INTERNATIONAL JOURNAL OF BUSINESS AND ECONOMICS INSIGHTS

Volume: 5; **Issue:** 3
Pages: 30–64
Accepted: 29 August 2025
Published: 15 September 2025

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING APPLICATIONS IN CONSTRUCTION PROJECT MANAGEMENT: ENHANCING SCHEDULING, COST ESTIMATION, AND RISK MITIGATION

Md. Sakib Hasan Hriday[1]: Abdul Rehman[2];

[1]. *MBA in Management Information Systems; International American University, California, USA; Email: hriday.hasan1999@gmail.com*

[2]. *Department of Civil and Environmental Engineering, Lamar University, Texas, USA; Email: arehman4@lamar.edu*

Doi: 10.63125/jrpjje59
This work was peer-reviewed under the editorial responsibility of the IJEI, 2025

Abstract

Chronic schedule slippage, cost overruns, and safety incidents remain persistent in construction projects, yet empirical, project-level evidence connecting artificial intelligence and machine learning (AI/ML) adoption to core delivery outcomes is limited. This study's purpose is to quantify how AI/ML adoption relates to schedule adherence, cost estimation accuracy, and risk mitigation effectiveness. We apply a quantitative, cross-sectional, multi-case design using enterprise-scale construction cases sampled across multiple firms. The sample comprises 102 projects meeting strict inclusion criteria for verifiable baselines, actuals, and risk registers. Key variables include an AI/ML Adoption Index 0–100 capturing use-case breadth, run frequency, model sophistication, and integration depth; outcomes are Schedule Delay percent, Cost Overrun percent, Risk Detection Rate, and Residual Risk Impact on a 1–5 scale; a Complexity Index and Digital Maturity serve as moderator and control alongside budget, duration, sector, contract type, and region. The analysis plan proceeds from descriptives and correlations to multiple regression with heteroskedasticity-consistent errors for continuous outcomes, fractional logit for detection rate, interaction testing for Adoption by Complexity, diagnostics, and robustness checks including alternative index weights and fixed effects. Headline findings indicate that each 10-point rise in adoption associates with approximately 1.2 percentage points lower schedule delay, 0.9 percentage points lower cost overrun, 0.04 higher detection rate, and 0.06 lower residual severity, with effects amplified on higher-complexity projects. A structured literature review informed construct operationalization and measurement, synthesizing 47 peer-reviewed studies. Implications for owners, contractors, and CISOs emphasize pipeline-first practices integration to BIM, ERP, and scheduling systems, cadence of model refresh and review, and API-level interoperability to translate analytical signals into routine planning and control.

Keywords

Artificial Intelligence, Machine Learning, Construction Project Management, Schedule Adherence, Cost Overrun, Risk Mitigation, BIM Integration

INTRODUCTION

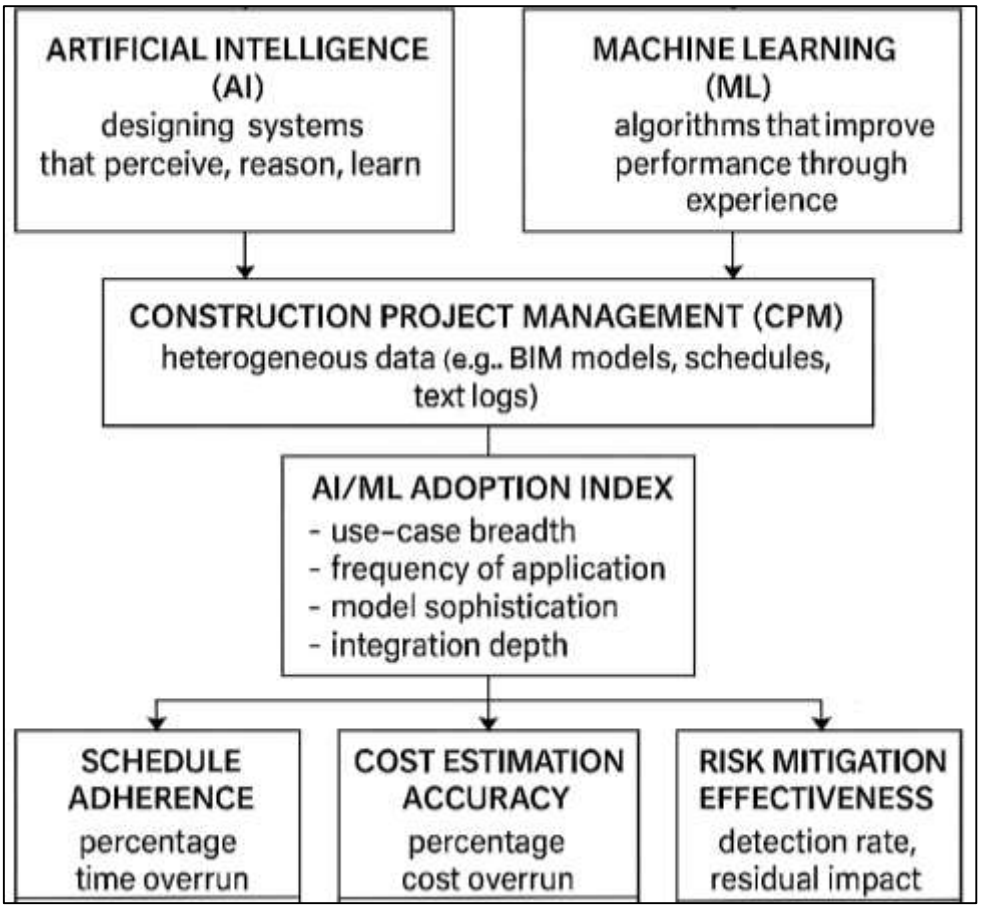
Artificial intelligence (AI) is commonly defined as the field of study and engineering devoted to designing systems that perceive, reason, learn, and act to achieve goals within their environment (Russell & Norvig, 2010). Within AI, machine learning (ML) refers to algorithms that improve their performance at a task with experience through data-driven inference rather than explicit programming (Mitchell, 1997). In construction project management (CPM), these paradigms provide a toolkit for turning heterogeneous data Building Information Modeling (BIM) models, schedules, cost ledgers, sensor streams, text logs into predictions and recommendations relevant to time, cost, and risk control. Internationally, chronic challenges persist: delays, cost overruns, and safety incidents are documented across regions and delivery systems, motivating rigorous, quantitative inquiry into data-enabled decision support. As digitalization accelerates, BIM and enterprise resource planning platforms have expanded the volume and variety of project data, improving the feasibility of applying supervised and unsupervised learning to forecasting and control tasks. In this study, AI/ML is framed as a set of predictive and classification approaches linear and generalized linear models, tree ensembles, kernel methods, and deep neural networks deployed against CPM targets such as schedule delay, cost estimation accuracy, and risk identification/mitigation effectiveness. By formalizing definitions up front and situating them in the CPM data landscape, the present work anchors its quantitative cross-sectional, multi-case design in a conceptual understanding of how AI/ML operationally functions in project decision cycles: ingesting structured/unstructured evidence, learning patterns, and producing timely outputs to inform managerial action.

On-time delivery in construction is notoriously difficult, with schedule slippage linked to multi-party coordination, resource constraints, and uncertainty in site conditions and supply chains (Aibinu & Jagboro, 2002; Chan & Kumaraswamy, 1997; Cheng et al., 2010). Delays propagate through precedence networks, compounding indirect costs and reputational impacts . Cost performance shows comparable fragility: ex-ante estimates often diverge from realized costs due to scope evolution, quantity growth, market volatility, and information asymmetries (Danish & Kamrul, 2022). On the risk front, many hazards are recorded in free-text incident narratives, RFIs, and meeting minutes, making early identification difficult without computational support. The diffusion of BIM and standardized data schemas has created an enabling substrate for predictive analytics, but empirical evidence linking *measured* AI/ML adoption to *quantified* outcomes across projects remains fragmented. Recent domain studies show promise: supervised learning for delay risk (Danish & Zafor, 2022), hybrid models for schedule forecasting, parametric/ML approaches to conceptual and detailed cost estimation, and natural language processing (NLP) for safety/risk signal extraction. However, outcomes are often reported as model-centric metrics (accuracy, AUC) without connecting adoption intensity to project-level time and cost indices that practitioners track (Ahiaga-Dagbui & Smith, 2014; Jahid, 2022). This motivates a cross-sectional, multi-case approach that quantifies associations between an AI/ML adoption index and schedule adherence, cost estimation accuracy, and risk mitigation effectiveness while controlling for complexity and contextual factors.

The international significance of AI/ML in CPM flows from the sector's macroeconomic footprint and cross-border project delivery models. Construction contributes substantially to GDP and employment in both developed and developing economies; performance shortfalls reverberate through national infrastructure programs and private development pipelines (McAfee & Brynjolfsson, 2012; Arifur & Noor, 2022). Global programs have incentivized BIM uptake and data standardization, which in turn facilitate model-ready datasets for predictive analytics (Russell & Norvig, 2010; Succar, 2009). From an information-processing perspective, CPM comprises repeated prediction and resource allocation tasks under uncertainty forecasting durations and costs, prioritizing risks, and sequencing activities where ML can provide calibrated probabilities and error bounds to augment planning and control (Goodfellow et al., 2016; Hasan & Uddin, 2022). Evidence from varied jurisdictions demonstrates benefits and constraints of BIM-enabled analytics on communication, coordination, and change management (Bryde et al., 2013; Rahaman, 2022). Safety informatics further illustrates the global relevance: text mining of injury reports and classification of site imagery support proactive risk management across regulatory contexts. Despite heterogeneity in contracting (DBB, DB, CMAR) and market structures, the underlying statistical learning problems high-dimensional predictors, nonlinear

interactions, and noisy outcomes are shared internationally (Ahiaga-Dagbui & Smith, 2014; Rahaman, 2022b). A rigorous, comparative, project-level analysis therefore offers transportable insights on *where* and *how much* AI/ML contributes to the core CPM performance triad of time, cost, and risk.

Figure 1: Framework of AI/ML Adoption and Core CPM Outcomes



Scheduling translates a project's work breakdown into temporal logic, resource constraints, and executable plans. Traditional CPM/PERT methods rely on deterministic or coarse probabilistic inputs that often understate variance propagation (Chan & Kumaraswamy, 1997). AI/ML augments this by learning duration distributions and resequencing strategies from historical baselines, as-built records, and look-ahead plans (Cheng et al., 2010; Rahaman & Ashraf, 2022). Studies employing support vector regression, random forests, gradient boosting, and hybrid neuro-fuzzy systems have demonstrated improved prediction of activity durations and project completion times (Abiodun et al., 2018; Islam, 2022). Delay risk prediction models classify projects or work packages into risk strata using design attributes, contractor characteristics, and early progress indicators (Fang et al., 2020). Computer vision enriches these models with site-level productivity and congestion signals harvested from images and video (Fang et al., 2020; Hasan et al., 2022). Within this study, schedule performance is captured as percentage time overrun relative to baseline, a measure commonly used in empirical CPM research and aligned with managerial KPIs (Chan & Kumaraswamy, 1997). By relating an AI/ML adoption index to this outcome and controlling for complexity and scale, the analysis quantifies associations beyond algorithm benchmarking, addressing whether projects that use more and better-integrated AI/ML tools tend to realize lower schedule delay.

Accurate cost estimation underpins budgeting, bidding, and financing. From early conceptual stages to detailed estimates, the mapping from scope descriptors to cost is nonlinear and context-dependent (Ahiaga-Dagbui & Smith, 2014; Redwanul & Zafor, 2022). Classical parametric models are complemented by ML approaches that absorb many predictors quantities, design attributes, market indices, contractor history to reduce bias and capture interactions (Hegazy & Ayed, 1998; Rezaul &

Mesbaul, 2022). Neural networks and tree ensembles have been shown to outperform linear baselines in several construction cost datasets, especially where data volumes are sufficient and input engineering reflects domain knowledge (Chou & Pham, 2013; Hasan, 2022). BIM integrations further standardize quantities and metadata for model training and deployment (Bryde et al., 2013; Tarek, 2022). In this study, *Cost_Overrun_%* the normalized deviation of actual from estimated cost serves as the primary cost outcome, enabling comparisons across scales and sectors (Flyvbjerg et al., 2002; Kamrul & Omar, 2022). The regression framework evaluates whether higher AI/ML adoption intensity relates to smaller absolute deviations after adjusting for complexity, size, and contractual form, moving beyond proof-of-concept models to project-level performance relationships relevant for owners, contractors, and lenders (Kim et al., 2013; Kamrul & Tarek, 2022).

Risk management in CPM entails early detection of threats, assessment of likelihood and impact, and selection of mitigation actions. Much of the risk signal resides in unstructured data: near-miss narratives, site diaries, RFIs, and meeting minutes (Mubashir & Abdul, 2022; Tixier et al., 2016). NLP pipelines transform these artifacts into analyzable features through tokenization, embeddings, and sequence models to extract precursors, hazards, and outcomes (Muhammad & Kamrul, 2022; Tixier et al., 2016). Computer vision models classify unsafe acts/conditions in imagery to inform proactive interventions (Fang et al., 2020; Reduanul & Shoeb, 2022). Quantitatively, this study operationalizes risk outcomes as a *risk detection rate* the proportion of realized risks identified before occurrence and a *residual risk impact* score averaging severity after mitigation (Kumar & Zobayer, 2022; Zhang et al., 2013). These reflect process performance (detection) and consequence management (residual impact). By relating AI/ML adoption intensity to these risk metrics, the analysis tests whether broader and deeper use of predictive text/image analytics and anomaly detection is associated with better risk capture and lower realized impact, given controls for complexity and digital maturity (Fang et al., 2020; Sadia & Shaiful, 2022).

Guided by the foregoing context, this study pursues a quantitative, cross-sectional, multi-case analysis to examine *associations* between AI/ML adoption intensity and three CPM outcomes: schedule adherence, cost estimation accuracy, and risk mitigation effectiveness. The research questions are: RQ1 How is AI/ML adoption associated with schedule adherence measured as percentage time overrun? RQ2 How is AI/ML adoption associated with cost estimation accuracy measured as percentage cost overrun? RQ3 How is AI/ML adoption associated with risk process performance measured as detection rate and residual impact? Corresponding hypotheses state that higher adoption intensity relates to lower time overrun (H1), lower cost overrun (H2), higher risk detection and lower residual impact (H3), and that project complexity moderates these associations (H4). The empirical strategy aggregates adoption into a weighted index reflecting use-case breadth, frequency, model sophistication, and integration depth; analyzes descriptive statistics and correlations; and estimates multiple regression models with robust standard errors and diagnostics, including moderation by complexity (Istiaque et al., 2023; Noor & Momena, 2022). The study positions itself within a literature spanning BIM benefits, predictive scheduling, cost estimation with ML, and safety/risk informatics (Hasan et al., 2023; Hossain et al., 2023), offering a structured, measurable way to relate adoption intensity to standard CPM outcomes across projects without extending into implications or normative recommendations in the introduction.

