create operational burden or customer harm [13]. The decision problem is therefore cost-sensitive from the beginning. Accuracy alone may be misleading if it hides false positive review burden, missed high-value fraud, or drift in behavior after the system is deployed. In smart infrastructure, AI systems are linked to real-time or near-real-time decision-making. Traffic prediction, routing, energy management, and urban planning tools must balance efficiency, reliability, equity, and privacy [14,16]. A technically accurate prediction may still be unacceptable if it shifts burden toward disadvantaged communities, depends on intrusive data collection, or fails under unusual conditions. Infrastructure AI also requires public accountability because the decisions often affect shared spaces and public resources.

In industrial operations, predictive maintenance illustrates the value of actionability. A model that predicts a failure but provides no useful maintenance window may have limited value. A model that triggers too many false alarms can increase downtime rather than reduce it. Transferability is also difficult because equipment, operating conditions, sensors, and maintenance histories differ across sites [15]. The final question is whether the prediction changes maintenance decisions in a way that improves reliability or cost.

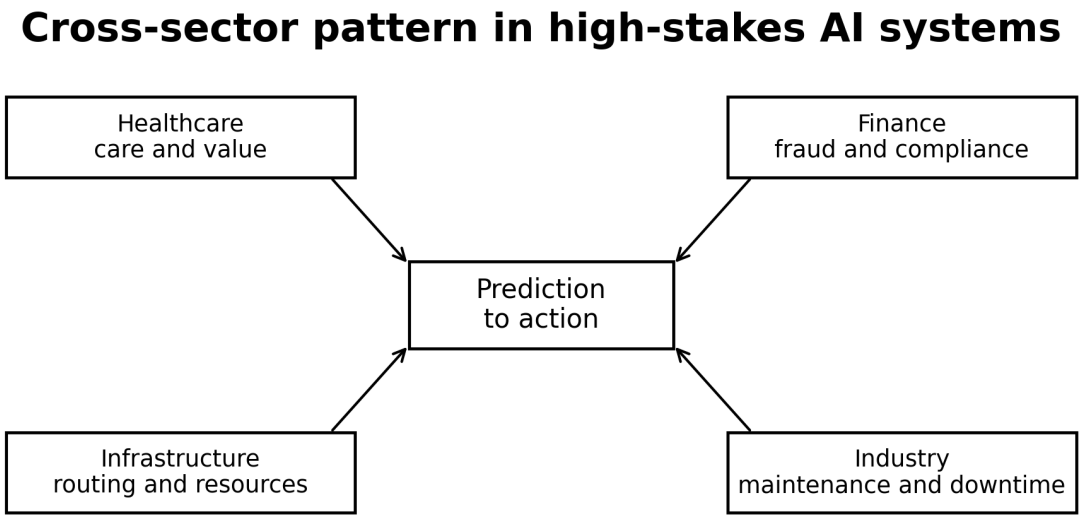

Figure 4. Cross-sector pattern showing how different high-stakes domains share the same prediction-to-action problem.

Domain	Typical data	Dominant model classes	Decision target	Key deployment risks
Healthcare	Electronic health records, claims, notes, images, labs	Boosting, survival models, multimodal models, language models	Risk, diagnosis support, utilization, care coordination	Dataset shift, calibration failure, bias, workflow mismatch
Finance	Transactions, credit histories, documents, network data	Ensembles, anomaly detection, graph ML, cost-sensitive models	Default, fraud, compliance, pricing	Class imbalance, adversarial behavior, drift, regulatory scrutiny
Smart infrastructure	Sensors, mobility traces, weather, geospatial data	Time-series models, optimization hybrids, reinforcement learning	Routing, congestion, resource allocation, control	Multi-objective trade-offs, equity, privacy, real-time reliability

Industry	Machine telemetry, logs, controls, maintenance history	Time-series models, boosting, deep sequence models, anomaly detection	Failure prediction, maintenance timing, process control	Alarm fatigue, transferability, downtime cost, interpretability

Table 6. Cross-sector patterns in integrative AI decision systems.

