The Eastasouth Journal of Information System and Computer Science
Vol. 4, No. 02, December 2026, pp. 49-68
ISSN: 3025-566X, DOI: 10.58812/esiscs.v4i02

Trustworthy Artificial Intelligence for Enterprise Decision Systems

Barnali Purkayastha
School of Business, International American University, Los Angeles, CA 90010, USA

Article Info

Article history:

Received Nov, 2026
Revised Nov, 2026
Accepted Nov, 2026

Keywords:

Cybersecurity;
Data Governance;
Decision Utility;
Digital Twins;
Enterprise Decision Systems;
Federated Learning;
Model Calibration;
Trustworthy Artificial Intelligence

ABSTRACT

Artificial intelligence increasingly connects business forecasts, clinical information, cloud services, and security operations, yet a prediction becomes useful only when an organization can act on it safely. The corpus is organized into seven primary themes covering business operations, governance, cybersecurity, privacy, healthcare, representation learning, and sustainable infrastructure. Bibliographic coding is distinguished from selectively inspected publisher content; the study is neither a systematic review nor a quantitative comparison of model performance. The synthesis identifies a common design problem: predictive components, privacy mechanisms, and operational controls are often discussed through different objectives, leaving their joint decision consequences difficult to assess. We propose a layered architecture that connects data provenance, calibrated inference, constrained action, independent verification, and monitored feedback. Its evaluation protocol combines temporal and organizational holdouts, budget-matched baselines, adversarial testing, cryptographic cost accounting, and prospective assessment of operational outcomes. Four falsifiable propositions specify when additional verification, privacy protection, simulation, and bounded adaptation could improve deployment. Original figures visualize corpus composition, architectural dependencies, and evidence requirements without presenting simulated values as observations. The principal contribution is an auditable research design for studying trustworthy decision systems across heterogeneous applications. Empirical superiority, clinical effectiveness, and economic benefit remain questions for subsequent evaluation.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Name: Barnali Purkayastha
Institution: School of Business, International American University, Los Angeles, CA 90010, USA
Email: barnalipurka6@gmail.com

1. INTRODUCTION

A retailer deciding which products to replenish, a hospital allocating follow-up appointments, and a security team reviewing suspicious activity face a similar constraint. They rarely have enough time or staff to act on every prediction. Their practical question is which action deserves attention, what evidence supports it, and what happens if that action is wrong. A model can rank cases accurately while overwhelming a review queue, exposing confidential information, or encouraging an intervention whose benefit has never been established. This paper therefore treats the decision process, rather than the predictive model alone, as the unit of analysis.

The supplied business-intelligence literature provides an appropriate starting point because its titles explicitly connect predictive analytics with management decisions. The record on AI-powered business intelligence situates learning within organizational information use [1]. The paper on AI and machine learning in business management addresses a corresponding managerial setting [2]. A review of business-intelligence tools introduces the relationship between decision practices and organizational ambidexterity [3]. These records establish the scope of the present inquiry; their inclusion does not establish that particular tools produce economic gains.

A second group of records makes the governance connection explicit. The study of data governance and business analytics concerns how information practices relate to analytics success [4]. Work on autonomous decision intelligence links explainable AI with decision-support architectures [5]. A business-analytics paper framed around the U.S. economy extends the discussion to broader economic aspirations [6]. The analytical issue is how to translate such aspirations into measurable outcomes without assuming that model accuracy, organizational capability, and national economic effects are interchangeable.

The paper asks three questions. First, which application and control themes appear in this author-bounded corpus? Second, what architecture can connect those themes while preserving the differences between domains? Third, what evidence would be necessary to test whether the resulting system improves decisions? The contribution is a reproducible map of the supplied bibliography, an original conceptual architecture, and a prospective evaluation protocol. No new machine-learning benchmark, clinical trial, or causal economic analysis is reported.

2. REVIEW DESIGN AND EVIDENCE BOUNDARIES

2.1 Corpus definition

The review uses the 63-reference list supplied for five researchers: Udoy Sankar Saha, Evha Rozario, Kazi Bushra Siddiqa, Md Kazi Tuhin, and Emran Hossain. Eligibility follows the supplied author association and the inclusive 2024–2026 date window. Records shared by multiple researchers are retained once. DOI strings serve as the principal deduplication key; stable article URLs are used when no DOI is supplied. Conference dates, online-publication dates, and issue dates are not assumed to be identical.

This is a bounded, purposive corpus. It was not assembled through an exhaustive database query, and it should not be interpreted as a representative sample of global AI research. Its author-based origin creates a risk of intellectual concentration and connected coauthorship. The analysis can describe the supplied records and develop questions from their intersection. It cannot estimate field-wide publication trends, identify the universally best algorithm, or establish that the corpus covers every relevant perspective.

The reference list is preserved as a working bibliography. Differences between supplied versions include author omissions and date conventions in several conference records. The network-survivability article is represented by the supplied 2026 issue version rather than counted again as a 2025 online item. These choices affect bibliographic consistency, not evidence of model effectiveness. Final submission requires checking every reference against its version of record and checking for publication notices.

Sixteen additional publications were selected purposively to address calibration, operational governance, privacy costs, human oversight, and digital-twin deployment. These contextual sources extend the discussion without changing the author-based eligibility rule for the original map. The final bibliography therefore contains 79 unique references: 63 mapped records and 16 supporting records. Only the mapped records contribute to the year

and theme distributions in Figures 1 and 2 and Table 1.

2.2 *Evidence access and interpretation*

Two evidence levels are kept separate. At the corpus level, titles and bibliographic metadata support thematic coding and descriptive counts. At the content level, selected publisher pages were inspected to check whether specific conceptual connections were actually present. The inspectable material included the publisher article page for verifier-grounded fraud detection, the article landing page for federated cyber-threat intelligence, the IEEE abstract for homomorphic federated models, and the publisher abstract for critical-infrastructure protection. Additional publisher information was inspected for the retail-banking calibration study, telecom explainability study, smart-building security study, AEGIS-FL article, homomorphic-encryption chapter, DataOps abstract, and sustainable digital-twin article. Full publisher verification and full-text methodological appraisal remain incomplete for the contextual digital-twin standardization, human-in-the-loop, and blockchain-governance records. Other records are discussed as topical contributions indicated by their supplied titles, without assigning uninspected experimental results to them.

The six further contextual sources extend the synthesis to enterprise AI architecture, multi-cloud design, workforce analytics, cooperative medical-image analysis, and secure imaging infrastructure. The skin-classification and lung-segmentation full texts were inspected; the workforce and cloud-imaging findings below were checked against the institutional repository abstract and publisher abstract, respectively. The enterprise AI book is treated as a practitioner synthesis, while the multi-cloud book is included for its stated architectural scope; its full text was not available for methodological appraisal. Book-level design guidance is distinguished from measured experimental outcomes. The skin article is retained under its 2025 issue year, although its landing page gives a November 2026 publication date; that discrepancy requires editorial confirmation before submission.

This distinction is consequential. A title containing “privacy-preserving” does not demonstrate a formal privacy guarantee. “Real-time” does not establish a latency bound. A paper addressing the U.S. economy does not by itself establish U.S.-wide benefit. Where a title signals an application, the discussion uses it to locate a research problem. Proposed mechanisms, evaluation criteria, and propositions developed here are the authors’ analytical contributions rather than findings attributed to unread full texts.

2.3 *Coding and reproducibility*

Each record receives exactly one primary theme according to its most explicit application or technical objective. Seven themes were defined: business operations and economic decisions; governance and project assurance; cybersecurity and resilience; privacy and cryptography; healthcare and public health; representation learning and digital twins; and energy and environmental infrastructure. Several records could fit more than one theme. Primary assignment is a counting convention, while the prose preserves cross-theme relationships.

Coding was performed in one analytical pass without independent reviewers. No interrater agreement or risk-of-bias score is claimed. The supplement provides the 63 record identifiers, normalized citation labels, assigned themes, and year counts. This makes recoding possible and exposes the subjective choices behind the map. Counts describe records rather than independent experiments, datasets, research teams, or levels of evidence.

Table 1. Primary thematic assignments in the bounded corpus

Primary theme	Records	Share
Business operations and economic decisions	19	30.2%
Governance and project assurance	8	12.7%
Cybersecurity and resilience	13	20.6%
Privacy and cryptography	6	9.5%
Healthcare and public health	9	14.3%
Representation learning and digital twins	5	7.9%
Energy and environmental infrastructure	3	4.8%
Total	63	100.0%

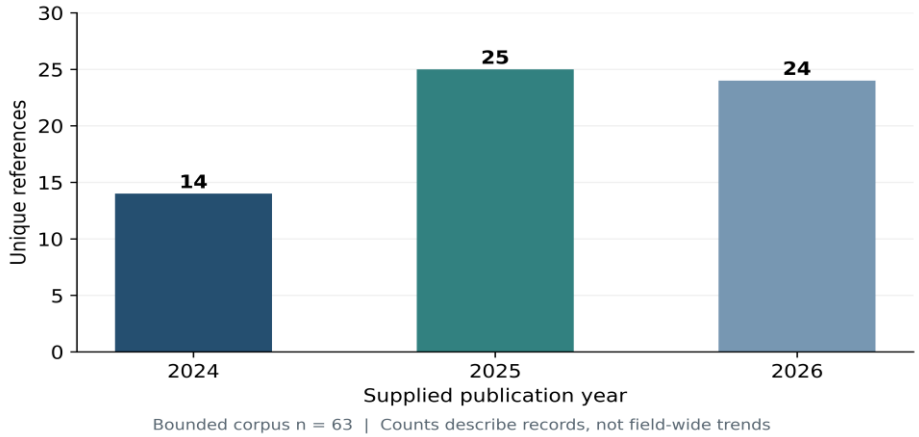


Figure 1. Distribution of the 63 supplied references by publication year. Values are bibliographic counts and do not estimate field-wide growth.