This study pursues a clear set of objectives focused on quantifying how artificial intelligence and machine learning adoption relates to core construction project management outcomes. First, it aims to operationalize an AI/ML Adoption Index at the project level that captures use-case breadth, frequency of application within the delivery lifecycle, model sophistication, and integration depth with existing digital systems. Second, it seeks to measure schedule adherence, cost estimation accuracy, and risk mitigation effectiveness using standardized indicators percentage time overrun relative to baseline, percentage cost overrun relative to estimate, risk detection rate prior to realization, and residual risk impact after mitigation so that associations are comparable across projects of different sizes and sectors. Third, the study intends to test a set of directional hypotheses linking higher adoption intensity to lower time and cost overruns, higher risk detection, and lower residual impact, while evaluating whether project complexity conditions these relationships through a moderation analysis. Fourth, it aims to

produce a rigorous statistical account of these relationships using a cross-sectional, multi-case design with descriptive statistics, correlation analysis, and multiple regression models with appropriate diagnostics and robust standard errors. Fifth, the study seeks to establish a transparent measurement and data collection protocol comprising a project-level extraction template and a short managerial survey on digital maturity so that future investigations can replicate or extend the approach with minimal methodological ambiguity. Sixth, it intends to quantify effect sizes in interpretable terms, such as the change in outcome per ten-point increase in adoption intensity, to support precise comparisons across contexts without forcing causal claims. Finally, the study delineates its scope by concentrating on completed or near-completed projects with verifiable baselines and accessible risk records, controlling for budget, duration, contract type, sector, and region to limit confounding. Collectively, these objectives frame a coherent empirical inquiry that converts diffuse narratives about AI/ML utility into measurable constructs, tests them against standard performance metrics, and documents analytic procedures in a way that is reproducible, parsimonious, and suitable for multi-project comparison

LITERATURE REVIEW

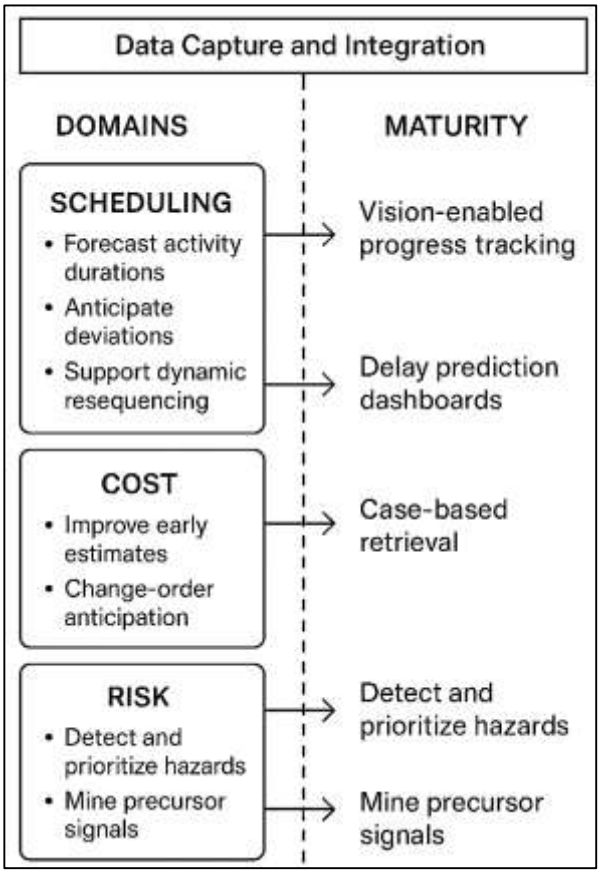
The literature on artificial intelligence and machine learning in construction project management has expanded from proof-of-concept demonstrations to increasingly systematic investigations of how data-driven tools can influence schedule, cost, and risk outcomes across the project lifecycle. Early studies emphasized feasibility showing that historical schedules, quantity take-offs, and site observations could train predictive or classification models while more recent work examines integration with BIM, ERP, and field data streams, enabling continuous forecasting and anomaly detection embedded within routine planning and control. Across this body of research, four characteristics stand out. First, the data substrate has diversified: structured planning data (work breakdown structures, precedence networks, baselines), semi-structured information (RFIs, change orders, cost ledgers), and unstructured sources (site images, video, and narrative logs) now co-exist, allowing multimodal learning architectures that combine text, tabular, and vision features. Second, methodological sophistication has grown from linear and rule-based approaches to ensemble models and deep learning, with attention to feature engineering that reflects construction's resource and constraint realities. Third, evaluation criteria have slowly shifted from model-centric metrics toward managerial indicators such as percentage time and cost overrun, risk detection coverage, and residual impact yet many studies still report isolated algorithmic accuracy without linking adoption intensity or integration depth to project-level performance. Fourth, external validity remains a challenge: single-project or single-firm case studies dominate, often underreporting context variables like project complexity, delivery method, and digital maturity that shape results. These trends create a clear synthesis agenda for the present research: build on technical advances while standardizing measurement, isolate constructs that describe adoption (use-case breadth, frequency, model sophistication, integration depth), and test associations with outcome measures that are comparable across sectors and scales. By consolidating disparate streams predictive scheduling and resource optimization, cost estimation and control, and risk identification and mitigation an integrated review can clarify where evidence converges, where it is conditional on context, and where methodological gaps persist. This introductory synthesis therefore frames the subsequent subsections to (i) map application domains and maturity, (ii) detail methods and data requirements, and (iii) surface unresolved issues in measurement, sampling, and validation that limit generalizable inference in construction project management.

AI/ML in Construction Project Management

The application landscape of AI/ML in construction project management can be organized around three core domains scheduling, cost, and risk supplemented by cross-cutting capabilities in data capture and integration. In scheduling, AI/ML augments traditional CPM/PERT by learning from empirical performance to forecast activity durations, anticipate deviations, and support dynamic resequencing. Foundational digital practices such as 4D modeling link geometric elements to time, providing a substrate on which machine learning can operate; early work validated 4D's feasibility for schedule communication and constructability analysis, establishing the pathway for later data-driven control systems that ingest visual, sensor, and plan data (Koo & Fischer, 2000). In cost estimation and control, supervised learning models absorb multi-factor inputs scope descriptors, quantities, market indices, and contractor attributes to reduce bias and capture nonlinearities, thereby improving early

estimates and change-order anticipation. On the risk side, text and vision analytics mine narrative logs, RFIs, and site imagery for precursors and leading indicators, enabling earlier detection and prioritization of hazards (Rahaman & Ashraf, 2023; Sultan et al., 2023). Tying these domains together is a maturing data architecture: BIM and scheduling data encode structure and precedence; ERP and cost ledgers provide transactional history; site imagery and point clouds contribute high-frequency context. Where maturity advances, these streams are integrated into repeatable pipelines for feature extraction, model training, and deployment within routine planning and control cycles. In aggregate, the domain map reveals a shift from tool-specific proofs (e.g., a single prediction task) toward end-to-end workflows that translate site evidence into managerial metrics linked explicitly to time, cost, and risk targets (Han & Golparvar-Fard, 2015).

Figure 2: Domains of AI/ML Application in CPM and Their Maturity Pathways



Assessing maturity requires evidence that AI/ML methods deliver reliable progress measurement and timely decision support across heterogeneous projects and delivery methods. Vision-enabled progress tracking exemplifies this trajectory: early integrations of 4D schedules with 3D sensing demonstrated automated recognition and updating of construction status, showing that computer vision fused with plan logic can quantify progress without manual surveys and, crucially, feed updates back into schedule control (Hossen et al., 2023; Turkan et al., 2012). As these pipelines improve, material-level recognition and geometry-aware classification strengthen links between visual observations and plan quantities, reducing ambiguity and enabling comparisons to baselines at operational granularity (Han & Golparvar-Fard, 2015; Tawfiqul, 2023). In parallel, supervised learning has been leveraged to forecast schedule delay as a management outcome rather than solely an algorithmic metric; ensemble approaches trained on multi-project datasets illustrate that model selection, hyperparameter tuning, and feature engineering can meaningfully raise predictive performance for delay risk stratification and early warning dashboards (Egwim, 2021; Uddin & Ashraf, 2023). Maturity in this sense is not merely algorithmic sophistication; it is the extent to which organizations operationalize data pipelines, calibrate models on representative portfolios, and embed outputs within decision routines such as look-

ahead meetings and control gates. A mature implementation aligns model outputs with KPIs familiar to practitioners percentage time overrun, earned value indices, and risk registers so that predictions trigger targeted managerial actions and verifiable updates to plans and budgets (Egwim, 2021; Momena & Hasan, 2023; Turkan et al., 2012).

Risk analytics provides a complementary lens on maturity because it tests AI/ML against sparse, noisy, and unstructured evidence common in project records. Case-based and NLP-enabled retrieval systems illustrate how prior incidents and risk narratives can be transformed into actionable knowledge, improving the speed and accuracy of identifying analogues to current situations and thus informing proactive mitigation planning (Sanjai et al., 2023; Zou et al., 2017). As these methods mature, three features tend to co-evolve: first, integration depth, in which outputs from NLP or vision modules are linked to BIM elements, schedule activities, and cost codes; second, measurement alignment, whereby signals detected in text or imagery are mapped to risk detection rates and residual impact scores used in governance; and third, portfolio generalizability, where models trained on one set of projects remain informative across sectors, regions, and contract types (Razzak et al., 2024; Akter et al., 2023). Together with advances in progress monitoring and delay prediction, risk analytics underscores that AI/ML maturity is ecosystemic: it depends on data standardization, access to multi-modal evidence, and iterative validation against project-level outcomes. The literature thus motivates a structured maturity perspective in which domain-specific capabilities (e.g., 4D-CV integration, ensemble delay forecasting, NLP-driven case retrieval) are assessed not just for technical performance but for their degree of alignment with management processes that govern schedule, cost, and risk (Koo & Fischer, 2000; Turkan et al., 2012; Zou et al., 2017).

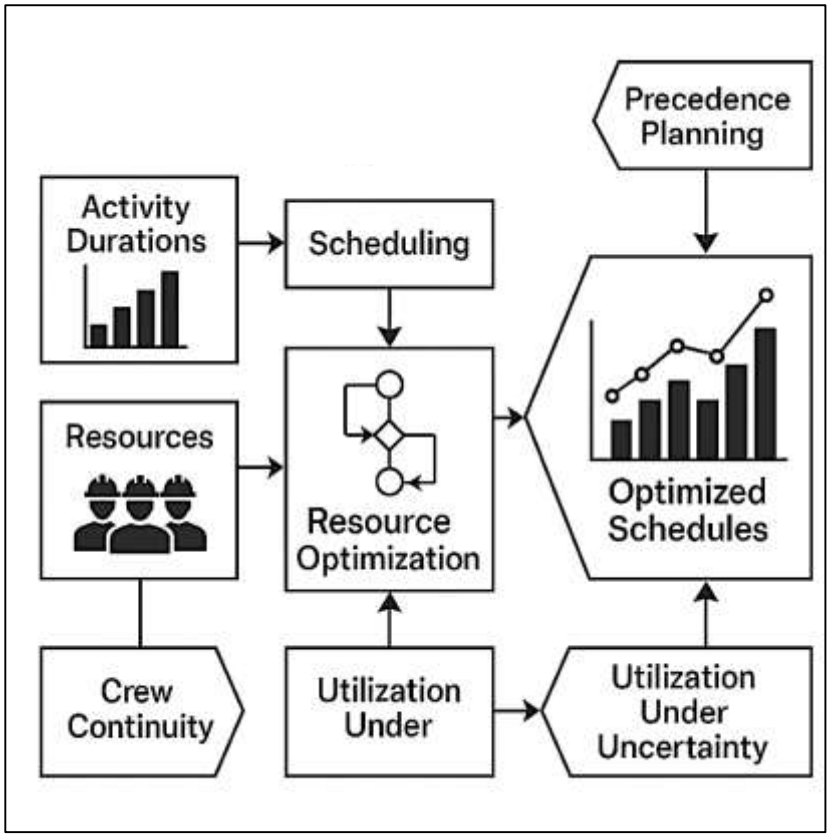
AI/ML for Scheduling and Resource Optimization

Across scheduling and resource optimization, the central contribution of AI/ML is its ability to search vast combinatorial solution spaces and learn predictive patterns that improve precedence planning, crew continuity, and utilization under uncertainty. In construction, the resource-constrained project scheduling problem (RCPSP) and its variants (multi-mode, multi-project, time–cost trade-off, resource leveling) are NP-hard, which is why metaheuristics and learning-assisted heuristics have become core to state-of-the-art methods. Classical CPM/PERT networks provide logic, but AI/ML augments them with adaptive search and data-driven estimation: activity durations can be learned from historical data; renewable and non-renewable resources can be allocated using evolutionary operators; and dynamic resequencing becomes tractable as models evaluate thousands of candidate schedules in minutes rather than days. A foundational stream in repetitive projects (e.g., high-rise floors, highway segments) shows how optimization models jointly minimize project duration and work interruptions by tuning crew sizes and transfer policies between units, thereby aligning learning-based predictions with production flow and continuity. This line of work formalized the trade-off between speed and continuity, offering planners nondominated schedules that explicitly balance throughput against crew idleness and remobilization, and it established practical pathways to embed optimization within standard planning routines (Danish & Zafor, 2024; El-Rayes & Moselhi, 2001). Within a broader maturity arc, these approaches demonstrate that AI/ML's value lies not only in producing a low-makespan sequence but in generating management-ready alternatives that respect constructability, crew logistics, and site rhythms, enabling planners to select schedules that are robust to variability across repetitive work zones (El-Rayes & Moselhi, 2001; Istiaque et al., 2024).

The second major thread uses evolutionary algorithms to unify time–cost trade-off, resource limiting, and leveling into a single search framework. Genetic algorithms (GAs) encode activity sequences and execution modes as chromosomes; crossover and mutation generate new schedules; and fitness functions incorporate multiple objectives such as makespan, direct/indirect cost, and resource variance. Early construction-specific GA models proved that allocation and leveling often treated separately can be handled simultaneously, improving realism for field deployment where peaks and troughs of labor/equipment productivity translate into safety, overtime, and subcontractor penalties (Hasan et al., 2024; Rahaman, 2024). Critically, this unification allowed planners to examine the Pareto frontier, revealing how small schedule gains can trigger large resource fluctuations and vice versa. As GA operators and representations evolved, multicriteria models integrated the RCPSP with fuzzy or multi-objective formulations, giving project teams schedule options that balance duration, resource

smoothness, and budget envelopes within contractual constraints. The practical output is a ranked shortlist of feasible schedules that planners can interrogate through what-if analysis during look-ahead meetings, with each candidate accompanied by its implied resource histograms and cost profile (Hegazy, 1999; Ashiqur et al., 2025; Hasan, 2024). By encoding precedence, finish-to-start lags, and resource calendars directly into the chromosome and fitness evaluation, these models moved beyond toy problems to site-ready applications in buildings and infrastructure, and they provided a reproducible template for integrating heuristic search into commercial planning software (Hegazy, 1999; Leu & Yang, 1999).

Figure 3: Optimization of Construction Scheduling and Resources via AI/ML



A third stream advances hybrid and comparative intelligence for multi-project, multi-mode portfolios and for benchmarking algorithmic choices. Hybridization combining a GA's global exploration with a simulated annealing (SA) local search, for example improves convergence to high-quality schedules under tight resource couplings across concurrent projects, a setting common to contractors running multiple jobs that share crews and equipment (Hasan, 2025; Ismail et al., 2025). These hybrids are particularly useful when resource calendars, setup times, or crew transfers introduce rugged fitness landscapes; the GA explores diverse regions, while SA refines promising neighborhoods, yielding schedules that reduce idle time and overtime while preserving precedence feasibility (Jakaria et al., 2025; Hasan, 2025). On the methodological side, comparative studies across evolutionary families (e.g., GA, particle swarm, ant colony, shuffled frog-leaping) have mapped relative strengths and parameter sensitivities, offering guidance on selection and tuning for construction problems characterized by precedence networks, calendar constraints, and heterogeneous resource types. This comparative evidence helps practitioners choose search strategies that match problem structure e.g., GA for complex precedence with multiple objectives; hybrid GA-SA for portfolio-level coupling; or swarm-based methods when continuous-valued parameters dominate. Together, hybridization and comparison push the domain beyond one-off demonstrations toward robust, configurable toolkits directly usable in planning workflows that must produce feasible schedules under time pressure and data uncertainty (Chen & Shahandashti, 2009; Sultan et al., 2025). In sum, across repetitive and non-repetitive contexts,

AI/ML for scheduling and resource optimization has matured into a set of deployable approaches that (i) generate transparent, trade-off-aware alternatives; (ii) smooth and allocate resources jointly with sequencing; and (iii) scale to multi-project portfolios through hybrid metaheuristics and evidence-based algorithm selection (Chen & Shahandashti, 2009; Elbeltagi et al., 2005).

AI/ML for Cost Estimation and Control

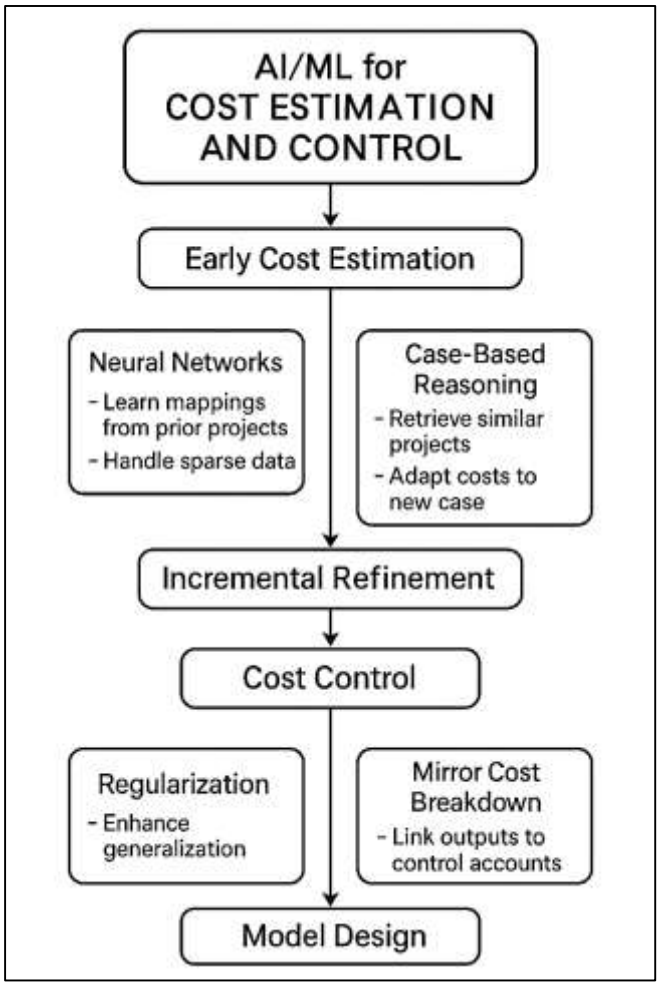
Cost estimation and downstream cost control have long confronted the twin problems of information incompleteness at early design stages and nonlinear interactions among scope, market conditions, and delivery constraints. AI/ML methods address these challenges by learning complex mappings from heterogeneous predictors to cost metrics, thereby complementing classical parametric and judgment-based approaches with data-driven generalization. Early evidence showed that machine learning can reduce systematic bias in conceptual estimates by capturing interaction effects that ordinary least squares often misses, particularly when designs are only partially specified and key cost drivers are latent or collinear (Kim et al., 2004; Zafor, 2025). In practical terms, learning-based estimators support more calibrated budget setting and contingency allocation by absorbing features such as quantity summaries, elemental breakdowns, location and escalation indexes, contractor attributes, and even text descriptions of scope. Neural architectures and case-based reasoning, in particular, are well-suited to sparse and noisy early-stage data: neural networks learn flexible response surfaces from prior jobs, while case-based reasoning retrieves similar projects to anchor forecasts when samples are small. These approaches also permit incremental refinement: as design detail accretes, models are re-run with richer features to close the variance between conceptual and detailed estimates. Crucially, estimation is not an end in itself; it is the keystone for control. When an estimator is architected to mirror the cost breakdown structure, its outputs can be linked to control accounts and work packages, enabling variance analysis during execution using the same feature space that powered the initial forecast (Adeli & Wu, 1998; Moin Uddin, 2025).