Applications in Healthcare and Value-Based Care

The IADS framework is especially relevant to healthcare systems that are moving toward value-based care. In traditional fee-for-service settings, data may be used mainly for billing, documentation, and retrospective reporting. In value-based arrangements, the same data must support prospective risk identification, episode management, quality improvement, utilization review, and care coordination. This changes the role of AI from prediction to operational decision support [6,12].

One application is population risk stratification. AI systems can help identify patients who may benefit from outreach, care management, or clinical review. However, the useful question is not simply who has the highest predicted risk. The more important question is which patients can be helped through a specific action that is available within the organization. A risk score should therefore be paired with intervention capacity, equity review, and monitoring of outcomes after action. A second application is episode-level transparency. Value-based care often depends on assigning services, providers, costs, and outcomes to a coherent episode of care. This is technically difficult because patient journeys cross multiple providers, sites, claims systems, and time windows. AI and language-model methods may help organize provider directories, identify relationships, and make episode structures more transparent [12]. The governance requirement is high because attribution decisions can affect accountability, payment, and provider behavior.

A third application is utilization management. Predictive systems may identify avoidable admissions, readmissions, emergency department use, post-acute care variation, or inefficient referral patterns. These uses can support cost reduction, but only if they are designed carefully. A system focused only on cost could create inappropriate incentives. A value-based system should evaluate cost, quality, access, and fairness together. The IADS framework therefore treats value capture as a balanced outcome rather than a simple reduction in spending. A fourth application is clinical workflow support. AI outputs can support imaging review, triage, documentation, diagnosis, and patient monitoring [1-5,10]. These systems require careful attention to human oversight. Clinicians should not be forced to accept opaque outputs, but they also should not be burdened by poorly timed alerts. A good decision system provides the right information at the moment when it can improve judgment.

Challenges of Integrative AI Decision Systems

The first challenge is data quality. High-stakes AI systems often depend on data generated for purposes other than the decision being supported. Claims data are produced for reimbursement, EHR data for clinical documentation, and industrial logs for operations. These data can be useful, but they contain missingness, measurement variation, coding artifacts, and institutional patterns. A decision system must therefore include data provenance and quality checks as part of its design. The second challenge is actionability. A prediction is not useful unless the organization can respond. In healthcare, a model may identify patients at high risk of hospitalization, but the system must still have staffing, outreach capacity, clinical pathways, and follow-up processes. In finance, a fraud score must connect to review queues and customer handling. In infrastructure, a forecast must connect to routing or control decisions. Actionability links model performance to operational capacity.

The third challenge is fairness. Fairness concerns arise when historical data encode unequal access, unequal treatment, or unequal measurement. In healthcare, a model trained on cost may underestimate need in populations that historically had less access to care. In infrastructure, mobility data may overrepresent people with certain devices or travel patterns. In finance, fraud or credit data may reflect prior institutional decisions. Fairness review should therefore examine both the data and the downstream decision. The fourth challenge is interpretability. In high-stakes decisions, users need to understand enough about the output to decide whether and how to act. This

does not mean every system must use a simple model. It means that the explanation, documentation, and interface must be appropriate for the decision context. Post-hoc explanations can be helpful, but they do not replace interpretable design when the decision is sensitive or when users need direct reasons for action [22]. The fifth challenge is governance. Many AI projects fail because ownership is unclear after deployment. A system should have named owners for data quality, model monitoring, user training, incident response, and periodic review. Governance also requires boundaries. A model should not be used for populations, settings, or decisions beyond its validated scope. If use changes, validation should be repeated.

Guideline or standard	Primary focus	Relevance to IADS
TRIPOD+AI	Reporting clinical prediction models using regression or machine learning	Supports transparent reporting of model development and validation [17]
PROBAST+AI	Risk of bias and applicability assessment for prediction models	Supports structured review of model quality and deployment fit [18]
CONSORT-AI	Clinical trial reports for interventions with an AI component	Supports reporting of clinical evaluation and human-AI interaction [19]
SPIRIT-AI	Clinical trial protocols for interventions with an AI component	Supports prospective planning for AI evaluation [20]
DECIDE-AI	Early-stage clinical evaluation of AI decision support systems	Emphasizes real-world use, human factors, and early clinical evaluation [21]
PRISMA-ScR, JBI, and SWiM	Scoping review conduct and transparent synthesis without meta-analysis	Support the review design, screening logic, extraction process, and narrative synthesis [24-26]

Table 7. Selected reporting and governance standards relevant to AI decision systems.