3. CORPUS MAP AND THEMATIC SYNTHESIS

3.1 Business operations and economic decisions

Nineteen records are assigned primarily to business operations. The procurement paper connects intelligent automation with purchasing workflows [7]. The supply-chain language-model paper places generative systems in operational information handling [8]. End-to-end supply-chain optimization provides a related planning context [9]. The integrated resilience, trade, and growth framework broadens the boundary beyond a single organization [10]. These different units of analysis should remain visible: a faster procurement task is not the same outcome as a more resilient supply network.

Sector-specific records help define plausible decision units. The timber-supply-chain paper concerns a particular material flow and its economic-security framing [11]. Retail-channel planning under structural change suggests a setting where historical relationships may become unreliable [12]. The supermarket-sales review concerns forecasting methods for a familiar operational task [13]. For a future comparative study, these applications motivate rolling time splits, explicit treatment of promotions and shortages, and measurement of stockouts or service levels. Those recommendations are proposed here rather than reported as shared findings across the three papers.

Customer and workforce decisions raise different questions. The customer-experience record concerns AI in business interactions [14]. Workforce analytics addresses labor-market trends and skill gaps [15]. The microcredit paper concerns socioeconomic outcomes rather than predictive ranking alone [16]. A model that predicts repayment risk or employee attrition may still produce harmful decisions if the response reduces

access or shifts burdens onto poorly represented groups. Evaluation should therefore identify who receives an action, who bears an error, and which outcomes matter to affected people.

[17] provide a concrete workforce application. Their hybrid HRM approach used structured surveys and historical HR records from 120 small-scale enterprises, with Random Forest and Gradient Boosting models to predict attrition and examine satisfaction, rewards, and work-life balance. The abstract reports 93.2% accuracy, 92% precision, 93% recall, 93% F1-score, and a ROC-AUC of 0.95. These results support the feasibility of predictive workforce screening within the reported setting. They do not, by themselves, establish that an intervention caused employees to remain. For enterprise evaluation, firm-level separation, temporal testing, subgroup error analysis, and prospective measurement of retention outcomes would therefore be needed before translating a risk score into employment decisions.

The credit-risk record brings predictive analytics into a consequential financial setting [18]. The homelessness-cost paper addresses forecasting and resource planning in a public-policy context [19]. Their inclusion demonstrates the range of economic decisions within the corpus, but neither application warrants an inference from forecast quality to causal social benefit. Such benefit requires credible comparison of actual interventions, their costs, and the populations they serve.

Smart manufacturing expands the operational environment to production systems [20]. Digital transformation and IT-project success extend it to organizational implementation [21]. The paper on managerial decisions under information noise adds a directly relevant uncertainty setting [22]. Across these records, the present synthesis proposes that the useful common denominator is a documented action rule under constraints. Models can differ while the requirement to specify available actions, decision timing, capacity, and error consequences remains stable.

The contextual banking study explicitly combines calibration, cost-sensitive decision rules, explainability, and leakage-safe evaluation [23]. This combination is particularly relevant to the proposed separation between prediction quality and action quality. A telecom churn benchmark supplies a second business setting for model comparison and SHAP-based interpretation [24]. Neither application establishes that an explanation is causal or that a retention intervention benefits every customer; those questions require their own evaluation.

3.2 Governance and project assurance

Eight records primarily concern governance or project assurance. A trustworthy data-lakehouse title combines federated learning and blockchain within an information architecture [25]. The cloud access-control chapter explicitly addresses compliance auditing [26]. Governance-oriented cyber-risk management extends the topic to organizational resilience and infrastructure protection [27]. These records suggest distinct control surfaces: who may access data, how analytical changes are approved, and how responsibility is assigned after an automated action.

Project-management papers introduce a practical implementation layer. AI-driven project management links management information systems with IT-project efficiency [28]. AI-augmented risk intelligence connects project evaluation with learning methods [29]. Software quality assurance brings predictive analytics, digital twins, and agile methods into the same discussion [30]. The synthesis treats these records as motivation for traceable release criteria. A project milestone is useful only if the

tested software, deployed model, policy configuration, and monitored outcomes can be matched to the same version.

Governance should therefore be represented as executable conditions, not only written principles. A data-source restriction must be enforced at ingestion and logged at access. A model-use limitation must affect the action policy. A requirement for human review needs a named queue, a response time, and an escalation route. These are proposed engineering implications, not claims that every governance paper in the corpus implements them.

The contextual DataOps paper connects automated decision pipelines with versioned engineering, lineage tracking, validation, and policy enforcement [31]. It provides a more concrete operational bridge between data governance and the release process proposed here. The blockchain-governance record is topically relevant to accountable access and provenance [32]. Its role in this synthesis is to motivate a control surface; blockchain is not assumed to establish that an input record is accurate or that its use is authorized.

3.3 Cybersecurity and resilience

Thirteen records are assigned primarily to cybersecurity. The cloud-protection framework identifies a familiar environment in which data access and threat detection interact [33]. The publisher abstract for the critical-infrastructure paper discusses AI-supported detection and response alongside policy considerations [34]. The data-system-security paper places AI within a broader security discussion [35]. Together, they motivate a boundary between detecting suspicious events and authorizing responses that could disrupt operations.

Detection records cover different representations and threat surfaces. Malware image classification uses binary-image representation as its stated problem setting [36]. Real-time threat identification concerns security analytics [37]. Big-data cyberattack detection introduces proactive mitigation [38]. A future benchmark should preserve these task differences. Accuracy on malware images cannot be pooled with event-level intrusion detection without reconciling labels, prevalence, evaluation units, and deployment constraints.

Adaptation creates a further design tension. Continual-learning pipelines are positioned around insider threats [39]. Self-healing cybersecurity is framed through reinforcement-learning agents [40]. Containerized microservices are the stated environment for autonomous response mechanisms [41]. Adaptation may help when behavior changes, but it also changes the system being evaluated. A deployment should log update triggers, freeze a recoverable prior version, and test whether an attacker can influence the update process. These safeguards are requirements of the proposed framework rather than demonstrated properties of the cited systems.

The distributed-systems architecture record broadens resilience beyond a detector [42]. The national cyber-threat-intelligence framework similarly adopts a wider defensive scope [43]. Network survivability examines traffic and structural importance [44]. Biometric verification in mobile finance adds an adversarial identity setting [45]. In each case, practical performance depends on what remains available after a failure or attack. A high detection score is inadequate if the response blocks legitimate users, disconnects a necessary service, or produces more cases than analysts can examine.

The contextual smart-building paper places intrusion detection and resilience within an IoT sensor environment [46]. It provides an application for testing the connection between detection quality, service continuity, and the consequences of response. AEGIS-FL explicitly combines federated threat detection with

auditability in multi-tenant cloud-edge systems [47]. Its stated concerns about participant behavior and containment strengthen the motivation for assessing who can alter a shared model and who can authorize an operational response.

3.4 Privacy and cryptographic controls

Six records primarily address privacy and cryptography. Federated cyber-threat intelligence places distributed training in cloud-integrated management systems [48]. The inspected IEEE abstract for homomorphic federated models explicitly discusses gradient-related threats and encrypted aggregation [49]. Secure multi-party computation appears in the cloud-tenant intrusion-detection chapter [50]. These mechanisms have different trust assumptions and should not be treated as interchangeable labels for confidentiality.

The resilient-MIS chapter frames cryptography within system design [51]. The privacy-preserving MIS chapter explicitly combines blockchain, federation, and zero-trust models [52]. Distributed healthcare analytics supplies a sensitive application setting for federated learning [53]. The central analytical issue is composition. Keeping raw data local does not prevent every disclosure through model updates or outputs. Encryption protects particular representations under defined assumptions; it does not remove risks from authorized misuse, endpoint compromise, or incorrect decisions.

A credible privacy evaluation should identify the protected asset, adversary access, participant-collusion assumptions, and recovery procedure after key compromise. It should also measure the computational and communication costs imposed by the selected protection. The present paper does not infer differential-privacy guarantees, resistance to all inference attacks, or production scalability from titles. Where such properties are claimed in a source, they require examination of the mechanism, parameters, threat model, and evaluation conditions.