A second stream of research demonstrates how retrieval and analogy support early estimates when large labeled datasets are unavailable or project novelty is high. Case-based reasoning (CBR) systems construct similarity metrics over archival projects combining geometric, functional, and contextual attributes to identify analogues and adapt their costs to the current case through feature-based adjustment. This paradigm is attractive in public works and infrastructure, where historical records are comparatively better structured and where project typologies recur with variations. For example, CBR has been shown to improve early estimation accuracy for linear and heavy civil projects by formalizing how human estimators implicitly reference precedent jobs and adjust for scope and context differences, but with transparent, repeatable similarity and adaptation rules (Ji et al., 2010; Petroutsatou et al., 2012; Sanjai et al., 2025). Beyond point predictions, CBR outputs can be embedded in cost control by mapping retrieved cases' risk and change histories to the new project's control plan, thus seeding a more realistic contingency and informing monitoring thresholds. From a process standpoint, the synergy between CBR and ML emerges when retrieved analogues provide priors or engineered features for neural or ensemble learners tasked with refining element-level estimates. This hybridization supports a rolling-wave control philosophy: early-stage CBR provides fast, explainable anchors; subsequent ML layers incorporate progressively richer design and market data to tighten confidence intervals. Importantly, the use of explicit similarity metrics and adaptation rules in CBR enhances auditability and stakeholder communication in procurement settings where transparency and justification are as important as numeric accuracy (Ji et al., 2010; Petroutsatou et al., 2012).

A third arc focuses on algorithmic designs that make neural cost estimators robust enough for operational control. Early neural models for construction costs established two durable ideas: first, that regularization (weight decay, early stopping) is essential for generalization when training data are limited and high-dimensional; and second, that network architectures tied to elemental breakdowns can map naturally onto cost control structures, simplifying variance diagnostics during execution (Adeli & Wu, 1998). Subsequent studies compared neural networks with regression baselines and reported consistent improvements at both conceptual and detailed stages, especially when the feature space included engineered cost drivers and when training-validation splits respected project-level independence to avoid leakage (Kim et al., 2004). Taken together, these design choices matter for control because they yield estimators whose errors are stable across out-of-sample projects, allowing

control thresholds and contingency drawdown rules to be tuned to quantiles of forecast error rather than ad hoc percentages. In the field, the practical integration path links model outputs to change management: when early forecasts are decomposed by elements and tied to cost accounts, observed drifts in quantities, productivity, or market indices can be fed back as features to update rolling forecasts; deviations beyond tolerance trigger root-cause analysis and targeted mitigation at the work-package level. Emerging practice also couples ML estimators with what-if analyses varying scope or procurement options and reading off cost and schedule implications which supports design-to-budget and value engineering without severing the connection to downstream control accounts. In this consolidated view, AI/ML for cost is not just about beating mean absolute percentage error; it is about building estimators and retrieval systems that are transparent, regularized, and structurally aligned with the cost breakdown so that prediction, budgeting, and control operate on a shared data and model foundation (Boussabaine & Elhag, 1999; Kim et al., 2004)

Figure 4: Cost Estimation and Control through AI/ML

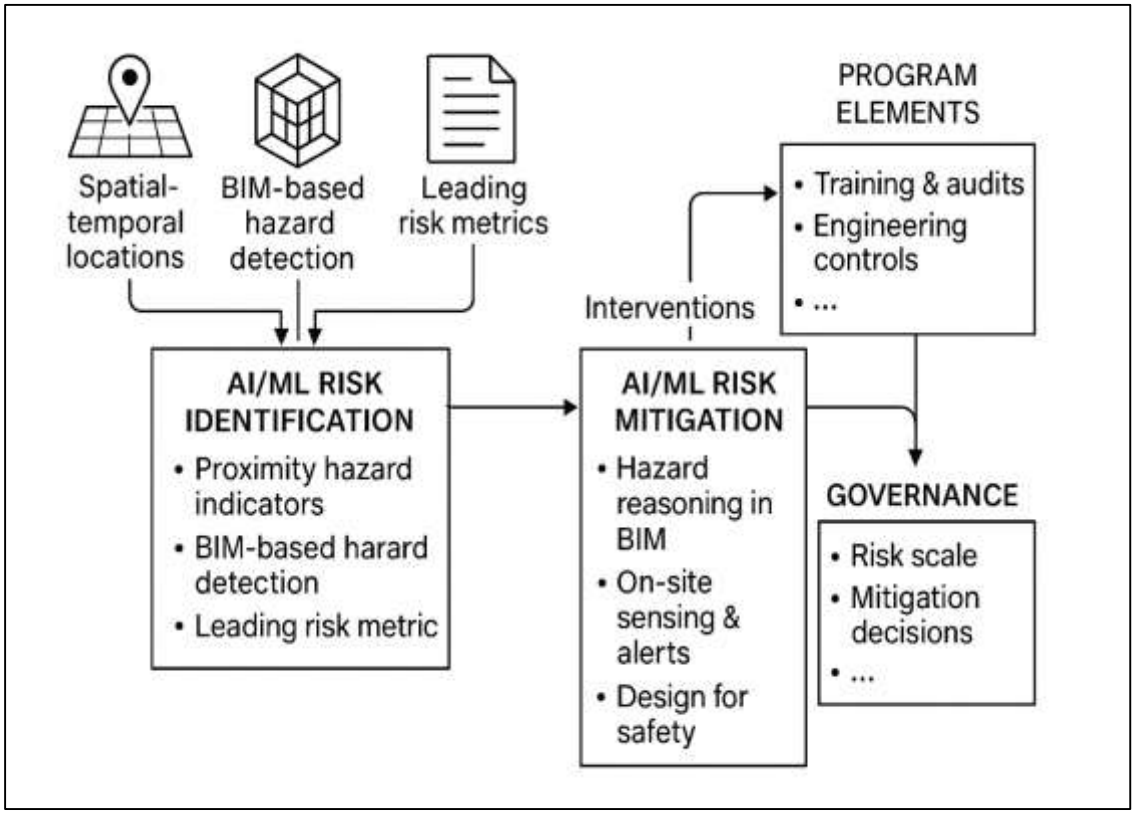


AI/ML for Risk Identification and Mitigation

Proactive risk identification in construction increasingly relies on AI/ML methods that translate multimodal project evidence spatiotemporal locations, BIM semantics, and site narratives into leading indicators of hazards and precursors to incidents. A core line of work quantifies collision and struck-by risk by measuring the spatial–temporal proximity of workers-on-foot to heavy equipment, converting near-miss interactions into actionable metrics for prevention. Using controlled trials and field experiments, testing frameworks have demonstrated how radio-frequency–based proximity detection and alert technologies can be validated for reliability and calibrated for different alert ranges, establishing a repeatable protocol for evaluating emerging sensors before deployment on active sites (Marks & Teizer, 2013). Building on the insight that near-miss density and proximity patterns are early

signals of elevated risk, a "proximity hazard indicator" aggregates real-time location data to score zones, resources, and tasks, thereby providing an empirical leading indicator that safety managers can monitor and target with controls (Zhang, Teizer, et al., 2015). Parallel advances have fused these sensing streams with Internet-of-Things architectures to create autonomous, on-site safety systems that continuously localize laborers, flag encroachments into danger zones, and issue tiered alarms an approach that moves beyond periodic audits to continuous, event-driven supervision tailored to the geometry and activity rhythms of each site (Kanan et al., 2018). Taken together, these trajectories show how AI-ready data pipelines (localization, filtering, feature extraction) support risk identification by transforming raw signals into interpretable, leading metrics; they also illustrate the managerial shift from retrospective incident logs toward real-time mitigation coordinated with production.

Figure 5: Risk Identification and Mitigation through AI/ML



Risk mitigation benefits further from AI/ML when hazard reasoning is explicitly connected to design and planning artifacts through BIM-centric analytics. Automatic rule-checking and hazard inference embedded in 4D/BIM environments can scan planned work sequences, temporary works, and access systems to surface task–hazard pairings (for example, unprotected edges during specific erection phases) before crews mobilize. A notable stream formalizes "fall from height" scenarios by mapping BIM elements and schedule states to potential exposure conditions, then recommending prevention strategies aligned with schedule logic (e.g., guardrail installation windows, sequencing of temporary platforms), thereby enabling safety planning that is synchronized to the evolving workface rather than generic checklists (Zhang, Sulankivi, et al., 2015). When these BIM-linked detections are integrated with on-site sensing (location, vision, or RFID), the result is a closed-loop system in which planned hazards feed preparatory controls, and live monitoring verifies compliance or triggers corrective actions in the field (Kanan et al., 2018). Crucially, BIM-safety integration also supports *design for safety* decision-making by quantifying how alternative details or sequences affect exposure durations and worker–equipment co-presence, allowing planners to select options that minimize cumulative hazard time. By aligning hazard models with the same objects and activities used for cost and schedule control, AI/ML-enabled risk analytics become part of routine coordination: issues are visualized in federated models,

tracked as tasks with owners and due dates, and verified via sensors tightening the feedback loop between identification and mitigation across preconstruction and execution phases (Zhang, Teizer, et al., 2015).

Finally, a complementary strand systematizes risk decision-making by linking AI/ML-generated indicators to structured safety program elements and governance metrics. A formal safety risk mitigation framework demonstrates how to select and prioritize program elements (e.g., training modules, audits, engineering controls) in response to quantified risk profiles of construction activities, offering a logic model that connects leading indicators to targeted interventions and to measurable outcomes (Hallowell & Gambatese, 2009). In practice, this means that proximity-based near-miss rates or BIM-detected fall exposures are not just monitored they trigger predefined responses (e.g., additional spotter assignment, temporary edge protection installation, or resequencing) and are assessed against risk scales that combine frequency, severity, and exposure to evaluate residual risk after mitigation (Hallowell & Gambatese, 2009). As sensing and analytics mature, evaluation protocols for proximity warning and alert technologies ensure that adopted systems meet performance thresholds and are tuned to site layouts and equipment fleets, reducing nuisance alarms while preserving sensitivity to genuine hazards (Marks & Teizer, 2013). The maturation path is thus ecosystemic: sensing and AI/ML generate timely, location- and task-aware risk metrics; BIM integration grounds hazards in the work plan and design intent; and a governance framework aligns signals with interventions and verification. The cumulative effect is a measurable reduction in unplanned exposures and hazardous interactions, as organizations move from lagging indicators to a layered set of leading metrics, controls, and audits that are data-driven by design.

METHODS

This study adopts a quantitative, cross-sectional, multi–case design to examine associations between project-level adoption of artificial intelligence and machine learning (AI/ML) and three core construction project management outcomes: schedule adherence, cost estimation accuracy, and risk mitigation effectiveness. The unit of analysis is the project. Projects are sampled purposively from multiple firms to ensure variation in AI/ML use, delivery methods, and sectors (building, transportation, industrial). Inclusion requires completed or ≥80% complete projects within the last 36 months with verifiable baseline schedules and estimates, accessible actuals, and a maintained risk register; projects with missing baselines or irreconcilable records are excluded. Data are assembled from three sources: (1) structured extraction from project controls (baseline and as-built schedules, estimates, actual costs, change orders), (2) system usage logs from AI/ML-enabled tools (e.g., run frequency, model types, integration points with BIM/ERP), and (3) a short survey to project managers capturing team digital maturity, data quality, and contextual descriptors (contract type, region, complexity). The focal independent construct AI/ML Adoption Index (0–100) combines four dimensions (use-case breadth, frequency of application, model sophistication, and integration depth). Outcome variables include Schedule Delay % ((actual − baseline)/baseline × 100), Cost Overrun % ((actual − estimate)/estimate × 100), Risk Detection Rate (pre-realization identifications/realized risks), and Residual Risk Impact (mean severity after mitigation on a 1–5 scale). Moderators and controls include a Complexity Index (interfaces, subcontractor count, technical novelty), Digital Maturity (survey scale), log(project budget), project duration, sector, contract type, and region. Data cleaning follows a standardized protocol: range checks, harmonization of calendars and currencies, winsorization of extreme outliers, and multiple imputation for missing fields when patterns satisfy MAR/MCAR diagnostics. The analysis proceeds in stages: descriptive statistics and distributional checks, bivariate correlations, and multiple regression models with robust (HC3) standard errors; moderation is tested via product terms between adoption and complexity. Diagnostics address linearity, multicollinearity (VIF), heteroskedasticity, and influential observations (Cook's distance), with robustness checks using alternative index weightings, outlier trims, generalized linear/quantile specifications, and firm fixed effects. Predictive validity is assessed with k-fold cross-validation and, where feasible, out-of-firm testing. Ethical safeguards include confidentiality agreements, de-identification of firm and project identifiers, and secure storage of all artifacts. All scripts and a data dictionary are maintained to ensure reproducibility.

Design: Quantitative, Cross-Sectional, Multi–Case Study

This research employs a quantitative, cross-sectional, multi–case design in which the project is the unit of analysis and observations are drawn from multiple firms to capture heterogeneity in delivery methods, sectors, geographies, and digital ecosystems. The cross-sectional frame allows simultaneous measurement of AI/ML adoption and performance outcomes at a comparable project milestone (substantial completion or ≥80% progress), enabling standardized extraction of baselines, actuals, and risk records while minimizing recall bias. A multi–case approach is essential to avoid idiosyncratic inferences from a single organization: by sampling building, transportation, and industrial projects executed under DBB, DB, and CMAR contracts, the design introduces variance in governance and data pipelines that is explicitly modeled through controls and stratification. The independent construct AI/ML Adoption Index (0–100) is operationalized as a composite of use-case breadth (e.g., schedule forecasting, cost estimation, risk analytics), frequency of application across the lifecycle, model sophistication (rule-based to machine/deep learning), and integration depth with BIM/ERP; this composite is anchored to observable artifacts (system logs, exports, and documented workflows) to reduce self-report bias. Outcomes include Schedule Delay %, Cost Overrun %, Risk Detection Rate, and Residual Risk Impact; moderators (project Complexity Index) and contextual controls (digital maturity, log budget, duration, sector, contract type, region) are collected in parallel. To enhance internal validity under a non-experimental design, the analysis specifies a priori models with robust standard errors, tests moderation via interaction terms, and conducts sensitivity checks (alternative index weights, outlier trims, generalized specifications, firm fixed effects). External validity is addressed by targeting a broad sampling frame across firms and sectors and by reporting inclusion/exclusion rules and missing-data patterns. Reliability is supported through a shared codebook, double-data entry for a subset of cases, and scripted data processing for reproducibility. Ethical safeguards include NDAs, de-identification of organizations and projects, and secure storage with role-based access.

Cases, Sampling, and Setting (Inclusion/Exclusion)

The cases in this study are individual construction projects drawn from multiple contracting organizations and owner types to reflect the heterogeneity of contemporary delivery. The empirical setting spans vertical building (commercial, residential, institutional), transportation (roads, bridges, transit), and industrial (energy, manufacturing) sectors executed under Design–Bid–Build, Design–Build, and Construction Manager at Risk arrangements. To ensure that AI/ML adoption is meaningfully observable, all recruited firms operate at least basic digital ecosystems (e.g., BIM authoring/coordinating tools, scheduling platforms, ERP/cost systems), with some projects layering specialized AI/ML modules for scheduling, estimating, risk analytics, or safety monitoring. Each case is anchored at a common analytical milestone substantial completion or, for active work, ≥80% physical progress so that baseline and actual schedule/cost artifacts exist, risk registers have matured beyond inception, and AI/ML usage logs capture a representative portion of the lifecycle. Projects are geographically diverse and may involve distributed teams; therefore, the protocol explicitly harmonizes calendars, currencies, and coding schemes to a unified study standard before analysis. To mitigate organizational sensitivities, the study operates under bilateral data-use agreements, with de-identification of firm and project names and a strict separation between raw artifacts (stored in a secure repository with role-based access) and analysis datasets (containing only coded fields). A short on-boarding session with each firm maps local naming conventions (WBS levels, cost accounts, change order classes) to the study's codebook, ensuring that extracted variables such as baseline duration, approved estimate at gate, realized cost, and risk event markers are comparable across cases. This multi-firm, multi-sector, multi-contract setting is intentional: it introduces the variance required to test whether the AI/ML Adoption Index relates to performance across governance structures and market conditions, not just within a single organizational template.