Discussion

The IADS framework highlights a practical gap in the way AI systems are often discussed. Many articles describe model development, but fewer describe the full decision pathway. In high-stakes environments, that pathway matters. A prediction can be statistically impressive and still fail to create value if it is not calibrated, if it reaches the user too late, if it produces too many alerts, if it performs unevenly across groups, or if no accountable process exists for acting on it.

Healthcare makes this problem visible because AI outputs can affect patient care and resource allocation. A clinical prediction system may influence who receives outreach, who is reviewed for admission risk, who receives a care management intervention, or how an episode of care is interpreted. These decisions are not purely technical. They are connected to patient safety, equity, clinical judgment, organizational capacity, and payment incentives [6-12].

The cross-sector comparison strengthens the framework. Finance, infrastructure, and industrial operations face different data and decisions, yet they share the same deployment structure. In each sector, AI must move through data integration, engineering, learning, validation, workflow, and value capture. In each sector, monitoring is necessary because the environment changes. In each sector, governance is necessary because the output affects people, resources, or safety.

The framework also clarifies why accuracy-only reporting is insufficient. Accuracy can be useful during model development, but deployment requires additional questions. Are probabilities calibrated? Was the model externally validated? Does performance hold across subgroups? What happens when the data distribution changes? What action follows the prediction? What is the cost of a false positive or false negative? Who is responsible when the system behaves unexpectedly?

For value-based care, the framework suggests that AI should be evaluated by its ability to support accountable and equitable action. A system that identifies expensive patients without improving care coordination may reduce trust and create poor incentives. A system that identifies actionable episodes, supports transparent attribution, and connects predictions to appropriate interventions may better align with value-based payment goals. The difference is not simply the model. It is the decision system around the model.

This study has limitations. Although the search was structured and the screening process was documented, the review was designed as a scoping technical synthesis rather than a registered systematic review or statistical meta-analysis. The included sources use different methods, domains, data types, and outcomes, which limits direct quantitative comparison. The screening was conducted for framework development rather than pooled effect estimation. Future empirical work should test the IADS framework in specific healthcare deployments and compare decision-system outcomes before and after implementation.

Conclusions

AI should be evaluated as a decision system when it is used in healthcare or another high-stakes environment. The model is only one component. Data integration, engineering, validation, workflow design, monitoring, governance, and value capture determine whether the system can support responsible action.

The IADS framework organizes these requirements into six linked components: data integration, engineering pipelines, learning, validation, serving, and value capture. Governance cuts across every component because privacy, fairness, interpretability, security, documentation, and human oversight cannot be separated from deployment. This structure helps distinguish a prediction model from a usable decision system.

Healthcare shows why this distinction matters. AI outputs may influence diagnosis, triage, utilization management, care coordination, episode attribution, and value-based care performance. Finance, smart infrastructure, and industry show the same underlying pattern in different forms. Across sectors, AI creates durable value only when predictions are connected to validated workflows, monitored use, and accountable decisions.

Future work should apply the IADS framework to empirical case studies, evaluate decision-system readiness across organizations, and develop reporting standards that make actionability as visible as model performance. The goal should not be to deploy more predictions. The goal should be to build systems that help people make better decisions under real-world constraints.

Author Contributions

The author confirms sole responsibility for conceptualization, methodology, investigation, writing, review, and editing of this manuscript.

Funding

This research received no external funding.

Data Availability Statement

All sources included in the framework synthesis are published works listed in the reference section. No patient records, protected health information, or private datasets were accessed. The screening flow and evidence map are summarized in Figure 1 and Table 3.

Acknowledgments

The author acknowledges Fordham University for academic support.