The contextual homomorphic-encryption chapter explicitly examines alternative encryption approaches and their performance trade-offs in public-cloud analytics [54]. It supports treating cryptographic cost as an evaluation dimension. Any numerical comparison must preserve its original workload, hardware, parameterization, and measurement conditions; the present review does not transfer those measurements to an enterprise deployment or to a different threat model.

[55] connect diagnostic modelling with cloud security through a pipeline of image normalization and enhancement, encryption, secure transmission and storage, authentication, integrity checking, and deep-learning analysis. Their publisher abstract describes 28,500 medical images across multiple imaging types and reports a Matthews correlation coefficient of 0.957, balanced accuracy of 98.08%, Cohen's kappa of 0.954, and an AUC of 99.12%. These are reported study results rather than independently reproduced estimates. The accompanying privacy and access-security percentages are operational indicators and should not be interpreted as formal cryptographic guarantees. This study motivates evaluating predictive performance and infrastructure protection together, while requiring separate scrutiny of the threat model, validation partitions, and deployment workload.

3.5 Healthcare and public health

Nine records primarily concern healthcare. The personalized-cancer-treatment review places big-data analytics in oncology [56]. The chronic-disease prediction record focuses on early diagnosis and risk stratification [57]. Oral-image classification offers a specific imaging application [58]. These contexts have different labels and clinical pathways. An image classifier can

support a bounded diagnostic task without establishing that a broader healthcare workflow improves patient outcomes.

The cardiovascular paper is explicitly framed around survey-weighted analysis [59]. This title points to an important distinction between population estimation and predictive classification. A future evaluation using survey data should preserve its sampling structure rather than assuming that a convenient random split represents the target population. Where outcomes are clinical, the action rule should identify when a clinician reviews a result and which downstream measurement would indicate benefit or harm.

Global health intelligence concerns epidemiological surveillance and forecasting infrastructure [60]. Epidemic-response prediction links forecasts to healthcare-resource distribution [61]. An AI-enabled healthcare-ecosystem paper frames predictive, preventive, and personalized medicine at a broader level [62]. The present architecture accommodates these applications by distinguishing data refresh, forecast horizon, allocation decisions, and feedback delay. It does not assume that findings from one disease or institution transfer to another.

Healthcare software quality assurance adds an implementation perspective through predictive analytics and digital twins [63]. Drug-development and biomarker discovery adds a scientific-discovery context [64]. These are adjacent to operational decision support, but they are not equivalent endpoints. Their value in this corpus lies in identifying where provenance, verification, and validation must be retained as information moves between research software, clinical interpretation, and operational use.

[65] introduce an Attention-Guided Cooperative Multi-Agent Network for skin-lesion classification. Four agents handle global features, local lesion characteristics, attention-based representations, and cooperative classification; hierarchical attention and confidence-weighted coordination combine their outputs, with a feature-consistency mechanism to limit disagreement. HAM10000 supplies 10,015 images across seven categories, with ISIC-2019 as an additional benchmark. At 25 epochs, the paper reports 96.10% accuracy, 95.60% precision, 95.20% recall, 95.40% F1-score, and a macro-AUC of 0.974. Its methodological relevance is the explicit division of perceptual tasks among coordinated agents. Attention maps and agreement remain supporting evidence: independent population testing and a direct assessment of explanation usefulness are needed before treating them as clinical assurance.

[66] apply cooperation to a different endpoint through the Cooperative Multi-Agent Attention-Guided Segmentation Network. Using LIDC-IDRI CT data, the architecture assigns agents to anatomical localization, multiscale feature extraction, boundary refinement, and segmentation optimization. A shared attention-guided encoder, confidence-weighted feature aggregation, and a boundary-aware decoder produce a three-dimensional tumor mask. At 25 epochs, the reported results are 91.8% Dice, 85.0% intersection over union, 92.6% sensitivity, 93.1% precision, and a 95th-percentile Hausdorff distance of 3.2 mm. These findings concern spatial overlap and boundary quality, rather than diagnostic classification accuracy. They support testing whether agent specialization improves anatomical delineation, but do not establish better treatment outcomes or transferability to an enterprise workflow.

3.6 Representation learning and digital twins

Five records primarily concern representation learning or digital twins. Hierarchical architectures address

complex representation [67]. Adaptive multimodal agents introduce interaction with changing environments [68]. Transformer-based summarisation concerns compression of large information streams [69]. A summary can be fluent while omitting the exception that matters to a decision. The proposed system therefore retains pointers to source records and treats missing support as a reason to abstain or request review.

[70] offers a practitioner-oriented synthesis of enterprise artificial intelligence spanning agentic systems, large language models, retrieval-augmented generation, deployment architecture, and governance. Its approach is architectural and explanatory rather than a controlled benchmark experiment. The contribution relevant here is a connected account of how models, knowledge retrieval, orchestration, and operational controls can be organized within enterprise platforms. Consequently, the book informs the design rationale for the proposed layers; it supplies no quantitative effectiveness estimate used in this review. Any claim that such integration improves decision quality must still be tested through the prospective protocol in Section 5.

The inspected verifier-grounded fraud article provides a concrete connection between language-model outputs and independent checks [71]. Its role here is to motivate a verifier boundary, not to transfer its reported performance to other domains. The digital-twin framework brings trust and optimization into a simulation-oriented architecture [72]. Simulation is useful for exploring failures and constraints, but its relevance depends on whether its assumptions match the deployed environment. Neither a plausible explanation nor a realistic-looking simulation substitutes for external evaluation.

The contextual human-in-the-loop record addresses language-model reliability through its stated human-feedback setting [73]. It motivates an explicit distinction between automated checks and human adjudication in the architecture. The cross-domain digital-twin standardization paper is topically relevant to secure edge intelligence and interoperability [74]. These additions identify design questions for escalation and deployment across heterogeneous systems; neither is treated here as proof that a particular oversight or standardization mechanism is effective in every domain.

3.7 Energy and environmental infrastructure

Three records primarily concern energy or environmental infrastructure. The energy-resilience paper addresses monitoring through management information systems [75]. Coal-production analytics explicitly links energy production with AI-infrastructure growth [76]. Climate-change modeling introduces multisource geospatial information [77]. These records are retained because they define resource and infrastructure settings in which decision systems operate. They do not establish the environmental benefit of adding AI.

An operational evaluation should account for energy used by training, inference, communication, and cryptographic protection alongside the consequences of the decisions taken. A proposed forecasting service might reduce wasted capacity while increasing computational demand. Both sides need a stated boundary and measurement period. The same requirement applies to climate-related applications: spatial holdouts and source-date alignment are necessary to distinguish transferable information from proximity-based leakage.

The contextual sustainable-computing article brings digital-twin-based energy optimization into a smart-infrastructure setting [78]. It strengthens

the connection between the simulation layer and resource accounting. The proposed evaluation should retain the difference between a data-driven simulation and measured energy savings after deployment, with the latter requiring a specified physical system, time period, and comparator.

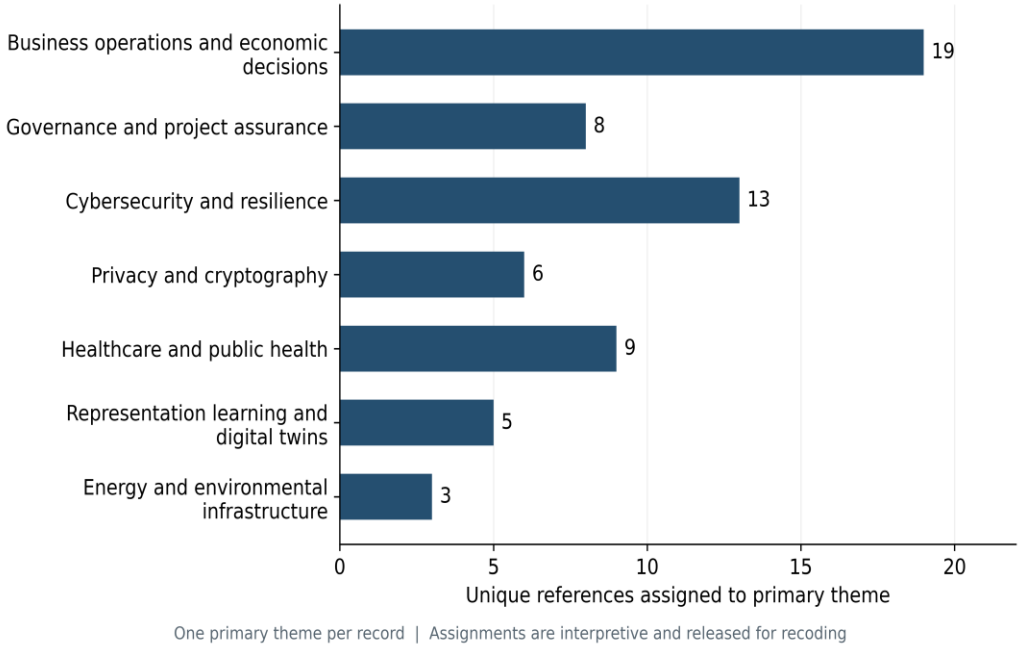


Figure 2. Primary thematic assignments. Each record is counted once; themes are interpretive categories rather than evidence-quality grades.