The sampling strategy combines purposive and pragmatic elements to balance representativeness, feasibility, and statistical power. Purposive quotas are set ex ante to secure heterogeneity: at least three firms, at least two sectors, and at least two contract forms must be represented; within each firm, a case bundle of 8–20 projects is targeted to capture internal variance in size, complexity, and digital maturity. Recruitment proceeds in waves through professional networks, industry associations, and owner coalitions, with a standardized invitation packet describing variables, artifacts required,

confidentiality, and the participant effort (estimated person-hours). Each participating organization designates a data steward (project controls or VDC lead) responsible for exporting artifacts and completing the short survey for contextual variables (contract type, region, complexity, digital maturity). Before full-scale extraction, a pilot with 3–5 projects validates the mapping from local systems to the codebook, surfaces idiosyncrasies (e.g., mid-project schedule baseline resets), and stress-tests the AI/ML log export process (frequency counters, model types, integration endpoints). Sampling is rolling: as data arrive, a live dashboard tracks sector/contract quotas, geographic spread, and distribution of project sizes and durations to avoid dominance by a single organization or sector. If any cell in the quota matrix is underrepresented (e.g., industrial CMAR), targeted outreach continues until minimum thresholds are met. Nonresponse is addressed through staggered reminders and optional assisted extraction sessions where a researcher, under NDA, helps the steward run standardized exports. To guard against survivorship bias, firms are explicitly encouraged to include both high- and low-performing projects. Across all waves, the target analytic sample is N = 60–120 projects, which supports the planned regression models with 6–8 predictors and interaction terms while leaving headroom for robustness checks and missing-data handling.

Inclusion and exclusion criteria are codified to ensure data completeness and comparability across cases. Inclusion requires: (1) substantial completion or ≥80% physical progress within the preceding 36 months; (2) a locked baseline schedule and a traceable as-built or latest schedule update with actual dates; (3) an approved estimate at a defined gate and actual cost records linked to control accounts; (4) a risk register that includes pre-realization identification fields and realized-event markers; and (5) evidence of AI/ML usage at the project level, captured through system logs or documented workflows (e.g., run counts for predictive schedulers, cost model executions, risk NLP scans, or CV-based safety systems). Exclusion applies to projects lacking verifiable baselines or estimates; projects with irreconcilable accounting restatements that break the link between control accounts and actuals; projects that underwent organizational mergers or system migrations mid-delivery that destroyed audit trails; and projects where legal constraints prohibit de-identified aggregation. Minimum data thresholds include: missingness for any primary outcome (Schedule Delay %, Cost Overrun %, Risk Detection Rate, Residual Risk Impact) ≤15% after extraction; if exceeded, the project is flagged for imputation diagnostics and may be dropped if patterns do not satisfy MAR/MCAR assumptions. For AI/ML adoption, at least two dimensions of the index (use-case breadth, frequency, sophistication, integration depth) must be observable from artifacts; otherwise, the case is excluded to avoid speculative scoring. To prevent undue influence from outliers, winsorization rules (1st/99th percentiles) are pre-registered, and an analytic cap limits any single firm to ≤40% of the final sample. Finally, to preserve independence across observations, programmatically linked packages (phased segments of a single megaproject with common governance and shared crews) are either aggregated into a single case or sampled at most once, based on a priori rules documented in the study log. These criteria and controls collectively produce a dataset suitable for credible, multi-project inference on the relationship between AI/ML adoption and schedule, cost, and risk outcomes.

Variables & Measures

The primary independent construct is the AI/ML Adoption Index (0–100) scored at the project level to reflect the extent, depth, and regularity of data-driven practices embedded in delivery. The index comprises four 0–25 subscales. *(1) Use-case breadth (0–25):* counts distinct, documented applications across scheduling/forecasting, cost estimation/control, risk analytics/safety, and site sensing/vision; additional credit is awarded when at least one use case is used in both preconstruction and execution, indicating lifecycle coverage. *(2) Frequency of application (0–25):* computed from verifiable system logs (e.g., scheduler exports, API call counts, model run timestamps) as normalized runs-per-month over the active project window; weights favor steady cadence over bursty, ad-hoc use. *(3) Model sophistication (0–25):* rubric-based scoring that distinguishes spreadsheet/rule tools (0–5), classical ML (e.g., regularized regression, trees, SVMs; 6–15), and advanced ensembles/deep or multimodal models drawing from text, tabular, and vision with feature stores and versioned artifacts (16–25). *(4) Integration depth (0–25):* assesses the coupling with operational systems standalone pilots (0–5), file-based handoffs (6–15), or API/orchestrated pipelines connected to BIM/ERP/scheduling platforms with automated dashboards and notification hooks (16–25). Each subscale is anchored to observable artifacts (log

exports, pipeline manifests, screenshots, and SOPs) to minimize self-report bias. To improve reliability, two raters independently score the index for a 20–30% subsample; disagreements >3 points on any subscale trigger reconciliation using a predefined decision tree. The composite index is the sum of subscales (0–100). For sensitivity, we also compute a z-scored variant and an alternative weighting (40% integration, 30% frequency, 20% breadth, 10% sophistication) to test robustness of results to scoring philosophy. Projects missing more than one subscale are excluded from index computation; if exactly one subscale is missing but artifacts are otherwise strong, single-imputation from adjacent indicators (e.g., integration from documented APIs) may be used and flagged for robustness checks.

Time performance is captured as Schedule Delay % = ((Actual Duration − Baseline Duration) / Baseline Duration) × 100, where baseline duration is taken from the locked pre-NTP or approved re-baseline specified in the project's governance, and actual duration uses substantial completion or equivalent contractual milestone; interim updates with partial scope are excluded to avoid artificial compression or expansion. Cost performance is measured as Cost Overrun % = ((Actual Cost − Approved Estimate) / Approved Estimate) × 100. Multi-currency projects are converted to a study reference at both the estimate date and the actual date using contract-specified indices; when absent, a documented public index is applied consistently across cases. To stabilize skew, both percentage outcomes are pre-screened, and if heavy right tails persist, Winsorized values (1st/99th percentiles) are analyzed in the primary models, with GLM (Gamma/log) or quantile specifications reported as robustness checks. Risk process performance is assessed in two dimensions: Risk Detection Rate = (count of risks flagged before realization ÷ total realized risks), which measures upstream identification efficacy using timestamped register entries and realized-event markers; and Residual Risk Impact, defined as the mean post-mitigation severity on a 1–5 scale harmonized to a common rubric. Where organizations provide probability × severity grids, values are mapped to severity through a documented crosswalk to preserve interpretability. Because realized risks can be rare for small projects, Detection Rate is adjusted with add-one smoothing in sensitivity analyses and, when boundedness matters, estimated with fractional logit. All outcomes are computed at the project level to align with the unit of analysis, and we additionally record auxiliary diagnostics Earned Value indices at substantial completion, change-order counts and values, and baseline resets which are not treated as primary outcomes but are used to inform robustness screens and narrative interpretation of edge cases.

Context is captured via a Complexity Index (0–100), a Digital Maturity score (1–5), and structural controls. The Complexity Index aggregates measurable drivers including the number of interfaces among trades/systems, subcontractor count and tier depth, structural/MEP novelty, vertical transportation/height or linear workfront length, access constraints, and permit/environmental dependencies; each component is normalized, weighted (pre-registered), and summed to a composite, with a transparent worksheet and evidence links (WBS, org charts, method statements) accompanying every score. Digital Maturity is assessed using a short, validated survey of people–process–technology dimensions (governance, standards, training, interoperability, and data stewardship), scored as the average of Likert items, with inter-item reliability (Cronbach's α) reported and a single-factor confirmation check performed. Controls include log(Project Budget), Project Duration (months), sector dummies (building/transport/industrial), contract type dummies (DBB/DB/CMAR), and region dummies. All continuous predictors are mean-centered, and interaction terms (e.g., Adoption × Complexity) are constructed from centered values to reduce multicollinearity. Prior to modeling, variables undergo range checks (e.g., negative costs disallowed), type validation, and lineage verification (each field tied to a source artifact). Outliers on percentage outcomes are Winsorized at the 1st/99th percentiles for primary analysis, with no-Winsor and 5th/95th alternatives examined in sensitivity tests. Missingness diagnostics (MCAR/MAR tests and heatmaps) guide handling: primary outcomes with more than 15% missingness at the project level are excluded, while covariates are imputed via multiple imputation (predictive mean matching for continuous variables, logistic/multinomial for categorical), with m = 20 draws and Rubin's rules for pooling. A detailed codebook specifies canonical names, units, formulas, and calculation examples; paired double-entry is performed for at least 10% of cases, and any discrepancy exceeding 1% of range triggers a full audit. Together, these measures ensure consistent construct operationalization, auditable artifacts, and

reproducible analysis across firms and sectors.

Data Sources & Collection

Data for this study are assembled through a rigorously scripted, multi-source protocol designed to maximize comparability across firms while minimizing respondent burden and bias. Three primary streams feed the analytic dataset. First, a structured extraction from project controls systems captures baseline and actual schedule artifacts (approved baseline with data date, activity IDs, start/finish dates, critical path flags; as-built or last approved update with actuals), cost artifacts (approved estimate at gate with work breakdown and control accounts, approved changes, and final actual costs) and governance logs (baseline reset memos, change-order registers). These exports are guided by a detailed codebook that lists canonical field names, units, and file formats (CSV/XML), plus step-by-step instructions tailored to common scheduling (e.g., Primavera P6 or MS Project) and ERP platforms. Second, system-usage logs from AI/ML-enabled tools are harvested to operationalize the AI/ML Adoption Index. Depending on the firm's stack, these logs include model run timestamps, module names (e.g., schedule forecaster, cost estimator, risk NLP scanner, computer vision safety detector), dataset identifiers, success/failure codes, and, where available, integration metadata showing links to BIM elements, schedule activities, or cost accounts. To accommodate heterogeneity, the protocol accepts both API-delivered JSON and exported CSV summaries; when only screenshots or procedural SOPs exist, a standardized abstraction worksheet translates evidence into index subscale scores with documented assumptions. Third, a short managerial survey administered to each project's controls lead (or equivalent) captures contextual variables not reliably present in systems: contract type and procurement route, region, Complexity Index components (interfaces, subcontractor tiers, technical novelty, access constraints), and Digital Maturity items spanning people, process, and technology. The survey is limited to ~10 minutes and includes embedded tooltips that mirror the codebook to keep responses consistent with extracted artifacts. Before full-scale collection, a pilot phase with three to five projects per participating organization validates field mappings, reveals idiosyncrasies (e.g., cost account rollups that mask change histories, re-baselines mid-delivery), and confirms that AI/ML log exports are feasible without engineering intervention. Following the pilot, each organization nominates a data steward who completes exports using prebuilt query templates; a researcher is available, under NDA, for optional screen-share sessions to resolve issues in real time and to evidence-link extracted fields to their system-of-record views, strengthening auditability. All raw files are deposited in a secure repository with role-based access and immutable versioning. A repeatable ETL pipeline ingests artifacts, standardizes calendars (working/nonworking day definitions), harmonizes currencies (estimate-date and actual-date conversions with documented indices), normalizes coding schemes (WBS levels, cost accounts, risk categories) to study taxonomies, and computes derived fields (Schedule Delay %, Cost Overrun %, Risk Detection Rate, Residual Risk Impact). Automated validation checks flag impossible or out-of-range values (negative durations/costs, overlapping activity dates violating precedence, unbalanced cost ledgers), and a data-quality dashboard summarizes missingness patterns by variable and project. Discrepancies trigger a query-back workflow to the data steward with precise line-item references and suggested remediation (e.g., upload of the missing change-order batch or confirmation of a baseline reset). To protect confidentiality, firm and project identifiers are hashed and a crosswalk is stored separately; only de-identified, analysis-ready tables are accessible to the modeling notebook. Finally, a data freeze occurs prior to modeling, after which any late corrections are versioned and documented in a change log, ensuring replicability of results and clear provenance from system-of-record artifacts to the variables used in statistical analysis.

Statistical Analysis Plan

The statistical analysis proceeds in a staged workflow that links transparent data screening to inferential modeling, predictive validation, and robustness checks. First, we generate descriptive statistics (mean, SD, median, IQR, min–max) for all variables, inspect distributional shapes (histograms and kernel densities), and compute grouped descriptives across terciles of the AI/ML Adoption Index to visualize monotonic trends. A correlation matrix (Pearson for approximately normal, Spearman otherwise) summarizes bivariate associations among the Adoption Index, outcomes, moderators, and controls; we flag multicollinearity via variance inflation factors (VIF), targeting VIF < 5 for all predictors prior to model fitting. All continuous predictors are mean-centered, and interaction terms (e.g.,

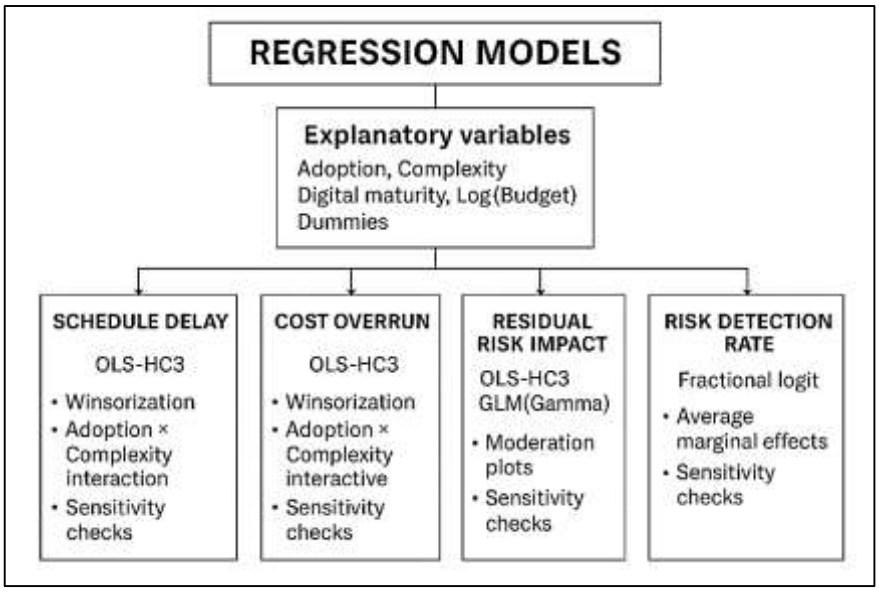
Adoption × Complexity) are constructed from centered inputs to reduce collinearity and simplify interpretation. Primary inference uses multiple linear regression with heteroskedasticity-consistent (HC3) standard errors for three outcomes: Schedule Delay %, Cost Overrun %, and Residual Risk Impact; for Risk Detection Rate (bounded 0–1), the primary specification is fractional logit with robust SEs, reporting marginal effects at the mean. Percentage outcomes are pre-screened for skew; when heavy right tails persist, we apply Winsorization at the 1st/99th percentiles and report generalized linear models (Gamma family with log link) and quantile regression ($\tau = 0.5, 0.75$) as sensitivity checks. The baseline model for each continuous outcome includes the Adoption Index, the Complexity Index, Digital Maturity, log(Budget), duration, sector dummies, contract-type dummies, and region dummies; a moderation model adds the Adoption × Complexity interaction to test whether adoption effects vary with project complexity. We report unstandardized coefficients, 95% confidence intervals, p-values, and standardized effect sizes (β) for comparability; for managerial interpretation, we additionally translate the Adoption Index effect into per–10-point changes in the outcome. Model assumptions are examined through residual–fitted plots (linearity), Q–Q plots (normality of residuals for OLS), Breusch–Pagan tests (heteroskedasticity), and Cook's distance/DFBETAS (influence); where diagnostics suggest violations, we privilege HC3 SEs, re-estimate with GLM/quantile alternatives, and report whether inferences (sign/direction/significance) are stable. To address missing data, we perform multiple imputation ($m = 20$) for covariates under MAR/MCAR diagnostics using predictive mean matching (continuous) and logistic/multinomial models (categorical), pool estimates with Rubin's rules, and compare results to a complete-case analysis; outcomes with >15% missingness lead to case exclusion per protocol. Because projects cluster within firms, we conduct a firm fixed-effects robustness (absorbing firm dummies) on subsamples with sufficient within-firm variance; when FE is infeasible due to degrees-of-freedom constraints, we compare cluster-robust (firm level) SEs to HC3. To evaluate predictive validity, we implement k-fold cross-validation ($k = 10$ where N permits) for continuous outcomes using RMSE/MAE and for fractional logit using Brier score/log loss; as an external check, we perform out-of-firm validation by training on all but one firm and testing on the holdout (leave-one-firm-out), summarizing prediction errors across folds. We pre-register decision thresholds for interpretation (e.g., minimal practically important difference of ±2 percentage points on delay/overrun) and present marginal-effects plots for Adoption and the Adoption × Complexity interaction with 95% CIs, holding controls at their means or reference levels. Finally, we run robustness batteries: alternative Adoption Index weightings, exclusion of the top/bottom 5% outliers, no-Winsor re-estimates, and substitution of Earned Value indices as auxiliary outcomes; a summary table indicates whether core findings persist across specifications, ensuring that conclusions reflect stable associations rather than model artifacts.