Conflicts of Interest

The author declares no conflicts of interest.

References

1. Rajkomar A, Dean J, Kohane IS. Machine learning in medicine. N Engl J Med. 2019;380:1347-1358. https://doi.org/10.1056/NEJMra1814259
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25:44-56. https://doi.org/10.1038/s41591-018-0300-7
3. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28:31-38. https://doi.org/10.1038/s41591-021-01614-0
4. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. 2022;28:1773-1784. https://doi.org/10.1038/s41591-022-01981-2
5. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA. 2018;319(13):1317-1318. https://doi.org/10.1001/jama.2017.18391
6. Kodan A. Artificial intelligence as a catalyst for value-based health insurance in the United States: narrative review and policy perspective. JMIR AI. 2026;5:e84698. https://doi.org/10.2196/84698
7. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med. 2021;385:283-286. https://doi.org/10.1056/NEJMc2104626
8. Huang Y, Zhu S, Li Y, Lu X, Bao Y, Wu Y. A scoping review of fair machine learning techniques when using real-world data in health care. J Biomed Inform. 2024;151:104633. https://doi.org/10.1016/j.jbi.2024.104633
9. Yang Y, Zhang H, Katabi D, Ghassemi M. The limits of fair medical imaging AI in real-world generalization. Nat Med. 2024;30:2838-2848. https://doi.org/10.1038/s41591-024-03113-4
10. Vorontsov E, Bozkurt A, Casson A, Shao M, Schaub N, Zeng W, et al. A foundation model for clinical-grade computational pathology and rare cancers detection. Nat Med. 2024;30:2924-2935. https://doi.org/10.1038/s41591-024-03141-0
11. Rajagopal A, Goh ES, Kohane IS. Machine learning operations in health care: a scoping review. Cell Rep Med. 2024;5:101679. https://doi.org/10.1016/j.xcrm.2024.101679
12. Kodan A. Enabling episode-level transparency in value-based care through large language model-driven provider directories. Cureus. 2026;18(1):e102219. https://doi.org/10.7759/cureus.102219
13. Compagnino AA, Maruccia Y, Cavuoti S, Riccio G, Tutone A, Crupi R, Pagliaro A. An introduction to machine learning methods for fraud detection. Appl Sci. 2025;15(21):11787. https://doi.org/10.3390/app152111787
14. Shaygan M, Meese C, Li W, Zhao XG, Nejad M. Traffic prediction using artificial intelligence: review of recent advances and emerging opportunities. Transp Res Part C Emerg Technol. 2022;145:103921. https://doi.org/10.1016/j.trc.2022.103921
15. Carvalho TP, Soares FAAMN, Vita R, Francisco RDP, Basto JP, Alcala SGS. A systematic literature review of machine learning methods applied to predictive maintenance. Comput Ind Eng. 2019;137:106024. https://doi.org/10.1016/j.cie.2019.106024
16. Almulhim AI. Integrating artificial intelligence into smart infrastructure management for sustainable urban planning. Technologies. 2025;13(11):481. https://doi.org/10.3390/technologies13110481
17. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. https://doi.org/10.1136/bmj-2023-078378
18. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models and algorithms. BMJ. 2025;388:e082505. https://doi.org/10.1136/bmj-2024-082505
19. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020;26:1364-1374. https://doi.org/10.1038/s41591-020-1034-x
20. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med. 2020;26:1351-1363. https://doi.org/10.1038/s41591-020-1037-7

21. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022;28:924-933. https://doi.org/10.1038/s41591-022-01772-9
22. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1:206-215. https://doi.org/10.1038/s42256-019-0048-x
23. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17:195. https://doi.org/10.1186/s12916-019-1426-2
24. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews: checklist and explanation. Ann Intern Med. 2018;169:467-473. https://doi.org/10.7326/M18-0850
25. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth. 2020;18:2119-2126. https://doi.org/10.11124/JBIES-20-00167
26. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis in systematic reviews: reporting guideline. BMJ. 2020;368:l6890. https://doi.org/10.1136/bmj.l6890