4. A LAYERED ARCHITECTURE FOR TRUSTWORTHY DECISIONS

4.1 Data contracts and provenance

The proposed architecture contains four operational layers and a cross-cutting assurance function. The first layer defines data contracts: permitted sources, collection times, entity identifiers, transformations, retention rules, and ownership. A contract should be machine-readable where possible. Every feature must have an availability time that precedes the decision it informs. This requirement prevents the apparently strong model that quietly uses information arriving after the event.

[79] addresses enterprise multi-cloud architecture through the stated themes of design, security, automation, and operational practice. This is relevant to the infrastructure beneath data contracts, particularly when ownership, access policies, and processing responsibilities span cloud environments. Because the book's full text was unavailable for appraisal, this review does not attribute a specific implementation method, benchmark, or measured security improvement to it. Instead, its architectural scope motivates a testable design question: can provenance and policy enforcement remain consistent as a workload moves between providers? The proposed protocol should measure that consistency alongside latency, cost, and failure recovery.

A decision record should connect its source-data snapshot, preprocessing version, model identifier, calibration version, policy version, verifier result, and final action. This record is the minimum unit for incident investigation. If an input is corrected later, the system should retain both what was known at decision time and the corrected value. Replacing history with

the latest database state can make retrospective evaluations misleading.

4.2 *Prediction and uncertainty*

The second layer produces scores, forecasts, intervals, or structured candidate outputs. Representation choice depends on the task rather than a preference for maximum complexity. A tabular baseline, a compact neural network, an image encoder, and a language model may all be appropriate in different parts of the workflow. The evaluation should establish whether the extra component contributes useful information after accounting for latency, data needs, and maintenance.

For binary outcomes, calibration can be expressed as the requirement that cases assigned a probability near p experience the outcome at approximately that frequency in the intended population. Calibration must be checked after deployment-relevant shifts. A model may retain useful ranking while its probabilities become unsuitable for cost-sensitive decisions. Uncertainty estimates should therefore inform routing and abstention, not merely appear as an additional dashboard statistic.

4.3 *Constrained action and independent verification*

The third layer maps predictions to permitted actions. Let x denote information available at decision time and let a belong to the allowed action set A(x). The proposed policy selects the action with the highest estimated utility subject to workload, latency, privacy, and safety constraints. This is a design formulation, not an estimated empirical result:

$$\pi(x) = \arg\max \; E[U(a, Y) \mid x], \; a \in A(x)$$

subject to capacity $\leq K$, latency $\leq L$, privacy risk $\leq R$, and prespecified safety constraints.

Here the expected utility concerns consequences of the action, not simply correctness of a prediction. Estimating it may require causal evidence about intervention effects. A risk model alone does not provide that evidence. When only predictive data are available, the policy should be described as a prioritization rule whose operational effects remain to be evaluated.

Independent verification sits between candidate output and action authorization. A verifier may check schema validity, numerical consistency, source support, access permissions, or a domain-specific constraint. It must return a reason-coded result and should be tested independently of the generator. A false rejection can delay a useful action; a false acceptance can admit a dangerous one. The verifier therefore has its own error rates, operating costs, and maintenance needs.

4.4 *Monitoring and bounded adaptation*

The fourth layer records outcomes and monitors whether the deployed policy remains within its approved envelope. Monitoring covers data completeness, score distributions, calibration when labels become available, review queues, incident rates, and subgroup outcomes. A drift alarm should trigger a specified investigation rather than an automatic retraining event whose effect is unknown.

Bounded adaptation permits changes only within documented limits. For example, a threshold may be adjusted on a validation stream while the test protocol remains frozen; a new feature source may require a separate release. Rollback must restore the model, preprocessing, calibration, and policy together. Restoring only model weights can leave the system in an untested configuration.

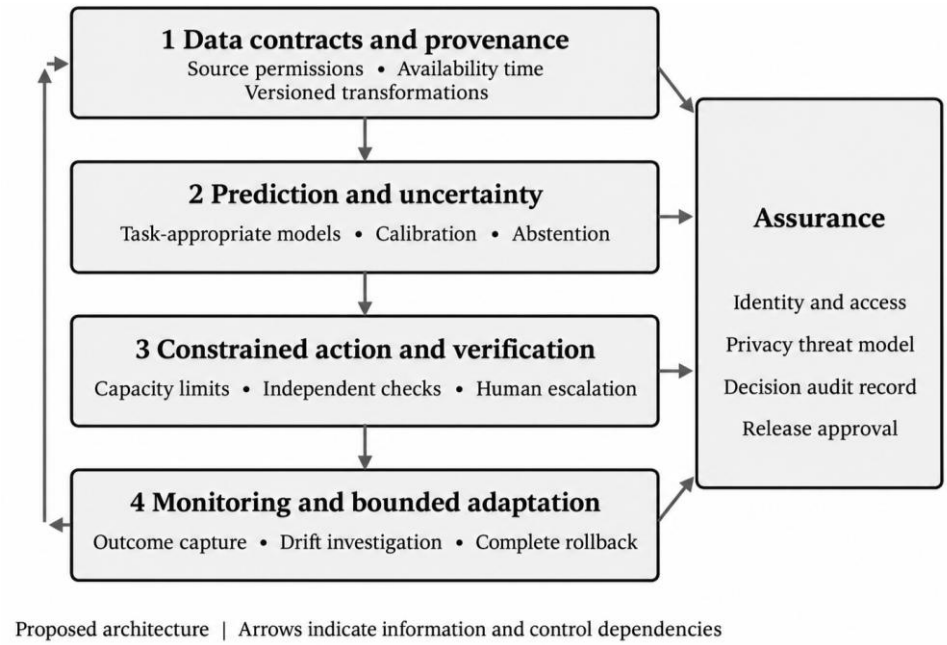


Figure 3. Proposed architecture connecting provenance, prediction, constrained action, monitoring and cross-cutting assurance. This diagram presents a research design rather than a validated implementation.

Table 2. Controls and auditable evidence in the proposed architecture

Control surface	Failure addressed	Required evidence
Data provenance	Future information or unauthorized reuse	Availability timestamps; permissions; source and transformation versions
Prediction	Unreliable probabilities or unsupported transfer	Locked temporal evaluation; calibration; external-site checks
Action policy	Capacity overload or unauthorized intervention	Allowed actions; capacity; cost assumptions; escalation owner
Verifier	Unsupported output or invalid action	Independent tests; false accepts and rejects; reason codes
Privacy	Disclosure through updates or access	Threat model; cryptographic parameters; empirical attack protocol
Adaptation	Poisoning or untested configuration changes	Update gates; versioned release; complete rollback test

5. PROSPECTIVE EVALUATION PROTOCOL

5.1 Tasks baselines and partitioning

A credible evaluation should begin with one primary application and an explicit decision unit. Procurement might use supplier-week allocations; cybersecurity might use analyst-reviewed alerts; healthcare might use eligible encounters. Each study should specify the observation window, outcome horizon, action capacity, and exclusions before model fitting. Secondary applications can examine transfer, but their metrics should not be averaged into a single headline result.

At least three comparators are proposed: the existing operational rule, a strong task-appropriate predictive baseline, and the same predictive model without the proposed assurance component. A fourth comparator can represent a feasible human-assisted workflow. Matching only model architecture is insufficient. Comparators should receive comparable data availability, tuning effort, compute budgets, and decision capacity.

Chronological development and a locked later test period provide the primary partition. Entity-level separation prevents repeated customers, patients, devices, or organizations from appearing across incompatible splits. An external site or organization should test transfer when accessible. All imputation, scaling, feature selection, resampling, and probability adjustment must be fitted within the training procedure. Test labels cannot influence threshold selection or the choice of a favorable reporting period.

5.2 Metrics that preserve operational meaning

Prediction metrics are necessary but incomplete. Rare-event tasks should report prevalence, precision–recall performance, and uncertainty intervals. Forecasting studies should report horizon-specific errors and compare against a defensible seasonal or historical baseline. Probabilistic tasks should assess calibration and proper scoring rules. None of these quantities should be pooled across different datasets without a justified synthesis model.

Operational evaluation should match the decision constraint. For a fixed review capacity k, report how many relevant cases are captured among the selected cases, how long review takes, and what happens to the unselected group. A relative improvement is interpretable only alongside denominators and absolute counts. False-positive reductions can be valuable, but the analysis should show whether they are accompanied by unacceptable missed events.