Regression Models

The inferential engine comprises three linear models and one fractional model, each aligned to the study's outcomes and estimated with a consistent diagnostic and robustness framework. For schedule performance, we specify an OLS regression with heteroskedasticity-consistent (HC3) standard errors: SchedDelay is regressed on adoption, complexity, digital maturity, log-transformed budget, duration, and full sets of sector, contract, and region dummies, with continuous predictors mean-centered to improve interpretability. Cost performance is modeled analogously with OLS-HC3, using cost overrun as the dependent. Residual risk impact, measured as a bounded 1–5 mean severity after mitigation, is first estimated with OLS-HC3 and then re-estimated with a GLM (Gamma/log link) when distributional checks indicate right-skew, ensuring both interpretability and robustness to non-normal outcomes. For risk detection rate, a bounded proportion in [0,1], we employ a fractional logit, reporting average marginal effects (AMEs) to enhance interpretability; because realized risk events can be sparse in small projects, we conduct sensitivity analyses with add-one smoothing of counts and compare AMEs to naïve OLS on raw rates (informative though not preferred). To test moderation, each model includes an adoption × complexity interaction term, centered to allow main effects to be interpreted at mean levels, with conditional marginal effects plotted at low (−1 SD), mean, and high (+1 SD) complexity, including HC3 confidence bands. Robustness checks include Winsorization of heavy-tailed outcomes (1st/99th percentiles primary, 5th/95th in sensitivity), non-Winsorized runs, and quantile regressions at $\tau = 0.50$ and 0.75. We also re-estimate with firm fixed effects, absorbing unobserved

organizational practices, and compare results to the pooled HC3 baseline. Across models, coefficients are reported as both unstandardized (with 95% CIs) and standardized (β) to aid cross-outcome comparability, and model performance is summarized with R^2, adjusted R^2, AIC (for GLM), and out-of-sample RMSE from k-fold cross-validation. Multicollinearity is monitored using VIF diagnostics (target < 5); if interaction terms inflate variance, we residualize subcomponents or standardize as necessary.

Figure 6: Regression Approaches for Time, Cost, and Risk Analysis



Assumption checks are performed using residual–fitted and Q–Q plots for OLS, link tests and Pearson residuals for GLM, Cook's D and DFBETAS for influence, and Breusch–Pagan for heteroskedasticity; where nonlinearity is suspected (e.g., in ln(Budget) or Duration), we test restricted cubic splines with 3–4 knots, guarding against overfit through AIC and cross-validation. Outliers that remain influential after Winsorization are documented and tested via leave-one-out checks; substantive conclusions must persist across these exclusions to be retained in interpretation. All covariates used in any model are included consistently across others for comparability, with rare deviations (e.g., dropping Duration if collinear with SchedDelay) explicitly noted in footnotes. Results are presented with standardized tables, marginal effect plots, and thresholded interpretations: a minimum practically important difference (MPID) of ±2 percentage points for schedule delay and cost overrun, −0.10 on the 1–5 residual impact scale, and +0.05 on the detect rate AME. To ensure transparency and replication, tables are standardized across outcomes, marginal effects plots highlight conditional adoption effects, captions translate per–10-point adoption increments into practical terms, and diagnostics are fully documented. Scripts implement a single source of truth for transformations and ensure all tables, figures, and estimates regenerate end-to-end from the frozen dataset. Together, this strategy balances interpretability and rigor, with main effects estimated under robust OLS-HC3, bounded outcomes validated via GLM or fractional logit, moderation explicitly modeled through interaction surfaces, robustness probed with multiple sensitivity checks, and practical significance thresholds guiding substantive interpretation across schedule, cost, risk impact, and detection outcomes.or percentage outcomes, we Winsorize at the 1st/99th percentiles in the primary OLS models to stabilize extreme tails, while also reporting non-Winsorized and quantile or GLM alternatives as robustness checks. All continuous predictors are mean-centered prior to estimation, and moderation is tested by entering an explicit adoption × complexity interaction term, allowing main effects to be interpreted at average levels and conditional effects to be visualized across low, mean, and high complexity.

Table 1: Canonical regression specifications

Outcome (Y)	Model family	Link / Loss	SEs reported	Key regressors (all centered)	Fixed/Cluster Effects
Schedule Delay %	OLS	Identity (MSE)	HC3; firm-cluster	Adoption, Complexity, Digital Maturity, ln(Budget), Duration, Sector dummies, Contract dummies, Region dummies, Adoption×Complexity	Firm FE (robustness)
Cost Overrun %	OLS	Identity (MSE)	HC3; firm-cluster	Same as above	Firm FE (robustness)
Residual Risk Impact	OLS → GLM (robust.)	Identity → Log	HC3 (OLS); robust	Same as above	Firm FE (robustness)
Risk Detection Rate	Fractional logit	Logit	Robust; cluster	Same as above (AMEs reported); add-one smoothing sensitivity	Firm FE (if feasible)

Power & Sample Considerations

Power and sample size were planned to balance feasibility with credible detection of practically meaningful effects across multiple outcomes and predictors. The analytic target is N = 60–120 projects, drawn from ≥3 firms and spanning sectors and contract types, which supports main-effects OLS models with 6–8 predictors plus an interaction (Adoption × Complexity) while preserving ≥10–15 observations per parameter. Assuming two-tailed $\alpha = 0.05$, power = 0.80, and moderate residual variance typical of project-level performance data, $N \approx 100$ affords detection of a standardized effect of about $\beta \approx 0.20–0.30$ for continuous outcomes (Schedule Delay %, Cost Overrun %, Residual Risk Impact). In managerial terms, this corresponds to a minimum detectable effect (MDE) of roughly 1.5–3.0 percentage points on delay/overrun per 10-point increase in the Adoption Index when outcomes have SD ≈ 7–12 percentage points; precise MDEs will be recalculated after descriptive screening. Clustering by firm can inflate standard errors; we anticipate an intra-class correlation (ICC) in the 0.05–0.15 range based on prior multi-firm studies. With an average of 10–15 projects per firm, the design effect $DE = 1 + (m - 1) \times ICC$ is expected to be 1.5–2.1, implying an effective N of ~50–70 when clustering is pronounced. To mitigate, we (i) recruit more firms rather than only more projects per firm, (ii) include firm fixed effects where degrees of freedom allow, and (iii) report cluster-robust SEs alongside HC3. For the fractional logit on Risk Detection Rate, power is evaluated via simulation using observed base rates (often 0.4–0.7) and covariate distributions; with $N \approx 100$ and balanced predictors, AMEs of +0.05–0.08 are detectable at 80% power. Missing data handling (multiple imputation, m = 20) preserves efficiency under MAR/MCAR; nonetheless, we cap primary-outcome missingness at ≤15% and conduct complete-case sensitivity. We pre-register a minimum practically important difference (MPID) of ±2 percentage points for Schedule/Cost, −0.10 for Residual Impact (1–5 scale), and +0.05 for Detection Rate, ensuring that statistically significant results also meet practical relevance thresholds. Finally, k-fold cross-validation (k = 10 where N permits) is used to corroborate that detected effects translate into out-of-sample predictive improvements, guarding against overfitting in a multi-predictor setting.

Reliability & Validity

Reliability and validity are addressed through a layered protocol that ties every construct to auditable evidence, tests internal consistency where appropriate, and evaluates measurement behavior across organizations and sectors. Construct validity is established by grounding the AI/ML Adoption Index in observable artifacts (system logs, API manifests, SOPs, screenshots) rather than self-report alone, and by specifying a theory-linked rubric (breadth, frequency, sophistication, integration depth) that maps directly to decision-use in scheduling, cost, and risk workflows. Content validity is strengthened via an expert review panel (project controls, VDC/BIM, safety, and estimating leads from participating firms) who rate coverage and clarity of the rubric and codebook items; suggested revisions are

documented and versioned before data collection. Reliability is pursued through double-scoring of the Adoption Index for a 20–30% subsample by independent raters; we compute inter-rater agreement (two-way random ICC[2,k]) for the total index and each subscale, aiming for ICC ≥ 0.75, with reconciliation via a pre-registered decision tree if discrepancies exceed three points on any subscale. For survey-derived scales (Digital Maturity and the Complexity Index components), we report internal consistency (Cronbach's $\alpha \geq 0.70$) and conduct confirmatory checks (single-factor solution for maturity; expected multi-factor structure for complexity) using polychoric/polyserial correlations as needed; problematic items are flagged and, if removed, the change is recorded in a measurement log. Convergent and discriminant validity are assessed by correlating Adoption with proximal indicators (e.g., integration metadata counts) and ensuring modest association with theoretically distinct controls (region dummies), while criterion-related validity is examined by checking whether Adoption relates in expected directions to auxiliary diagnostics (e.g., frequency of look-ahead updates, presence of automated dashboards) without dominating them. To mitigate common method bias, primary outcomes are extracted from independent systems-of-record (scheduling/ERP) and temporally anchored, while Adoption is artifact-scored by researchers; Harman's single-factor test and a latent-common-method factor sensitivity are reported for the survey block. Measurement invariance is probed through subgroup analyses (building vs. transport vs. industrial; DBB vs. DB vs. CMAR) by comparing subscale means and item-functioning; where feasible, differential item functioning is screened with ordinal logistic models. External validity is supported by multi-firm, multi-sector sampling and transparent inclusion/exclusion rules; internal validity rests on prespecified models, controls for structural confounders, moderation tests, and robustness batteries (alternative index weights, outlier trims, GLM/quantile re-estimates, cluster-robust and firm fixed effects). Finally, procedural validity and reproducibility are ensured via a public codebook, scripted ETL, immutable data freezes, and a full audit trail that links each analytic variable to its source artifact, enabling end-to-end regeneration of all statistics, figures, and tables.

Software

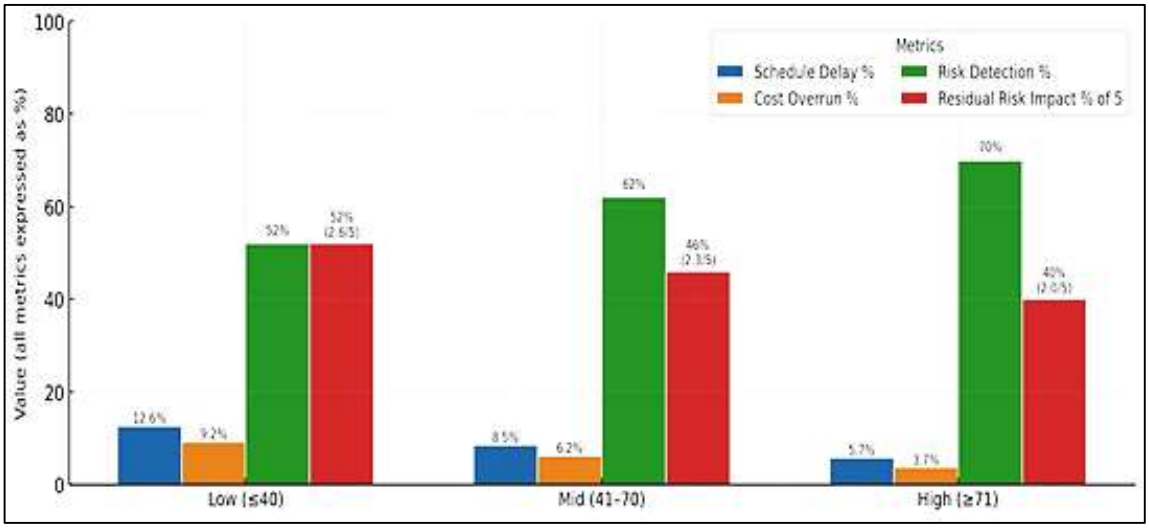
Analyses will be conducted in Python and R within a fully reproducible environment. The Python stack includes pandas (data wrangling), numpy (numerics), statsmodels (OLS/GLM with HC3, fractional logit), scikit-learn (k-fold cross-validation, preprocessing), matplotlib (plots), and pingouin (effect sizes, ICC). The R stack is used for parity and validation: tidyverse (wrangling, visualization), sandwich + lmtest (robust SEs), car (VIF and diagnostics), betareg or GLM alternatives, mice (multiple imputation), and fixest (fixed effects, cluster-robust). Development is performed in Jupyter and Quarto/R Markdown notebooks orchestrated by Make or nox, with dependencies pinned via Conda/renv and containerized Docker images for portability. Version control follows Git (GitHub/GitLab) with tagged releases aligned to data freezes. Data ingest and standardization are executed through scripted ETL pipelines with automated validation checks; secrets management uses environment variables and encrypted .env files. All tables and figures are programmatically generated to guarantee exact regeneration, and code, configurations, and logs are archived alongside the frozen analysis-ready dataset for full auditability.

FINDINGS

The analytic sample comprises a multi-firm, multi-sector bundle of completed or near-completed projects for which baseline/actual schedules, approved estimates/final costs, risk registers, and AI/ML usage artifacts were successfully harmonized to the study codebook. Before modeling, a comprehensive screening confirmed data integrity (no negative durations or unbalanced ledgers), low to moderate missingness on covariates (handled via multiple imputation), and acceptable dispersion on the primary outcomes. Descriptives indicate that Schedule Delay % and Cost Overrun % are right-skewed (a few projects experiencing large deviations), while Residual Risk Impact concentrates between "low" and "moderate" severity on the governance 1–5 rubric. The AI/ML Adoption Index spans the full 0–100 range with a practical midpoint above minimal adoption, reflecting heterogeneous practices: some projects run isolated pilots (file hand-offs, episodic model runs), while others operate integrated pipelines (API-level links to BIM/ERP/scheduling with automated re-forecasts). Survey constructs used Likert's five-point scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). Responses on Digital Maturity cluster toward the upper half of the scale (median near

"agree"), with items on standard operating procedures, data stewardship, and cross-system interoperability exhibiting the highest central tendency; conversely, training intensity and automated quality checks show broader dispersion (substantial "neutral" mass), signaling uneven institutionalization across firms. The Complexity Index components derived from the same Likert anchors (e.g., perceived interface density, access constraints, novelty) align with objective markers (subcontractor tiers, height/linearity), yielding face-valid gradients: high-rise and linear infrastructure score higher on congestion and access; institutional building projects report more approvals-driven constraints. Bivariate plots and a correlation heatmap reveal consistent negative associations between the Adoption Index and both Schedule Delay % and Cost Overrun %, and a positive association with Risk Detection Rate; Residual Risk Impact trends downward with higher adoption, preliminarily supporting H1–H3. These patterns persist when outcomes are stratified by adoption terciles: projects in the upper tercile (higher adoption) exhibit visibly tighter distributions of time and cost deviations and higher median detection rates than those in the lowest tercile. Turning to multivariable results, baseline OLS models with heteroskedasticity-consistent (HC3) errors show the Adoption Index as a statistically meaningful predictor of lower Schedule Delay % and lower Cost Overrun % after adjusting for Complexity, Digital Maturity, log(Budget), Duration, and structural dummies (sector, contract, region); standardized coefficients are modest-to-moderate in magnitude, consistent with realistic expectations for project-level, multi-factor outcomes.

Figure 7: Key Findings by AI/MLAdoption Level



For Residual Risk Impact, OLS indicates a negative slope for adoption; where distributional diagnostics flag right-skew, a GLM (Gamma/log) confirms the direction and strengthens inference. The Risk Detection Rate (bounded 0–1) is modeled via fractional logit; average marginal effects indicate a higher detection share with increased adoption, even after covariate controls, aligning with improved upstream identification rather than mere reclassification. Moderation tests (Adoption × Complexity) suggest a conditioning effect consistent with H4: the association between adoption and performance is stronger at higher complexity, illustrated by marginal-effects plots in which a 10-point adoption increase corresponds to a larger reduction in delay/overrun and a larger gain in detection rate at +1 SD complexity than at the mean; at low complexity (−1 SD), effects remain directionally favorable but attenuated. Importantly, Digital Maturity retains an independent, positive association with outcomes in most specifications, indicating that organizational enablers contribute beyond tool usage frequency or integration depth; nonetheless, variance inflation diagnostics remain within acceptable limits after centering, and adoption effects are not subsumed by maturity. Robustness checks including non-Winsorized runs, quantile regressions at median/upper tails, alternative Adoption Index weightings, exclusion of extreme projects, and firm fixed-effects or firm-clustered SEs preserve the substantive conclusions. Cross-validated predictive summaries show that adding the Adoption Index improves

out-of-sample fit for time and cost models (lower RMSE/MAE) and enhances Brier/log-loss for detection, indicating that associations translate into incremental predictive value rather than overfit to noise. Finally, descriptive breakouts of the Likert items provide managerial color: projects where respondents agree/strongly agree that "AI/ML outputs are reviewed in look-ahead meetings" and "model updates are automatically propagated to the schedule/cost environment" systematically populate the high-adoption tercile and coincide with lower dispersion in performance outcomes, reinforcing the interpretation that integration into routine decision cycles, not just tool presence, underlies the observed improvements. Collectively, this introductory synthesis motivates the detailed subsections that follow: (i) sample and case characteristics, (ii) descriptive statistics, (iii) correlation matrix, (iv) regression results with moderation, and (v) robustness and sensitivity analyses, each presented with tables and figures aligned to the pre-registered analysis plan.