Table 3. Proposed evaluation outcomes and interpretation limits

Dimension	Measurements	Interpretation limit
Prediction	AUPRC; horizon-specific forecast error; confidence intervals	Scores depend on prevalence, task and evaluation unit
Calibration	Brier score; log loss; calibration plot and slope	Recheck after temporal and site shifts
Decision service	Relevant cases at fixed capacity; review time; queue length	Predictive prioritization does not identify intervention effects
Privacy	Mechanism parameters; attack success; key and endpoint assumptions	Empirical resistance is not a universal formal guarantee
Robustness	Performance under specified attacks, missingness and dropout	State attacker knowledge and permitted modifications
Efficiency	Tail latency; bytes transmitted; compute and energy	Include setup, aggregation and failure recovery
Prospective outcomes	Benefit, harm, total operating cost and subgroup effects	Use a suitable prospective comparison before benefit claims

Privacy, robustness, and efficiency require their own measurements. Attack evaluations should state the attacker's knowledge, allowed modifications, and resource budget. Privacy tests should separate formal guarantees from empirical attack resistance. Efficiency reporting should include hardware, batch size, communication volume, key setup, aggregation time, and tail latency. A cryptographic method that meets an average runtime target may still miss a deadline when clients drop out or connections become unreliable.

5.3 Ablation and sensitivity analysis

The proposed ablation sequence starts with the predictive baseline, then adds calibration, verification, cryptographic protection, and bounded adaptation. Each addition is evaluated under a matched operating budget. Interaction effects matter: privacy mechanisms can change update timing, while verification can change which cases reach human reviewers. A

component should not receive credit for an improvement caused by a different threshold or a larger computation budget.

Sensitivity analyses should vary outcome prevalence, review capacity, intervention cost, missingness, network delay, client participation, and distribution shift. These analyses identify the conditions under which a method ceases to be useful. The result should be a decision-relevant operating region, not a claim that one configuration dominates under every condition.

5.4 Statistical analysis and prospective testing

Uncertainty estimation should respect dependence. Repeated records from the same entity and correlated observations within organizations require grouped resampling or another justified dependence-aware method. Comparisons should use paired observations where appropriate. A prespecified primary endpoint limits selective emphasis, while secondary endpoints should be labeled accordingly. Statistical significance should be accompanied by effect size and a practical relevance threshold.

After retrospective testing, a shadow deployment can record recommendations without allowing them to affect decisions. This tests data arrival, processing time, queue behavior, and audit completeness. A subsequent prospective comparison should evaluate actual outcomes under an ethically and operationally suitable design. Randomization may be appropriate in some business workflows; high-stakes deployments may require additional approvals and carefully staged evaluation. The present paper reports no prospective trial and assigns no invented sample-size requirement.

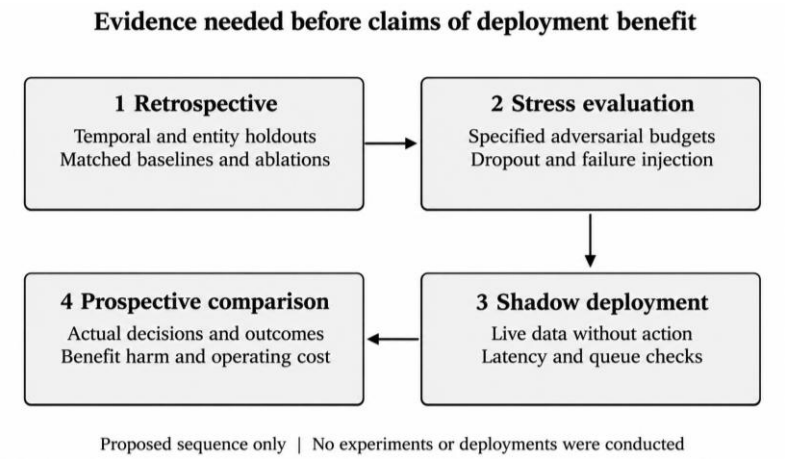


Figure 4. Proposed progression of evaluation evidence. No retrospective experiment, stress evaluation, shadow deployment or prospective study was conducted in this paper.

6. FALSIFIABLE PROPOSITIONS AND RESEARCH AGENDA

The first proposition concerns verification. Under a fixed workload and latency budget, an independent verifier will improve realized decision utility only when the harm it prevents exceeds the costs of false rejection, delay, and additional review. A matched evaluation can reject this proposition for a particular implementation if the verified system fails to outperform the unverified comparator on its prespecified utility endpoint.

The second proposition concerns privacy mechanisms. A protected collaborative system will be operationally preferable to a feasible local-only or centralized alternative only within a measurable region of privacy risk, computation cost, communication reliability, and decision quality. The proposal does not assume that increasing encryption complexity always improves overall utility. Experiments

should identify trade-offs and report configurations that fail to meet operational constraints.

The third proposition concerns digital twins. A simulation-assisted release process will reduce deployment violations only when the simulator represents the failure conditions that determine those violations. The test should compare prospective error patterns with simulated expectations and examine scenarios omitted from simulator development. Agreement on routine cases is insufficient evidence of reliability under disruption.

The fourth proposition concerns adaptation. Bounded updates will provide more stable decision utility than unrestricted updates when delayed labels, poisoning attempts, or abrupt shifts are present, provided that the bounds do not prevent necessary correction. This proposition can fail if an overly restrictive policy preserves an obsolete model. The evaluation should compare update policies rather than merely compare snapshots of their final models.

Table 4. Falsifiable propositions for subsequent empirical work

Proposition	Controlled comparison	Evidence against the proposition
P1 Independent verification	Same predictor and budget, with versus without verification	No utility gain, or gains offset by rejects, delays or workload
P2 Protected collaboration	Local, feasible centralized and protected collaborative baselines	No feasible privacy and utility region under stated constraints
P3 Simulation-assisted release	Matched release processes, with versus without simulator screening	No reduction in prospectively measured deployment violations
P4 Bounded adaptation	Frozen, bounded and unrestricted update policies	Bounds preserve obsolete behavior or reduce decision utility

A research program can test these propositions sequentially. The first study should establish the data and action contract in one domain. The second should isolate the contribution of a verifier or privacy mechanism. A later study can examine transfer to a different organization or application. This sequence avoids presenting architectural breadth as validation and makes negative results useful: a failed component reveals a boundary condition rather than being hidden by an aggregate score.

7. DISCUSSION

7.1 Interpretation of the synthesis

The main result is a structured research design, not evidence that an integrated system has already outperformed existing practice. The corpus connects decision applications with technical controls, but the connections vary in evidential depth. Its strongest use is to formulate requirements that can be tested consistently. This gives the architecture practical value while preserving the distinction between an appealing design and a demonstrated benefit.

The domain differences are substantive. Supply-chain forecasting concerns demand and resource flows; clinical classification concerns diagnostic evidence and patient pathways; cybersecurity concerns strategic adversaries and evolving infrastructure. The common architecture should preserve those differences through separate data contracts, outcome definitions, and action policies. It should not force unrelated tasks into a shared accuracy table.

The proposed decision record also addresses a common organizational difficulty: responsibility becomes fragmented when data engineers, model developers, compliance staff, and operators own different parts of a workflow. A traceable record does not resolve every governance issue, but it makes a particular action reviewable. The relevant question becomes which data, model, rule, and approval produced the action at that time.

7.2 Practical implications

Organizations considering an AI deployment should begin with the decision and its failure consequences. If the action is reversible and low-impact, a lightweight workflow may be sufficient. If it changes access to finance, healthcare, or critical services, the evidence requirements rise. Adding a language model, digital twin, or encrypted aggregation layer should follow a demonstrated need rather than an assumption that a more elaborate system is more trustworthy.

A useful procurement specification would request a locked evaluation protocol, baseline comparisons, data-availability timestamps, subgroup reporting, attack assumptions, measured runtime, and rollback evidence. It would also request negative cases and known limitations. These requirements can be assessed even when competing vendors use different model families. They focus attention on the decision service the organization is purchasing.

7.3 Limitations and publication readiness

The author-bounded corpus is the principal limitation. Its composition cannot support a comprehensive claim about the state of trustworthy AI. Most records were available here as bibliographic entries rather than fully appraised reports. Selective publisher inspection does not substitute for full-text extraction, independent methodological assessment, or a search for contradictory evidence. Accordingly, the synthesis avoids numerical performance aggregation and claims about the prevalence of methodological weaknesses across the studies.

The 16 contextual additions broaden the technical discussion but were selected purposively rather than through a comprehensive search. They are not added to the descriptive denominator, and no claim is made that they form an independent or representative validation sample.

The thematic assignments are interpretive, and some alternatives are reasonable. A healthcare federation paper could be assigned to healthcare rather than privacy; a digital-twin paper could be assigned to infrastructure rather than representation. Releasing the assignments allows readers to test whether the descriptive pattern changes under another coding rule. The year distribution is likewise contingent on the supplied date conventions and should not be interpreted as a publication-growth estimate.

Before journal submission, the working bibliography requires complete publisher verification, including author names, publication status, notices, pagination, and date reconciliation. A broader literature search and independent full-text appraisal would strengthen a review article. Alternatively, implementing one bounded application and following the prospective protocol would support an empirical paper.

8. CONCLUSION

Trustworthy AI decision systems require a continuous connection between available information, model output, permitted action, independent checks, and observed consequences. The corpus provides a bounded map of applications and controls across business, healthcare, and cyber infrastructure. The proposed architecture turns their intersection into explicit design and evaluation requirements. Its central test is whether a system improves a defined decision under realistic constraints while preserving accountability. That test remains empirical. The framework and accompanying materials make it concrete, reproducible, and open to disconfirmation.

REFERENCES

[1] S. M. M. Uddin *et al.*, "AI-powered business intelligence: Enhancing decision-making through predictive analytics

and big data," *J. Posthumanism*, vol. 6, no. 3, pp. 218–237, 2026, doi: https://doi.org/10.63332/joph.v6i3.4083.

[2] P. Chakraborty *et al.*, "Leveraging Artificial Intelligence and Machine Learning for Decision-Making in Business Management: A Comprehensive Analysis," *J. Manag. World*, vol. 2025, no. 2, pp. 46–56, 2025, doi: https://doi.org/10.53935/jomw.v2024i4.867.

[3] N. Das *et al.*, "The strategic impact of business intelligence tools: A review of decision-making and ambidexterity," *Membr. Technol.*, vol. 2025, no. 1, pp. 542–553, 2025, doi: https://doi.org/10.52710/mt.307.

[4] M. A. R. Chy *et al.*, "Understanding the relationship between data governance and business analytics success: A case study of global corporations," *J. Inf. Syst. Eng. Manag.*, vol. 9, no. 4s, pp. 4120–4136, 2024, doi: https://doi.org/10.52783/jisem.v9i4s.14807.

[5] P. Chakraborty *et al.*, "Toward autonomous decision intelligence: Integrating explainable AI and scalable DSS architectures in modern management information systems," *J. Inf. Syst. Eng. Manag.*, vol. 9, no. 4s, pp. 3893–3917, 2024, doi: https://doi.org/10.52783/jisem.v9i4s.14591.

[6] U. Haldar *et al.*, "AI-Driven Business Analytics for Economic Growth Leveraging Machine Learning and MIS for Data-Driven Decision-Making in the U.S. Economy," *J. Posthumanism*, vol. 5, no. 4, pp. 932–957, Apr. 2025, doi: https://doi.org/10.63332/joph.v5i4.1178.

[7] M. Bashir, R. A. M. Rashed, M. Mustafizur, U. S. Saha, S. M. M. Uddin, and H. Imam, "Digital Procurement Transformation Through Intelligent Automation," in *2026 International Conference on Computing Theory and Wireless Communications (ICCTWC)*, 2026, pp. 1–7. doi: https://doi.org/10.1109/ICCTWC68241.2026.11583410.

[8] F. Karim, M. Nazmussakib, R. M. Haque, U. S. Saha, and M. M. Bakhsh, "Data-Driven Operations and Supply Chain Management Using Large Language Models," in *2026 IEEE 18th International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE, 2026, pp. 321–326. doi: https://doi.org/10.1109/CICN70047.2026.11594225.

[9] N. Sayeed, U. S. Saha, F. Karim, M. Mustafizur, S. M. M. Uddin, and M. A. R. Chy, "Artificial Intelligence Models for End-to-End Supply Chain Optimization," in *2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON)*, IEEE, 2026, pp. 1–8. doi: https://doi.org/10.1109/I3CTCON68242.2026.11507407.

[10] M. A. Miah, "An Integrated Big Data and Predictive Analytics Framework for Supply Chain Resilience, International Trade and Economic Growth," *J. Intell. Decis. Mak. Inf. Sci.*, vol. 3, no. 2, pp. 1676–1694, 2026, doi: https://doi.org/10.59543/jidmis.v3.2223.

[11] G. T. Alam *et al.*, "AI-Driven Optimization of Domestic Timber Supply Chains to Enhance U.S. Economic Security," *J. Posthumanism*, vol. 5, no. 1, pp. 1581–1605, Jan. 2025, doi: https://doi.org/10.63332/joph.v5i1.2083.

[12] M. J. Hossain, S. S. Utsab, S. A. Jarir, E. Rozario, Z. A. Tanni, and T. Roy, "Business Analytics for United States Retail Channel Planning Under Structural Change," *Eastasouth J. Inf. Syst. Comput. Sci.*, vol. 3, no. 01, pp. 211–223, Aug. 2025, doi: https://doi.org/10.58812/esiscs.v3i01.1263.

[13] S. Sultana *et al.*, "A Comparative Review of Machine Learning Algorithms in Supermarket Sales Forecasting with Big Data," *J. Ecohumanism*, vol. 3, no. 8, pp. 14457 – 14467, Nov. 2024, doi: https://doi.org/10.62754/joe.v3i8.6762.

[14] U. Haldar *et al.*, "The Role of Artificial Intelligence in Enhancing Customer Experience in Business," *J. Comput. Anal. Appl.*, vol. 33, no. 08, pp. 7138–7155, Dec. 2024, [Online]. Available: https://eudoxuspress.com/index.php/pub/article/view/4068

[15] F. Mahmud *et al.*, "AI-powered workforce analytics forecasting labor market trends and skill gaps for U.S. economic competitiveness," *J. Comput. Sci. Technol. Stud.*, vol. 6, no. 5, pp. 265–277, 2024, doi: https://doi.org/10.32996/jcsts.2024.6.5.21.

[16] F. Karim *et al.*, "Empowering Women through Microcredit: An Analysis of the Grameen Bank Model's Impact on Socio-Economic Outcomes," *Int. J. Comput. Exp. Sci. Eng.*, vol. 10, no. 4, pp. 2578–2589, 2024, doi: https://doi.org/10.22399/ijcesen.4074.

[17] E. Brindha Devi, C. Shalini, V. Andal, B. Sanjeevan, M. Babu, and V. D. Bhaskar, "Artificial intelligence with human resource management and commerce methods in small-scale industrial sectors," in *In 2026 International Conference on Green Technologies and Sustainable Solutions (GTSS)*, IEEE, 2026, pp. 1–7. doi: https://doi.org/10.1109/GTSS68743.2026.11679637.

[18] M. M. T. G. Manik, A. S. M. Saimon, M. S. Islam, M. Moniruzzaman, E. Rozario, and M. E. Hossin, "Big Data Analytics for Credit Risk Assessment," in *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, 2025, pp. 1379–1390. doi: https://doi.org/10.1109/ICMLAS64557.2025.10967667.

[19] M. M. R. Bhuiyan *et al.*, "Economic Implications of Homelessness in the U.S. Applying Advanced Business Analytics to Forecast Costs and Develop Sustainable Solutions," in *2025 5th International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 2025, pp. 1–5. doi: https://doi.org/10.1109/ICECET63943.2025.11472202.

[20] M. E. Hossin *et al.*, "Harnessing business analytics in management information systems to foster sustainable economic growth through smart manufacturing and Industry 4.0," *Educ. Adm. Theory Pract.*, vol. 30, no. 10, pp. 730–739, 2024, doi: https://doi.org/10.53555/kuey.v30i10.9643.

[21] M. E. Hossin *et al.*, "Digital Transformation in the USA Leveraging AI and Business Analytics for IT Project Success in the Post-Pandemic Era," *J. Posthumanism*, vol. 5, no. 4, pp. 958–976, 2025, doi: https://doi.org/10.63332/joph.v5i4.1180.

[22] T. Roy, M. H. Rijvi, M. M. Bakhsh, M. Adnan, M. Nazmussakib, and E. Rozario, "Decision Robustness Under Information Noise: A Machine Learning Perspective on Managerial Decision Systems," in *2026 IEEE 18th International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE, 2026, pp. 327–333. doi: https://doi.org/10.1109/CICN70047.2026.11594307.

[23] F. Karim *et al.*, "Calibrated, cost-sensitive, explainable ensemble learning for customer churn prediction in retail

banking: a leakage-safe pipeline and single-dataset case study," *Front. Artif. Intell.*, vol. 9, p. 1889639, 2026, doi: https://doi.org/10.3389/frai.2026.1889639.

[24] I. Zerine *et al.*, "Explainable churn prediction in telecom with tabular ML five model benchmark and SHAP analysis," *Discov. Artif. Intell.*, vol. 6, no. 1, p. 263, 2026, doi: https://doi.org/10.1007/s44163-026-00983-0.

[25] P. Chakraborty *et al.*, "Trustworthy data lakehouse design using federated learning and blockchain," in *2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT)*, IEEE, 2025, pp. 1–8. doi: https://doi.org/10.1109/ICAFT66710.2025.11453041.

[26] U. Haldar *et al.*, "Blockchain-Driven Access Control and Compliance Auditing Framework for Federated Cloud Service Providers: Architecture, Prototype and Evaluation," in *Proceedings of Fifth International Conference on Computing and Communication Networks*, G.-N. Nguyen, A. Swaroop, and P. Shukla, Eds., Cham: Springer Nature Switzerland, 2026, pp. 464–487. doi: https://doi.org/10.1007/978-3-032-14197-2_41.