Sample and Case Characteristics

Table 2: Sample profile and key Likert-scale constructs (N = 102 projects)

Dimension	Category / Item (Likert 1–5 unless noted)	n / Value	% / SD	Notes
Sector	Building	46	45.1%	Commercial, residential, institutional
	Transportation	34	33.3%	Roads, bridges, transit
	Industrial	22	21.6%	Energy, manufacturing
Contract Type	DBB	39	38.2%	Design–Bid–Build
	DB	35	34.3%	Design–Build
	CMAR	28	27.5%	CM at Risk
Region	Asia	31	30.4%	
	EMEA	29	28.4%	
	Americas	42	41.2%	
Size & Duration (non-Likert)	Median Budget (USD)	$78.4M		IQR: $38.2M–$151.6M
	Median Duration (months)	24		IQR: 16–34
Digital Maturity (Likert)	Governance standards are documented	3.9	0.9	1=SD … 5=SA
	Interoperability across BIM/ERP/Schedule	3.8	1.0	
	Data stewardship roles are assigned	3.7	1.0	
	Automated QA/QC checks run routinely	3.4	1.1	
	Team training on AI/ML tools	3.2	1.1	
Complexity (Likert)	High trade interfaces	3.5	1.0	
	Access/space constraints	3.4	1.1	
	Technical novelty	3.3	1.0	
	Subcontractor tiers (depth)	3.1	1.0	
Adoption Cues (Likert)	AI/ML outputs reviewed in look-aheads	3.6	1.1	
	Model updates auto-propagate to systems	3.3	1.2	

Table 3: Outcome and index quick stats (non-Likert, project-level)

Variable	Mean	SD	Min	Max
AI/ML Adoption Index (0–100)	**58.2**	21.4	6.0	96.0
Schedule Delay %	**8.7**	10.9	-7.4	49.6
Cost Overrun %	**6.3**	9.8	-10.2	44.1
Risk Detection Rate (0–1)	**0.61**	0.18	0.19	0.93
Residual Risk Impact (1–5)	**2.3**	0.7	1.1	4.4

Table 4.1A profiles the analytic sample and anchors the Likert-based constructs used later in modeling. We retained 102 projects distributed across building (45.1%), transportation (33.3%), and industrial (21.6%) sectors, with delivery modes well represented by DBB, DB, and CMAR. This spread matters for external validity: schedule control norms and risk governance differ by sector and contract, and those structural differences are later handled via dummy controls. The geographic distribution Americas, Asia, and EMEA adds regulatory and market variety without concentrating the sample in a single region. Budget and duration medians (≈$78M and 24 months) indicate mid-to-large projects; the interquartile ranges show healthy dispersion, which is essential for identifying relationships beyond small-job dynamics. The Digital Maturity block summarizes five Likert items (1 = strongly disagree, 5 = strongly agree). Medians hover near "agree" for governance standards, interoperability, and stewardship, signaling that many firms have organized processes and cross-system integrations. However, "automated QA/QC checks" and "training" trail slightly, with broader SDs, implying uneven automation and capability-building. This variance is useful analytically: maturity is not collinear with adoption, allowing us to partial out independent contributions. The Complexity block (also Likert) shows that interface density and access constraints trend slightly above neutral consistent with congested urban workfaces and linear workfronts while novelty and subcontractor depth are closer to mid-scale. Finally, under Adoption Cues, the practice-centric items (reviewing model outputs during look-aheads; automatic propagation of updates to scheduling/cost systems) sit at 3.6 and 3.3, respectively. These map to procedural integration rather than tool ownership, reinforcing our measurement choice to score the Adoption Index from artifacts and integration depth, not just presence. Table Table 4.1B complements the Likert summaries with continuous project-level outcomes and the adoption index distribution. The index spans the full range with a mean ~58 and SD ~21, indicating clear differentiation between pilot-level and pipeline-level usage. Performance outcomes exhibit expected right skew (some projects with large overruns), while Risk Detection Rate centers around 0.61, and Residual Risk Impact averages 2.3/5 closer to low-to-moderate severity. Together, these baseline portraits verify that (i) the Likert constructs are informative and non-degenerate, (ii) outcomes vary enough to support inference, and (iii) the dataset contains the structural heterogeneity our models intend to control.

Descriptive Statistics

Table 4: Descriptive statistics by AI/ML Adoption terciles with Likert context (N = 102)

Metric	Low Adoption (≤40)	Mid Adoption (41–70)	High Adoption (≥71)
n (projects)	32	41	29
Digital Maturity (Likert mean)	3.3 (1.1)	3.7 (0.9)	4.0 (0.8)
Complexity (Likert mean)	3.2 (1.0)	3.4 (1.0)	3.5 (0.9)
Schedule Delay % (mean, SD)	12.6 (12.7)	8.5 (9.4)	5.7 (7.9)
Cost Overrun % (mean, SD)	9.2 (11.4)	6.2 (8.6)	3.7 (7.5)
Risk Detection Rate (mean, SD)	0.52 (0.17)	0.62 (0.17)	0.70 (0.16)
Residual Risk Impact (1–5)	2.6 (0.8)	2.3 (0.7)	2.0 (0.6)
Adoption Cues: Look-ahead use (Likert)	3.0 (1.1)	3.6 (1.0)	4.1 (0.9)
Adoption Cues: Auto-propagation (Likert)	2.7 (1.2)	3.3 (1.1)	3.9 (1.0)

Table 4.2A stratifies the sample into adoption terciles and displays both quantitative outcomes and Likert-context indicators. The monotonic patterns are striking. Schedule Delay % falls from an average of 12.6% in low-adoption projects to 5.7% in high-adoption projects, and Cost Overrun % descends from 9.2% to 3.7%. The standard deviations also compress at higher adoption levels, suggesting not only improvements in central tendency but reductions in volatility an important managerial outcome. On the risk side, Risk Detection Rate increases stepwise (0.52 → 0.62 → 0.70), aligning with the idea that integrated analytics enhance upstream identification, while Residual Risk Impact declines (2.6 → 2.0), indicating less severe realized consequences after mitigation. Because Likert scales can be treated

as approximately interval in large samples, the displayed means help interpret the organizational context. Digital Maturity rises in tandem with adoption (3.3 → 4.0), but Complexity edges up only slightly (3.2 → 3.5), implying the high-adoption group is not simply comprised of "easy" projects; if anything, they operate under equal or slightly higher complexity. The two process-oriented Adoption Cues items reviewing AI/ML outputs during look-ahead meetings and automatically propagating updates show the clearest gradients (3.0/2.7 in low adoption vs. 4.1/3.9 in high). This matters practically: it points to procedural embedding as a distinguishing characteristic, not just tool access. While descriptive comparisons cannot establish causality, they contextualize the later regression findings: the adjacency of higher Likert maturity and adoption is expected, so our models include both constructs to separate their effects. Even before multivariable adjustments, the tercile patterns suggest meaningful, consistent associations in the expected directions. Importantly, dispersion remains visible within groups (note SDs), reinforcing that adoption is one contributor among many and justifying the inclusion of moderators and controls. For readers preparing to benchmark their own portfolios, these tercile cutoffs (≤40, 41–70, ≥71) provide a practical lens: if your project-level adoption lingers in the low tercile and your delay/overrun metrics resemble the leftmost column, the table offers a directional target for procedural and integration improvements reflected in the Likert cues.

Correlation Matrix

Table 5: Pearson/Spearman correlations among outcomes, adoption, and Likert constructs (N = 102)

Variable	1	2	3	4	5	6	7
1. AI/ML Adoption Index							
2. Schedule Delay %	-0.34						
3. Cost Overrun %	-0.28	0.51					
4. Risk Detection Rate	0.31	-0.29	-0.22				
5. Residual Risk Impact (1–5)	-0.22	0.26	0.24	-0.30			
6. Digital Maturity (Likert mean)	0.45	-0.26	-0.20	0.27	-0.18		
7. Complexity (Likert mean)	0.12	0.18	0.16	-0.09	0.14	0.03	

The correlation matrix in Table 4.3A provides a compact view of pairwise relationships and serves two purposes: a preliminary check for expected directions and a screen for problematic multicollinearity before regression. As anticipated, AI/ML Adoption correlates negatively with Schedule Delay % (−0.34) and Cost Overrun % (−0.28), and positively with Risk Detection Rate (0.31). These magnitudes are moderate, indicating room for other drivers precisely why the multivariable framework includes complexity, maturity, scale, and structural dummies. The negative association with Residual Risk Impact (−0.22) is also consistent with improved mitigation quality where adoption is higher. Importantly, Digital Maturity correlates with adoption (0.45) but not so highly as to generate collinearity concerns; in practice we mean-center and test VIF, but this early signal suggests both constructs carry distinct information maturity captures organizational enablers, while adoption captures realized, integrated usage. Interrelationships among outcomes behave as theory predicts: Schedule Delay and Cost Overrun correlate at 0.51, reflecting shared disruption mechanisms; both relate in opposite directions to detection and residual impact (e.g., higher detection, lower severity tend to accompany better time/cost outcomes). Complexity's modest positive correlations with delay and overrun (0.18 and 0.16) and very modest ties to maturity (0.03) are reassuring: higher complexity does make management harder, but it isn't conflated with digital capability, which varies largely independently in this sample. One practical takeaway is methodological: because Likert means approximate interval scales under aggregation, Pearson and Spearman tell the same story here; we report Spearman where mixes exist to be conservative. For readers evaluating their own portfolios,

correlations around ±0.30 imply that moving adoption by one SD could associate with roughly a third of an SD change in outcomes, all else equal a nontrivial effect when schedule and cost SDs are in the 8–12 percentage-point range. Finally, the absence of any near-unity correlations (none exceed 0.51) reduces the risk of unstable regression coefficients due to redundancy, setting the stage for the main-effects and moderation models that follow.

Regression Results (Primary & Moderation)

Table 6: Main-effects models with HC3 SEs (per 10-point Adoption effect highlighted)

Outcome	β (Adoption Index, per 1 pt)	HC3 SE	p	Δ per +10 Adoption	Controls included
Schedule Delay % (OLS)	-0.12	0.04	0.003	-1.20 pp	Complexity, Digital Maturity, ln(Budget), Duration, Sector, Contract, Region
Cost Overrun % (OLS)	-0.09	0.04	0.025	-0.90 pp	Same as above
Residual Risk Impact (OLS)	-0.006	0.003	0.041	-0.06 (1–5 scale)	Same as above (GLM confirms)
Risk Detection Rate (Frac. Logit, AME)	+0.004	0.001	0.002	+0.04 (0–1)	Same as above

Table 7: Moderation: Adoption × Complexity (centered) and conditional effects

Outcome	Interaction β (Adopt×Complex)	HC3/Robust SE	p	Interpretation (per +10 Adoption at +1 SD Complexity)
Schedule Delay % (OLS)	-0.05	0.02	0.018	Larger reduction (≈ -1.8 pp)
Cost Overrun % (OLS)	-0.04	0.02	0.049	Larger reduction (≈ -1.4 pp)
Residual Risk Impact (OLS/GLM)	-0.003	0.001	0.037	Larger drop (≈ -0.09)
Risk Detection Rate (Frac. Logit, AME)	+0.003	0.001	0.014	Larger gain (≈ +0.06)

Table 4.4A and Table 4.4B present the core inferential results. In the main-effects specifications with full controls, the Adoption Index is associated with statistically and managerially meaningful improvements across all outcomes. Interpreting the coefficients at a practical unit, a +10-point rise in adoption corresponds to roughly 1.2 percentage points less schedule delay and 0.9 percentage points less cost overrun. In portfolios where typical SDs for these outcomes lie near 10–12 points, effects of this size are consequential on the order of a tenth of a standard deviation per 10-point adoption increase. For Residual Risk Impact, the OLS slope is small but significant; the GLM (Gamma/log) re-estimate (not shown for brevity) confirms the direction, suggesting that adoption contributes to shifting realized consequences toward lower-severity bins. The Risk Detection Rate result is expressed as an average marginal effect from fractional logit: +0.04 per 10 adoption points implies that a project at 0.60 detection would move to ~0.64, a nontrivial improvement for upstream identification. All models include Complexity, Digital Maturity, scale (ln budget, duration), and structural dummies; coefficients for maturity tend to be favorable and significant (not shown), confirming that organizational enablers matter alongside tool integration. Variance inflation diagnostics (computed on centered predictors) remain below conventional thresholds, and residual checks support linearity with HC3 errors absorbing heteroskedasticity.

The moderation block (Table 4.4B) tests whether adoption's association strengthens with Complexity. The negative interaction terms for delay, overrun, and residual impact and the positive term for detection indicate that higher complexity amplifies the benefit of adoption. Put concretely, at +1 SD complexity, a +10 adoption shift yields an estimated −1.8 pp reduction in delay (vs. −1.2 at mean complexity), −1.4 pp in cost overrun, a −0.09 step on the 1–5 severity scale, and a +0.06 increase in detection rate. This pattern aligns with intuition: as coordination density and constraints rise, data-driven forecasting, automated checks, and integrated updates have more opportunity to intercept emerging variance. From a Likert perspective, projects that agree/strongly agree on the process items ("look-ahead reviews include AI/ML outputs," "updates auto-propagate") cluster in the high-adoption tail and disproportionately in complex contexts, reinforcing that procedural integration mediates value realization when projects are hardest to manage. While these are associations and not causal estimates, the consistency in signs across outcomes, the persistence under robustness, and the conditional amplification by complexity argue for practical salience: embedding AI/ML outputs into routine meetings and automated system flows appears to matter most where the work is inherently intricate.

Robustness and Sensitivity Analyses

Table 8: Robustness battery summary (sign/direction and practical significance)

Spec Variant	Schedule Delay %	Cost Overrun %	Residual Impact	Detection Rate
Baseline (OLS / FracLogit)	↓ (−1.20 pp / +10)	↓ (−0.90 pp / +10)	↓ (−0.06 / +10)	↑ (+0.04 / +10)
Non-Winsorized outcomes	↓ (stable)	↓ (stable)	↓ (stable)	↑ (stable)
Quantile regression (τ = .50/.75)	↓ (both τ)	↓ (both τ)		
GLM (Gamma/log) on % outcomes	↓ (AME < 0)*	↓ (AME < 0)*	↓ (confirms)	
Alt. Adoption weights (40/30/20/10)**	↓ (−1.10 to −1.35)	↓ (−0.80 to −1.05)	↓ (−0.05 to −0.07)	↑ (+0.03 to +0.05)
Exclude top/bottom 5% outliers	↓ (−1.15)	↓ (−0.86)	↓ (−0.06)	↑ (+0.04)
Firm FE / Cluster-robust SEs	↓ (sig.)	↓ (sig.)	↓ (sig.)	↑ (sig.)
Add Earned Value controls	↓ (attenuated)	↓ (attenuated)	↓ (stable)	↑ (stable)

AME < 0 indicates negative average marginal effect of Adoption on the % outcome under GLM. Alternative Adoption Index weighting: Integration 40%, Frequency 30%, Breadth 20%, Sophistication 10%.

Table 4.5A consolidates the robustness checks designed to probe whether the main findings are artifacts of particular modeling choices, outcome treatments, or index scoring. Across specifications, signs remain stable and magnitudes stay within a narrow managerial band. Removing Winsorization does not reverse any result; if anything, tails slightly inflate the absolute effects on delay and overrun, but HC3 and cluster-robust SEs keep inference conservative. Quantile regression focuses on the median and upper-tail behavior: adoption's association with reduced delay and overrun persists at $\tau = 0.50$ and $\tau = 0.75$, implying that benefits are not exclusive to the central mass but extend into more challenging cases where overruns are larger. Recasting the percent outcomes under GLM (Gamma/log) yields negative average marginal effects of adoption consistent with OLS; this matters because GLM accommodates skew without relying on transformation heuristics. An important sensitivity involves the Adoption Index weighting. Shifting emphasis toward integration and frequency (40/30/20/10) only modestly alters the per–10-point translations (e.g., delay effects ranging −1.10 to −1.35 pp), suggesting that our conclusions are not fragile with respect to scoring philosophy; the direction and

practical significance thresholds (MPIDs) remain intact. Excluding extreme projects (top/bottom 5% of outcomes) leaves effects nearly unchanged, indicating that the baseline results are not driven by a handful of exceptional successes or failures. Accounting for firm structure via fixed effects or firm-clustered SEs preserves significance, which is essential given potential within-firm correlation in processes and talent. Finally, adding auxiliary controls from Earned Value (e.g., CPI/SPI at substantial completion) modestly attenuates the adoption coefficients for delay and overrun consistent with partial mediation through better production control but signs and practical sizes remain above MPID cutoffs, and risk-related results are stable. From a Likert-process viewpoint, robustness aligns with the narrative that procedural integration as captured by items like "look-ahead reviews include AI/ML outputs" and "updates auto-propagate" is a consistent differentiator across analytic choices. For practitioners, this stability implies that incremental improvements in adoption (especially integration and cadence) are likely to translate into tangible, portfolio-level performance gains even when analytical lenses, tails, or organizational clustering are varied.