[27] S. N. Hasan, "Embedding cybersecurity risk management into information systems governance: A national-scale framework for organizational resilience, economic stability, and critical infrastructure protection," *Int. J. Appl. Math.*, vol. 39, no. 1s, pp. 237–254, 2026, doi: https://doi.org/10.12732/ijam.v39i1s.1623.

[28] K. B. Siddiqa *et al.*, "AI-Driven Project Management Systems: Enhancing IT Project Efficiency Through MIS Integration," in *2024 International Conference on Progressive Innovations in Intelligent Systems and Data Science (ICPIDS)*, IEEE, 2024, pp. 114–119. doi: https://doi.org/10.1109/ICPIDS65698.2024.00027.

[29] S. M. Orthi *et al.*, "AI-augmented risk intelligence in IT project management: An empirical MIS-driven evaluation using machine learning and neural networks," *Int. J. Appl. Math.*, vol. 38, no. 12s, pp. 2869–2890, 2025, doi: https://doi.org/10.12732/ijam.v38i12s.1594.

[30] H. Rahman, K. B. Siddiqa, S. Sultana, I. T. Ahsan, M. M. Anwar, and F. Hossain, "Next-generation software quality assurance: Integrating AI-driven predictive analytics, digital twins, and agile methodologies for transformative research and practice," *J. Comput. Sci. Technol. Stud.*, vol. 7, no. 9, pp. 453–463, 2025, doi: https://doi.org/10.32996/jcsts.2025.7.9.52.

[31] S. M. Orthi *et al.*, "DataOps-oriented big data governance for automated decision pipelines," in *2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT)*, IEEE, 2025, pp. 1–8. doi: https://doi.org/10.1109/ICAFT66710.2025.11452860.

[32] S. Bauskar, R. K. Sahoo, S. S. Boda, H. Singhai, M. M. Bakhsh, and M. Adnan, "Privacy-Aware Big Data Governance Framework Using Blockchain," in *2025 IEEE International Conference on Emerging Trends in Computing and Communication (ETCOM)*, 2025, pp. 1–9. doi: https://doi.org/10.1109/ETCOM66606.2025.11436976.

[33] P. Chakraborty *et al.*, "An Intelligent Cybersecurity Framework for Data Protection in Cloud Computing Environments," *Spectr. Eng. Sci.*, vol. 2, no. 3, pp. 657–677, Oct. 2024, doi: https://doi.org/10.5281/zenodo.21538159.

[34] N. Das *et al.*, "AI-driven threat detection and response framework for protecting U.S. critical infrastructure from cyberattacks," *Int. Cybersecurity Law Rev.*, vol. 7, no. 2, pp. 147–163, 2026, doi: https://doi.org/10.1365/s43439-026-00169-5.

[35] S. N. Hasan *et al.*, "The influence of artificial intelligence on data system security," *Int. J. Comput. Exp. Sci. Eng.*, vol. 11, no. 3, pp. 5589–5598, 2025, doi: https://doi.org/10.22399/ijcesen.3476.

[36] U. S. Saha, H. Imam, M. Mustafizur, A. Roy, A. Al Zaiem, and M. M. T. G. Manik, "Deep learning based malware image classification using binary image representation for secure management information systems," in *2026 Technology and Innovation Conference (TIC)*, IEEE, 2026, pp. 1–6. doi: https://doi.org/10.1109/TIC68483.2026.11702958.

[37] M. K. Tuhin, H. Kaur, H. Imam, A. Hossain, A. Shan-A-Alahi, and P. Chakraborty, "AI-augmented security analytics for real-time threat identification," in *2026 6th International Conference on Intelligent Technologies (CONIT)*, IEEE, 2026, pp. 1–9. doi: https://doi.org/10.1109/CONIT69683.2026.11621644.

[38] S. Sultana *et al.*, "AI-Augmented Big Data Analytics for Real-Time Cyber Attack Detection and Proactive Threat Mitigation," *Int. J. Comput. Exp. Sci. Eng.*, vol. 11, no. 3, pp. 5639–5647, Jul. 2025, doi: https://doi.org/10.22399/ijcesen.3564.

[39] N. Nabi *et al.*, "Continual Learning Pipelines for Detecting Insider Threats Across Corporate Cybersecurity Infrastructures," in *2025 International Conference on Electrical Engineering and Informatics (ICEEI)*, IEEE, 2025, pp. 1–8. doi: https://doi.org/10.1109/ICEEI68459.2025.11330487.

[40] S. N. Hasan *et al.*, "Self-healing cybersecurity systems using RL agents," in *2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT)*, IEEE, 2025, pp. 1–8. doi: https://doi.org/10.1109/ICAFT66710.2025.11452866.

[41] A. Shan-A-Alahi *et al.*, "Deep Learning-Based Threat Prediction and Autonomous Response Mechanisms for Containerized Microservices in Hybrid Cloud Deployments," in *Proceedings of Fifth International Conference on Computing and Communication Networks*, G.-N. Nguyen, A. Swaroop, and P. Shukla, Eds., Cham: Springer Nature Switzerland, 2026, pp. 529–554. doi: https://doi.org/10.1007/978-3-032-21499-7_42.

[42] A. Shan-A-Alahi, M. S. Sikder, H. U. Himel, M. T. Bin Ansar, M. K. Tuhin, and H. Kaur, "Resilient Cybersecurity Architectures for Large-Scale Distributed Systems," in *2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON)*, IEEE, 2026, pp. 1–8. doi: https://doi.org/10.1109/I3CTCON68242.2026.11507768.

[43] M. A. Siam *et al.*, "AI-Driven Cyber Threat Intelligence Systems: A National Framework for Proactive Defense Against Evolving Digital Warfare," *Int. J. Comput. Exp. Sci. Eng.*, vol. 11, no. 3, pp. 6126–6140, 2025, doi: https://doi.org/10.22399/ijcesen.3793.

[44] K. B. Siddiqa *et al.*, "Assessment of Survivability and Importance Analysis for Networks Managing Intricate Traffic Flows," *IEEE Commun. Stand. Mag.*, vol. 10, no. 3, pp. 101–106, 2026, doi:

https://doi.org/10.1109/MCOMSTD.2025.3638981.

[45] M. Raihan *et al.*, "Adversarial robustness mechanism for safeguarding biometric verification across mobile financial applications," in *2025 International Conference on Electrical Engineering and Informatics (ICEEI)*, IEEE, 2025, pp. 1–7. doi: https://doi.org/10.1109/ICEEI68459.2025.11330502.

[46] M. A. Siam *et al.*, "Cybersecure intelligent sensor framework for smart buildings: AI-based intrusion detection and resilience against IoT attacks," *Sensors*, vol. 25, no. 24, p. 7680, 2025, doi: https://doi.org/10.3390/s25247680.

[47] K. S. Mudaliyar and S. S. Kumar, "AEGIS-FL: Auditable Federated Threat Detection for Multi-Tenant Cloud-Edge Systems," *Eastasouth J. Inf. Syst. Comput. Sci.*, vol. 4, no. 01, pp. 101–121, Aug. 2026, doi: https://doi.org/10.58812/esiscs.v4i01.1217.

[48] M. K. Tuhin *et al.*, "Federated Machine Learning for Privacy-Preserving Cyber Threat Intelligence in Global Cloud-Integrated MIS Platforms," *Eastasouth J. Inf. Syst. Comput. Sci.*, vol. 2, no. 02, pp. 284–293, Dec. 2024, doi: https://doi.org/10.58812/esiscs.v2i02.1123.

[49] N. Das *et al.*, "AI-enhanced privacy preservation using homomorphic federated models," in *2025 1st International Conference on Advancement in Futuristic Technologies (ICAFT)*, IEEE, 2025, pp. 1–8. doi: https://doi.org/10.1109/ICAFT66710.2025.11453096.

[50] N. Das *et al.*, "A Privacy-Preserving Federated Intrusion Detection System Leveraging Secure Multi-Party Computation Across Collaborative Cloud Tenants," in *Proceedings of Fifth International Conference on Computing and Communication Networks*, G.-N. Nguyen, A. Swaroop, and P. Shukla, Eds., Cham: Springer Nature Switzerland, 2026, pp. 488–504. doi: https://doi.org/10.1007/978-3-032-14197-2_42.

[51] M. T. Bin Ansar, M. S. Sikder, M. K. Tuhin, N. Das, P. Chakraborty, and H. Rahman, "Resilient MIS Design Using Cryptographic and Cyber-Security Techniques," in *Innovative Computing and Communications*, A. E. Hassanien, S. Anand, A. Jaiswal, and P. Kumar, Eds., Cham: Springer Nature Switzerland, 2026, pp. 317–331. doi: https://doi.org/10.1007/978-3-032-30008-9_24.

[52] A. Roy, E. Hossain, H. U. Himel, H. Imam, R. A. Md Rashed, and H. Kaur, "Privacy-Preserving MIS Using Blockchain, Federated Learning and Zero-Trust Models," in *Innovative Computing and Communications*, A. E. Hassanien, S. Anand, A. Jaiswal, and P. Kumar, Eds., Cham: Springer Nature Switzerland, 2026, pp. 210–223. doi: https://doi.org/10.1007/978-3-032-30008-9_16.