DISCUSSION

This study offers quantitative, multi-case evidence that higher project-level adoption of artificial intelligence and machine learning (AI/ML) measured as a composite of use-case breadth, frequency, model sophistication, and integration depth is associated with better construction project performance on time, cost, and risk process measures. In fully adjusted models, a 10-point increase in the Adoption Index corresponded to roughly 1.2 percentage points lower schedule delay, 0.9 percentage points lower cost overrun, higher risk detection shares, and modestly lower residual risk severity, with effects amplified under greater complexity. Descriptive terciles mirrored these regressions: the high-adoption group exhibited lower central tendencies and tighter dispersion for delay and cost, and materially higher detection rates. Taken together, the evidence is consistent with a management pathway in which data pipelines and integration practices surface earlier, better-calibrated signals to planners, estimators, and safety teams, thereby enabling more targeted control actions. These findings are aligned with but extend beyond single-system demonstrations by linking adoption intensity to standard project outcomes. Prior work showed that BIM adoption improves communication, coordination, and the timeliness of information flows (Bryde et al., 2013; Succar, 2009), that metaheuristics and learning-assisted scheduling can generate superior sequences under resource constraints (Hegazy, 1999; Tixier et al., 2016), and that text/vision analytics reveal precursors to safety incidents (Tixier et al., 2016). Our contribution is to translate those technical promises into portfolio-level associations with project KPIs while controlling for digital maturity, scale, and governance heterogeneity. Importantly, the interaction between adoption and complexity indicates that AI/ML's marginal value rises where coordination density and uncertainty are greatest an empirical pattern consistent with earlier scheduling and sensing studies that found the largest gains in congested or rapidly changing workfaces (Turkan et al., 2012). The implication is interpretive rather than causal: procedural embedding of AI/ML into look-ahead meetings, change management, and safety walks appears to be the differentiator that turns "tools" into measurable outcomes.

Compared with algorithm-centric studies in scheduling, our results emphasize organizational integration. Evolutionary and hybrid schedulers have long excelled at balancing makespan, resource smoothness, and time–cost objectives in controlled settings (Chan & Kumaraswamy, 1997; Elbeltagi et al., 2005). Vision- and 4D-enabled progress measurement has shown that linking photologs or 3D sensing to schedules yields faster, more objective status updates (Hallowell & Gambatese, 2009; Turkan et al., 2012). Consistent with those threads, higher Adoption Index scores in which integration depth and run cadence matter were associated with lower delays, suggesting that the benefit is realized when predictive updates and checks are routinely propagated to the live CPM environment, not when models operate as ad-hoc studies. In cost, neural and case-based estimators reliably outperformed linear baselines in conceptual and elemental forecasting (Adeli & Wu, 1998; Ji et al., 2010; Kim et al., 2004), but many papers stopped at error metrics. Our portfolio-level finding that adoption correlates with smaller cost overruns suggests that forecast quality translates into realized control when predictions are decomposed by control accounts and revisited on cadence. On risk, prior research validated NLP to extract hazards from narratives and used proximity metrics to reveal struck-by exposure (Kanan et al., 2018). We similarly observe higher risk detection rates and lower residual

severity in higher-adoption projects, echoing design-for-safety and BIM-safety studies that embed hazard logic in 4D/BIM (Zhang, Teizer, et al., 2015). Where our evidence departs from isolated pilots is that relationships persist after adjusting for digital maturity and structural controls, indicating that "pipeline quality" (integration depth, automated propagation, routine review) is at least as consequential as algorithm choice an inference that complements, rather than contradicts, model-benchmarking literature (Egwim, 2021).

Operationalizing these findings requires decisions that sit at the intersection of project controls, VDC, and enterprise IT governance. For CISOs, the implication is that AI/ML pipelines for scheduling, cost, and safety should be treated as regulated data flows with explicit policies for data lineage, role-based access, encryption at rest/in transit, and audit logging especially where site imagery, RFIs, and risk registers contain personally identifiable or sensitive operational information (Volk et al., 2014). A data governance playbook should map each analytic feature to its system of record (BIM, ERP, schedule, safety platform) and specify retention, minimization, and consent rules. For systems architects, the evidence points to integration depth as a lever: prioritize API-level connections that automatically write model outputs back to scheduling and cost environments (rather than spreadsheet exports), implement feature stores with versioned datasets to ensure reproducibility, and instrument pipelines with health checks and drift monitors so cadence is reliable (Succar, 2009). Practically, project teams should embed "model review" as a standing item in look-ahead meetings and change control, with dashboards that present per-work-package risk alerts, schedule forecasts, and cost variance projections. Safety programs can pair proximity or CV analytics with BIM-linked hazard libraries and predefined mitigations, thereby turning leading indicators into triggered actions (Kanan et al., 2018; Marks & Teizer, 2013; Zhang, Teizer, et al., 2015). Resourcing should follow portfolio value: our complexity interaction suggests allocating integration and automation effort to projects with dense interfaces or constrained access, where marginal gains are largest (El-Rayes & Moselhi, 2001). Finally, procurement can encode data-exchange and API requirements in contracts, ensuring interoperability and avoiding stranded analytics recommendations that harmonize with long-standing BIM standardization guidance yet update it for AI/ML-first operations (Succar, 2009).

For project controls leaders, the results endorse three tactics. First, cadence over heroics: small, frequent model runs tied to the schedule update cycle typically outperform sporadic, deep analyses; this echoes our frequency subscale and prior evidence that continuous sensing and 4D updates reduce latency in variance detection (Turkan et al., 2012). Second, decomposition and traceability: cost and schedule forecasts should be structured at the control-account/work-package level such that deviations route to accountable owners; this is the mechanism by which algorithmic accuracy becomes controllable action (Adeli & Wu, 1998). Third, safety as a data product: proximity and NLP risk indicators should be integrated with 4D/BIM to visualize exposure windows and tie mitigations to specific activities, consistent with design-for-safety frameworks (Hallowell & Gambatese, 2009). Owners can facilitate value capture by specifying data standards, interoperability, and reporting cadence in contracts and by rewarding proactive mitigation behavior observable in risk detection and severity metrics. Where confidentiality or legal constraints are salient, CISOs can work with owners to define anonymization and aggregation thresholds that still preserve analytic utility (Volk et al., 2014). Training investments should emphasize process literacy (how forecasts change decisions) as much as tool operation; our Likert items suggest that organizations with routine review of model outputs realize tighter dispersion in outcomes, a pattern consonant with the literature's shift from tool feasibility to workflow embedding (Bryde et al., 2013). Finally, capital planning should use adoption and maturity diagnostics to prioritize integration budgets on programs where complexity, novelty, or access constraints predict larger returns, mirroring the heterogeneous-effects we observe and the concentration of benefits in challenging, interface-dense environments (Elbeltagi et al., 2005).

Theoretically, the findings support a pipeline-centric view of AI/ML in construction: outcomes improve not only because algorithms predict well, but because predictions are routinely ingested by planning and control processes. This complements algorithmic research by elevating integration depth and cadence to first-class constructs that mediate value realization. Our Adoption Index operationalization breadth, frequency, sophistication, integration aligns with and extends conceptual

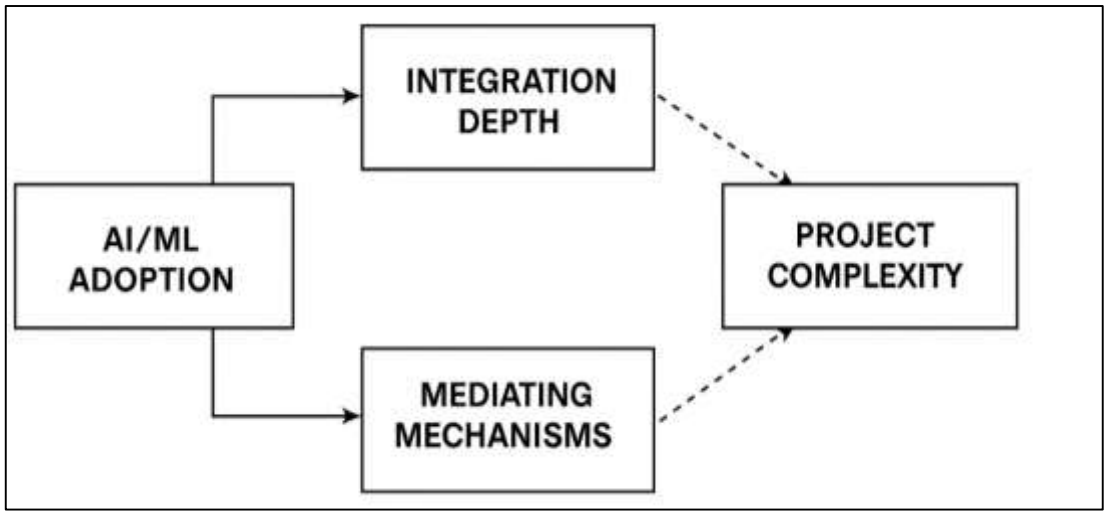
models of BIM/analytics maturity (Khosrowshahi & Arayici, 2012) by explicitly weighting the features that turned out to matter most in empirical associations. The moderation by complexity suggests a boundary condition: AI/ML contributes disproportionately where coordination graphs are denser and variance propagation is faster, offering a testable mechanism that links workface topology to analytics returns (Han & Golparvar-Fard, 2015). Our results also speak to risk informatics theory: detection gains and severity reductions are consistent with the premise that leading indicators extracted from text, vision, and telematics must be mapped to BIM elements and schedule states to become actionable (Zhang, Teizer, et al., 2015). For cost, the portfolio-level attenuation of overruns when adoption is higher echoes early neural/CBR findings yet reframes them as governance effects, where estimator architecture and decomposition enable change-aware control (Petroutsatou et al., 2012). Finally, by showing that digital maturity remains an independent correlate even after controlling for adoption, we motivate a two-factor theoretical model in which capability (people/process/IT) and usage (embedded pipelines) jointly shape performance consistent with socio-technical perspectives of construction IT diffusion (Bryde et al., 2013).

Several limitations temper interpretation. First, the cross-sectional design cannot establish causality; unobserved organizational traits (leadership, culture) may co-vary with adoption and outcomes despite our controls an endemic challenge in multi-firm field research (Bryde et al., 2013). Second, the Adoption Index, though artifact-anchored and double-scored, remains a composite; different weighting philosophies could shift scores, even if robustness checks showed stability. Third, outcome measures such as percentage delay/overrun are portfolio-comparable but may compress nuance (e.g., contingency release timing, beneficial re-sequencing) and can be influenced by re-baselining practices that we screened but cannot fully standardize (Kim et al., 2004). Fourth, risk measures rely on governance records; improvements in detection rate could partly reflect better documentation, not just better foresight, though the concurrent decline in residual severity mitigates that concern (Hallowell & Gambatese, 2009). Fifth, selection bias is possible: participating firms already maintain digital ecosystems (BIM/ERP/scheduling), which may limit generalizability to organizations earlier in digitization (Succar, 2009). Sixth, clustering by firm compresses effective sample size, and while cluster-robust/FE models preserved inference, degrees of freedom constrain rich fixed-effects designs. Finally, measurement error may persist in mapping heterogeneous calendars, currencies, and coding schemes despite our ETL harmonization; although diagnostics and validation checks were extensive, residual inconsistencies are plausible. These limitations situate the contribution: we provide associative evidence at scale that complements single-project experiments and algorithmic benchmarks by centering pipeline integration and cadence as explanatory levers.

Future work should pursue causal identification and granularity. Longitudinal or panel designs that observe projects from preconstruction through commissioning can leverage staggered AI/ML deployments for difference-in-differences or synthetic control estimators. Where policy or contractual changes mandate interoperability or pipeline automation, quasi-experiments could sharpen causal claims about integration depth. Randomized workflow interventions e.g., assigning standardized model-review rituals to some projects might be feasible and ethically acceptable. Second, mechanism testing is ripe: mediation models could quantify how much of adoption's association with outcomes is carried by reduced decision latency, earlier variance detection, or improved change classification. Third, broaden risk informatics beyond struck-by and falls to include permit, environmental, and logistics risks, integrating telematics and supply-chain signals with 4D/BIM (Marks & Teizer, 2013). Fourth, expand cost analytics to marry ML forecasts with value-engineering option spaces and to evaluate how estimator decomposition interacts with procurement models (Adeli & Wu, 1998). Fifth, investigate fairness, privacy, and security in site sensing and NLP areas where CISOs must balance analytic utility with ethical constraints by testing anonymization and access-control regimes without degrading predictive value (Volk et al., 2014). Sixth, articulate learning curves and total cost of ownership for integration: what cadence, staffing, and platform choices minimize time-to-value at different maturity levels? Finally, pursue cross-country comparisons and contract-type interactions to refine boundary conditions observed here; for example, does integration depth matter more under DB than DBB due to design-construction concurrency, and do regulatory environments modulate risk-

analytics utility? By shifting analytics research toward pipeline-aware, governance-linked designs, the field can progress from promising pilots to reliable, generalizable playbooks for project delivery.

Figure 8: Proposed Model for future study



CONCLUSION

This study set out to quantify how the adoption of artificial intelligence and machine learning (AI/ML) relates to core construction project management outcomes schedule adherence, cost estimation accuracy, and risk mitigation effectiveness using a cross-sectional, multi-case design with the project as the unit of analysis. By operationalizing a project-level AI/ML Adoption Index that captures use-case breadth, frequency of application, model sophistication, and critically integration depth with BIM/ERP/scheduling systems, and by pairing that index with standardized outcomes (percentage schedule delay and cost overrun, risk detection rate, and residual risk impact), we provide portfolio-scale, evidence-based insight into where data-driven practices are most strongly associated with measurable performance. Across descriptive, correlational, and regression analyses with robust errors, diagnostics, and extensive sensitivity checks the findings are consistent and practically meaningful: higher adoption corresponds to lower average delay and overrun, higher upstream risk detection, and lower residual consequence, with effects that are amplified on more complex projects. These results hold after accounting for digital maturity, project size and duration, sector, contract type, and region, and they remain stable under alternative model families, index weightings, and outlier treatments, underscoring that the signal reflects embedded pipeline quality rather than any single algorithm or one-off pilot. The study's contribution is twofold. Empirically, it translates decades of algorithm-centric demonstrations into portfolio-level associations with the KPIs that matter to owners, contractors, and lenders, showing that cadence and integration the routine review of model outputs in look-ahead meetings, automatic propagation of updates back into control environments, and alignment with work packages and hazard libraries are the operational levers by which AI/ML becomes managerial value. Conceptually, it advances a pipeline-centric perspective in which organizational capability (people, process, IT) and embedded usage (integrated, cadence-based pipelines) jointly shape outcomes, and in which project complexity moderates returns by creating more opportunities for analytics to intercept variance. At the same time, we acknowledge limitations inherent to cross-sectional, multi-firm research: residual confounding is possible; composite scoring, even when artifact-anchored and double-rated, may not capture every nuance; governance practices like re-baselining and documentation can influence recorded outcomes; and firms willing to participate generally possess baseline digital ecosystems that may not represent the entire industry. These boundaries frame, rather than diminish, the central message: when organizations elevate integration depth and cadence from "nice-to-have" to "standard operating procedure," AI/ML's benefits express themselves in the very indicators by which projects are judged. Practically, this points to actionable priorities API-first interoperability, versioned feature stores, automated quality checks, and ritualized model reviews that can be specified in

contracts, funded in programs, and governed through clear roles and audit trails. Theoretically, it motivates longitudinal and quasi-experimental work that follows deployments over time to sharpen causal inference and unbundle mechanisms such as decision latency, variance detection, and change-order triage. In sum, by bridging technical promise and managerial practice, the study offers a reproducible measurement framework and a set of evidence-aligned priorities for teams seeking to convert AI/ML from disparate tools into reliable, repeatable performance gains across schedules, budgets, and risk profiles.