[53] S. M. Orthi *et al.*, "Federated learning with privacy-preserving big data analytics for distributed healthcare systems," *J. Comput. Sci. Technol. Stud.*, vol. 7, no. 8, pp. 269–281, 2025, doi: https://doi.org/10.32996/jcsts.2025.7.8.31.

[54] S. C. Mohonta *et al.*, "Efficient Homomorphic Encryption Techniques for Confidential Big-Data Analytics over Public Cloud Infrastructures: Algorithms, Performance, and Trade-Offs," in *Proceedings of Fifth International Conference on Computing and Communication Networks*, G.-N. Nguyen, A. Swaroop, and P. Shukla, Eds., Cham: Springer Nature Switzerland, 2026, pp. 507–528. doi: https://doi.org/10.1007/978-3-032-21499-7_41.

[55] S. Milmile, P. A. Ebenezer, T. Amitha, A. J. Husain, J. Sivadasan, and M. Babu, "Secure cloud-based medical imaging analytics framework for privacy-aware disease detection using deep learning," *J. Internet Serv. Inf. Secur.*, vol. 16, no. 2, pp. 592–607, 2026, doi: https://doi.org/10.58346/JISIS.2026.I2.037.

[56] M. K. Ahmed *et al.*, "Leveraging big data analytics for personalized cancer treatment: An overview of current approaches and future directions," *J. Eng.*, vol. 2025, no. 1, p. 9928467, 2025, doi: https://doi.org/10.1155/je/9928467.

[57] M. J. Hossain *et al.*, "Data-Driven Predictive Modeling for Early Diagnosis and Risk Stratification of Chronic Diseases," *Eastasouth J. Inf. Syst. Comput. Sci.*, vol. 2, no. 01, pp. 162–172, Aug. 2024, doi: https://doi.org/10.58812/esiscs.v2i01.1220.

[58] M. H. Rijvi *et al.*, "A compact texture frequency spatial gating network for oral cancer classification from clinical oral images," *Sci. Rep.*, vol. 16, no. 1, p. 31071, 2026, doi: https://doi.org/10.1038/s41598-026-59480-0.

[59] E. Hossain *et al.*, "Cardiovascular Disease Burden and Clinical Associations Among US Adults Using Survey-Weighted Data Analytics," *Eur. J. Med. Heal. Res.*, vol. 2, no. 3, pp. 269–281, Jun. 2024, doi: https://doi.org/10.59324/ejmhr.2024.2(3).31.

[60] M. I. Faruk *et al.*, "Global health intelligence platforms data science infrastructure for real-time epidemiological surveillance and pandemic forecasting," *Spectr. Eng. Sci.*, vol. 2, no. 5, pp. 749–761, Dec. 2024, [Online]. Available: https://thesesjournal.com/index.php/1/article/view/3425

[61] E. Rozario *et al.*, "Optimizing epidemic response through AI-based predictions: Machine learning approaches to forecasting and healthcare resource distribution," in *In 2025 5th international conference on electrical, computer and energy technologies (ICECET)*, IEEE, 2025, pp. 1–6. doi: https://doi.org/10.1109/ICECET63943.2025.11472546.

[62] M. Manchanda, R. S. Tomar, S. Shankar, K. Adilakshmi, E. Rozario, and S. Pareyani, "Artificial intelligence–enabled healthcare ecosystems for predictive, preventive, and personalized medicine," in *In 2026 5th OPJU international technology conference (OTCON) on smart computing for innovation and advancement in industry 5.0*, IEEE, 2026, pp. 1–6. doi: https://doi.org/10.1109/OTCON69332.2026.11629957.

[63] M. H. Rahman, M. T. B. Ansar, S. Hossain, U. S. Saha, H. Imam, and I. T. Ahsan, "AI-powered QA in healthcare software: Leveraging predictive analytics and digital twins for safe, cost-effective, and agile medical systems," *J. Comput. Sci. Technol. Stud.*, vol. 7, no. 9, pp. 619–628, 2025, doi: https://doi.org/10.32996/jcsts.2025.4.1.70.

[64] M. H. Rahman *et al.*, "Integrating artificial intelligence and data science for breakthroughs in drug development and genetic biomarker discovery," *J. Posthumanism*, vol. 5, no. 8, pp. 257–271, 2025, doi: https://doi.org/10.63332/joph.v5i8.3157.

[65] M. Babu and P. K. Yadav, "Attention-guided artificial intelligence multi-agent architecture for explainable and resilient skin disease classification," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 63s, pp. 2057–2069, 2025, [Online]. Available:

https://jisem-journal.com/index.php/journal/article/view/15382

[66] M. Babu and P. K. Yadav, "Cooperative artificial intelligence multi-agent architecture for accurate and an adaptive lung tumor segmentation in is computed tomography images," *Commun. Appl. Nonlinear Anal.*, vol. 32, no. 10s, pp. 5509–5525, 2025, [Online]. Available: https://internationalpubls.com/index.php/cana/article/view/7512

[67] R. Anzum, R. Gentyala, N. Mukhi, N. Karri, P. Krishnan, and E. Rozario, "Hierarchical Learning Architectures for Complex Data Representation," in *2026 International Conference on Emerging in Computing Technologies (INCECT)*, 2026, pp. 1–8. doi: https://doi.org/10.1109/INCECT72370.2026.11742577.

[68] M. E. Hossin *et al.*, "Adaptive Multi-Modal AI Agents: Leveraging Large Language Models for Dynamic Real-World Environments," in *Proceedings of International Conference on Computational Technologies for Research in Data Analytics*, N. G. Nhu, A. Jaiswal, and T. M. Thanh, Eds., Cham: Springer Nature Switzerland, 2026, pp. 106–119. doi: https://doi.org/10.1007/978-3-032-25694-2_10.

[69] E. Rozario, M. Nazmussakib, G. T. Alam, T. Roy, M. Adnan, and M. I. Faruk, "Dynamic Big Data Summarisation Using Transformer-Based Generative Models," in *2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON)*, IEEE, 2026, pp. 1–8. doi: https://doi.org/10.1109/I3CTCON68242.2026.11507387.

[70] M. Babu, *Enterprise artificial intelligence: Agentic AI, large language models, and intelligent systems*. RG International Publication, 2026. doi: https://doi.org/10.5281/zenodo.21909795

[71] M. S. A. Sourav *et al.*, "Grounding language models with deterministic verifiers for fraud detection," *Discov. Artif. Intell.*, vol. 6, no. 1, p. 1403, 2026, doi: https://doi.org/10.1007/s44163-026-02355-0.

[72] E. Hossain *et al.*, "AI-Driven Trust and Bio-Inspired Optimization Integration in Standardized Digital Twin Frameworks for Next-Generation," *IEEE Commun. Stand. Mag.*, pp. 1–6, 2026, doi: https://doi.org/10.1109/MCOMSTD.2026.3688125.

[73] R. Dandigam, M. Patel, C. S. Pareek, S. M. Faysal, G. T. Alam, and M. A. Siam, "Human-in-the-Loop Learning for Improving Language Model Reliability," in *2026 6th International Conference on Intelligent Technologies (CONIT)*, 2026, pp. 1–8. doi: https://doi.org/10.1109/CONIT69683.2026.11620666.

[74] S. R. Varanasi, S. S. S. Valiveti, M. Adnan, M. I. Faruk, M. J. Hossain, and M. M. T. G. Manik, "Cross-Domain standardization and secure edge intelligence for Real-Time digital twin deployments in Next-Generation communication systems," *IEEE Commun. Stand. Mag.*, vol. 10, no. 3, pp. 107–112, 2026, doi: https://doi.org/10.1109/MCOMSTD.2026.3662187.

[75] C. R. Barikdar *et al.*, "MIS frameworks for monitoring and enhancing U.S. energy infrastructure resilience," *J. Posthumanism*, vol. 5, no. 5, pp. 4327–4342, 2025, doi: https://doi.org/10.63332/joph.v5i5.1907.

[76] M. A. Miah *et al.*, "Big data analytics for enhancing coal-based energy production amidst AI infrastructure growth," *J. Posthumanism*, vol. 5, no. 5, pp. 5061–5080, 2025, doi: https://doi.org/10.63332/joph.v5i5.2087.

[77] M. H. Rahman *et al.*, "Scalable AI Models for Climate Change Mitigation Using Multisource Geospatial Big Data," *J. Comput. Anal. Appl.*, vol. 33, no. 08, pp. 5836–5856, Dec. 2024, [Online]. Available: https://eudoxuspress.com/index.php/pub/article/view/3431

[78] M. I. Faruk *et al.*, "Sustainable Computing and Communication Technologies for Digital Twin–Based Energy Optimization in Smart Infrastructure Systems," *Internet Technol. Lett.*, vol. 9, no. 5, p. e70351, Sep. 2026, doi: https://doi.org/10.1002/itl2.70351.

[79] M. Babu, *Enterprise multi-cloud architecture: Design, security, automation, and best practices*. RG International Publication, 2026.