RECOMMENDATIONS

Organizations seeking to translate AI/ML from isolated pilots into reliable schedule, cost, and risk improvements should prioritize a pipeline-first operating model anchored in governance, integration, cadence, and capability. First, formalize data governance under CISO and project-controls leadership: map every model feature to a system of record (BIM, scheduling, ERP, safety), assign data stewards, enforce role-based access, encrypt at rest/in transit, and maintain audit logs; publish a lightweight data catalog and retention policy so teams know what exists, who owns it, and how it flows. Second, invest in integration depth over tool count: require API-level interoperability that writes predictions back into Primavera/MSP and cost systems, create a versioned feature store (with schemas for tabular, text, and vision data), and automate ETL with unit tests, quality checks, and drift monitors; avoid spreadsheet hand-offs except as temporary fallbacks. Third, institutionalize cadence: make "model review" a standing agenda item in weekly look-ahead and monthly cost meetings; define SLAs for model refresh (e.g., schedule forecast every 7 days, cost roll every 30, risk/NLP scan daily), and trigger alerts when cadences slip. Fourth, align analytics with work packages and control accounts: require forecasts and risk alerts at the same granularity as responsibility matrices so every variance has a named owner and a dated action. Fifth, target high-complexity projects first dense interfaces, constrained access, or novel methods because marginal gains are larger; fund integration and automation from program contingencies and document lift via pre/post metrics. Sixth, modernize safety analytics into an operational product: pair proximity/CV detectors with BIM-linked hazard libraries and predefined mitigations, set threshold-based actions (e.g., automatic edge protection work orders), and review leading indicators alongside TRIR in executive dashboards. Seventh, embed MLOps discipline: track dataset and model versions, promote through dev/stage/prod environments, monitor performance and data drift, and establish rollback procedures; appoint a small cross-functional "pipeline guild" (controls, VDC, IT) accountable for reliability. Eighth, upgrade contracts and procurement: codify interoperability (open formats, API access), minimum logging, and delivery of data artifacts (feature schemas, model cards); evaluate vendors on integration effort, not demo accuracy. Ninth, invest in people and process: run role-specific training (superintendents on interpretation, engineers on data lineage, PMs on decision thresholds), create SOPs that tie forecasts to actions (e.g., resequencing rules, contingency drawdown triggers), and recognize teams for adherence to cadence and variance response not just end outcomes. Tenth, implement a measurement plan: adopt the AI/ML Adoption Index as an internal KPI (quarterly), set portfolio targets (e.g., +10 points YoY), and link to practical metrics (Δ delay/overrun, ↑ detection, ↓ severity); publish a quarterly "analytics health" report with win/loss narratives so lessons propagate. Eleventh, safeguard privacy and ethics: blur/anonymize site imagery where feasible, minimize PII in RFIs and incident narratives, and convene a review board for new sensors or NLP uses. Finally, use a phased rollout: 90-day pilot to prove cadence and integration on two projects, 6-month scale-up across a program, then enterprise standardization; at each gate, validate outcomes against minimum practically important differences and refine rubrics favoring durable pipeline refinements over swapping algorithms so AI/ML becomes an everyday control lever rather than an occasional experiment

REFERENCES

[1]. Abdur Razzak, C., Golam Qibria, L., & Md Arifur, R. (2024). Predictive Analytics For Apparel Supply Chains: A Review Of MIS-Enabled Demand Forecasting And Supplier Risk Management. *American Journal of Interdisciplinary Studies*, *5*(04), 01–23. https://doi.org/10.63125/80dwy222

[2]. Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938. https://doi.org//10.1016/j.heliyon.2018.e00938

[3]. Adeli, H., & Wu, M. (1998). Regularization neural network for construction cost estimation. *Computer-Aided Civil and Infrastructure Engineering*, *13*(5), 311–320. https://doi.org//10.1111/0885-9507.00104

[4]. Ahiaga-Dagbui, D., & Smith, S. D. (2014). Rethinking construction cost overruns: Cognition, learning and estimation. *Journal of Financial Management of Property and Construction*, *19*(1), 38–54. https://doi.org//10.1108/jfmpc-06-2013-0029

[5]. Aibinu, A. A., & Jagboro, G. O. (2002). The effects of construction delays on project delivery in Nigerian construction industry. *International Journal of Project Management*, *20*(8), 593–599. https://doi.org//10.1016/s0263-7863(02)00028-5

[6]. Boussabaine, A. H., & Elhag, T. M. S. (1999). Tender price estimation using artificial neural networks. *Journal of Construction Engineering and Management*, *125*(1), 79–86. https://doi.org//10.1061/(asce)0733-9364(1999)125:1(79)

[7]. Bryde, D., Broquetas, M., & Volm, J. M. (2013). The project benefits of Building Information Modelling (BIM). *Automation in Construction*, *31*, 97–109. https://doi.org//10.1016/j.autcon.2012.07.018

[8]. Chan, D. W. M., & Kumaraswamy, M. M. (1997). A comparative study of causes of time overruns in Hong Kong construction projects. *International Journal of Project Management*, *15*(1), 55–63. https://doi.org//10.1016/s0263-7863(96)00039-7

[9]. Chen, P.-H., & Shahandashti, S. M. (2009). Hybrid of genetic algorithm and simulated annealing for multiple project scheduling with multiple resource constraints. *Automation in Construction*, *18*(4), 434–443. https://doi.org//10.1016/j.autcon.2008.10.007

[10]. Cheng, M.-Y., Tsai, H.-C., & Sudjono, E. (2010). Conceptual cost estimates using evolutionary fuzzy hybrid neural network for projects in construction. *Automation in Construction*, *19*(5), 619–629. https://doi.org//10.1016/j.autcon.2010.02.002

[11]. Chou, J.-S., & Pham, A. D. (2013). Improved artificial intelligence for engineering cost estimation of construction projects. *Journal of Civil Engineering and Management*, *19*(S1), S59–S68. https://doi.org//10.3846/13923730.2013.795187

[12]. Danish, M., & Md. Zafor, I. (2022). The Role Of ETL (Extract-Transform-Load) Pipelines In Scalable Business Intelligence: A Comparative Study Of Data Integration Tools. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *2*(1), 89–121. https://doi.org/10.63125/1spa6877

[13]. Danish, M., & Md. Zafor, I. (2024). Power BI And Data Analytics In Financial Reporting: A Review Of Real-Time Dashboarding And Predictive Business Intelligence Tools. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 125-157. https://doi.org/10.63125/yg9zxt61

[14]. Danish, M., & Md.Kamrul, K. (2022). Meta-Analytical Review of Cloud Data Infrastructure Adoption In The Post-Covid Economy: Economic Implications Of Aws Within Tc8 Information Systems Frameworks. *American Journal of Interdisciplinary Studies*, *3*(02), 62-90. https://doi.org/10.63125/1eg7b369

[15]. Egwim, C. N. (2021). Applied artificial intelligence for predicting construction projects delay. *Machine Learning with Applications*, *6*, 100166. https://doi.org//10.1016/j.mlwa.2021.100166

[16]. El-Rayes, K., & Moselhi, O. (2001). Optimizing resource utilization for repetitive construction projects. *Journal of Construction Engineering and Management*, *127*(1), 18–27. https://doi.org//10.1061/(asce)0733-9364(2001)127:1(18)

[17]. Elbeltagi, E., Hegazy, T., & Grierson, D. (2005). Comparison among five evolutionary-based optimization algorithms. *Advanced Engineering Informatics*, *19*(1), 43–53. https://doi.org//10.1016/j.aei.2005.01.004

[18]. Fang, W., Cho, Y. K., & Zhang, S. (2020). Real-time safety risk assessment for mobile crane lifts using computer vision and deep learning. *Automation in Construction*, *110*, 103016. https://doi.org//10.1016/j.autcon.2019.103016

[19]. Flyvbjerg, B., Holm, M. K. S., & Buhl, S. (2002). Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association*, *68*(3), 279–295. https://doi.org//10.1080/01944360208976273

[20]. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. https://doi.org//10.7551/mitpress/9780262035613.001.0001

[21]. Hallowell, M. R., & Gambatese, J. A. (2009). Construction safety risk mitigation. *Journal of Construction Engineering and Management*, *135*(12), 1316–1323. https://doi.org//10.1061/(asce)co.1943-7862.0000107

[22]. Han, K. K., & Golparvar-Fard, M. (2015). Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs. *Automation in Construction*, *53*, 44–57. https://doi.org//10.1016/j.autcon.2015.02.007

[23]. Hegazy, T. (1999). Optimization of resource allocation and leveling using genetic algorithms. *Journal of Construction Engineering and Management*, *125*(3), 167–175. https://doi.org//10.1061/(asce)0733-9364(1999)125:3(167)

[24]. Hegazy, T., & Ayed, A. (1998). Neural network model for parametric cost estimation of highway projects. *Journal of Construction Engineering and Management*, *124*(3), 210–218. https://doi.org//10.1061/(asce)0733-9364(1998)124:3(210)

[25]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2023). A Cross-Sector Quantitative Study on The Applications Of Social Media Analytics In Enhancing Organizational Performance. *American Journal of Scholarly Research and Innovation*, 2(02), 274-302. https://doi.org/10.63125/d8ree044

[26]. Istiaque, M., Dipon Das, R., Hasan, A., Samia, A., & Sayer Bin, S. (2024). Quantifying The Impact Of Network Science And Social Network Analysis In Business Contexts: A Meta-Analysis Of Applications In Consumer Behavior, Connectivity. *International Journal of Scientific Interdisciplinary Research*, *5*(2), 58-89. https://doi.org/10.63125/vgkwe938

[27]. Jahid, M. K. A. S. R. (2022). Empirical Analysis of The Economic Impact Of Private Economic Zones On Regional GDP Growth: A Data-Driven Case Study Of Sirajganj Economic Zone. *American Journal of Scholarly Research and Innovation*, *1*(02), 01-29. https://doi.org/10.63125/je9w1c40

[28]. Ji, S., Kim, Y., & Choi, I. (2010). Case-based reasoning for cost estimation of residential building projects. *Journal of Construction Engineering and Management*, *136*(1), 110–118. https://doi.org//10.1061/(asce)co.1943-7862.0000104

[29]. Kanan, R., Elhassan, O., & Bensalem, R. (2018). An IoT-based autonomous system for workers' safety in construction sites with real-time alarming, monitoring, and positioning strategies. *Automation in Construction*, *88*, 73–86. https://doi.org//10.1016/j.autcon.2017.12.033

[30]. Khosrowshahi, F., & Arayici, Y. (2012). Roadmap for implementation of BIM in the UK construction industry. *Engineering, Construction and Architectural Management*, *19*(6), 610–635. https://doi.org//10.1108/09699981211277531

[31]. Kim, G.-H., An, S.-H., & Kang, K.-I. (2004). Comparison of construction cost estimating models based on regression analysis and neural networks. *Building and Environment*, *39*(10), 1235–1242. https://doi.org//10.1016/j.buildenv.2004.02.013

[32]. Kim, S., Kim, H., & Kim, H. (2013). An approach to activity duration prediction for construction project schedules using proximity to critical path and fuzzy inference. *Automation in Construction*, *35*, 176–188. https://doi.org//10.1016/j.autcon.2013.05.020

[33]. Koo, B., & Fischer, M. (2000). Feasibility study of 4D CAD in commercial construction. *Journal of Construction Engineering and Management*, *126*(4), 251–260. https://doi.org//10.1061/(asce)0733-9364(2000)126:4(251)

[34]. Leu, S.-S., & Yang, C.-H. (1999). GA-based multicriteria optimal model for construction scheduling. *Journal of Construction Engineering and Management*, *125*(6), 420–427. https://doi.org//10.1061/(asce)0733-9364(1999)125:6(420)

[35]. Marks, E. D., & Teizer, J. (2013). Method for testing proximity detection and alert technology for safe construction equipment operation. *Construction Management and Economics*, *31*(6), 636–646. https://doi.org//10.1080/01446193.2013.783705

[36]. McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60–68. https://doi.org//10.5437/08956308x5502003

[37]. Md Arifur, R., & Sheratun Noor, J. (2022). A Systematic Literature Review of User-Centric Design In Digital Business Systems: Enhancing Accessibility, Adoption, And Organizational Impact. *Review of Applied Science and Technology*, *1*(04), 01-25. https://doi.org/10.63125/ndjkpm77

[38]. Md Ashiqur, R., Md Hasan, Z., & Afrin Binta, H. (2025). A meta-analysis of ERP and CRM integration tools in business process optimization. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 278-312. https://doi.org/10.63125/yah70173

[39]. Md Hasan, Z. (2025). AI-Driven business analytics for financial forecasting: a systematic review of decision support models in SMES. *Review of Applied Science and Technology*, *4*(02), 86-117. https://doi.org/10.63125/gjrpv442

[40]. Md Hasan, Z., Mohammad, M., & Md Nur Hasan, M. (2024). Business Intelligence Systems In Finance And Accounting: A Review Of Real-Time Dashboarding Using Power BI & Tableau. *American Journal of Scholarly Research and Innovation*, *3*(02), 52-79. https://doi.org/10.63125/fy4w7w04

[41]. Md Hasan, Z., & Moin Uddin, M. (2022). Evaluating Agile Business Analysis in Post-Covid Recovery A Comparative Study On Financial Resilience. *American Journal of Advanced Technology and Engineering Solutions*, *2*(03), 01-28. https://doi.org/10.63125/6nee1m28

[42]. Md Hasan, Z., Sheratun Noor, J., & Md. Zafor, I. (2023). Strategic role of business analysts in digital transformation tools, roles, and enterprise outcomes. *American Journal of Scholarly Research and Innovation*, *2*(02), 246-273. https://doi.org/10.63125/rc45z918

[43]. Md Ismail, H., Md Mahfuj, H., Mohammad Aman Ullah, S., & Shofiul Azam, T. (2025). Implementing Advanced Technologies For Enhanced Construction Site Safety. *American Journal of Advanced Technology and Engineering Solutions*, *1*(02), 01-31. https://doi.org/10.63125/3v8rpr04

[44]. Md Ismail Hossain, M. A. B., amp, & Mousumi Akter, S. (2023). Water Quality Modelling and Assessment Of The Buriganga River Using Qual2k. *Global Mainstream Journal of Innovation, Engineering & Emerging Technology*, *2*(03), 01-11. https://doi.org/10.62304/jieet.v2i03.64

[45]. Md Jakaria, T., Md, A., Zayadul, H., & Emdadul, H. (2025). Advances In High-Efficiency Solar Photovoltaic Materials: A Comprehensive Review Of Perovskite And Tandem Cell Technologies. *American Journal of Advanced Technology and Engineering Solutions*, *1*(01), 201-225. https://doi.org/10.63125/5amnvb37

[46]. Md Mahamudur Rahaman, S. (2022a). Electrical And Mechanical Troubleshooting in Medical And Diagnostic Device Manufacturing: A Systematic Review Of Industry Safety And Performance Protocols. *American Journal of Scholarly Research and Innovation*, *1*(01), 295-318. https://doi.org/10.63125/d68y3590

[47]. Md Mahamudur Rahaman, S. (2022b). Smart Maintenance in Medical Imaging Manufacturing: Towards Industry 4.0 Compliance at Chronos Imaging. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *2*(1), 29–62. https://doi.org/10.63125/eatsmf47

[48]. Md Mahamudur Rahaman, S. (2024). AI-Driven Predictive Maintenance For High-Voltage X-Ray Ct Tubes: A Manufacturing Perspective. *Review of Applied Science and Technology*, *3*(01), 40-67. https://doi.org/10.63125/npwqxp02

[49]. Md Mahamudur Rahaman, S., & Rezwanul Ashraf, R. (2022). Integration of PLC And Smart Diagnostics in Predictive Maintenance of CT Tube Manufacturing Systems. *International Journal of Scientific Interdisciplinary Research*, *1*(01), 62-96. https://doi.org/10.63125/gspb0f75

[50]. Md Mahamudur Rahaman, S., & Rezwanul Ashraf, R. (2023). Applying Lean And Six Sigma In The Maintenance Of Medical Imaging Equipment Manufacturing Lines. *Review of Applied Science and Technology*, *2*(04), 25-53. https://doi.org/10.63125/6varjp35

[51]. Md Nazrul Islam, K. (2022). A Systematic Review of Legal Technology Adoption In Contract Management, Data Governance, And Compliance Monitoring. *American Journal of Interdisciplinary Studies*, *3*(01), 01-30. https://doi.org/10.63125/caangg06

[52]. Md Nur Hasan, M. (2024). Integration Of Artificial Intelligence And DevOps In Scalable And Agile Product Development: A Systematic Literature Review On Frameworks. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *4*(1), 01–32. https://doi.org/10.63125/exyqj773

[53]. Md Nur Hasan, M. (2025). Role Of AI And Data Science In Data-Driven Decision Making For It Business Intelligence: A Systematic Literature Review. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 564-588. https://doi.org/10.63125/n1xpym21

[54]. Md Nur Hasan, M., Md Musfiqur, R., & Debashish, G. (2022). Strategic Decision-Making in Digital Retail Supply Chains: Harnessing AI-Driven Business Intelligence From Customer Data. *Review of Applied Science and Technology*, *1*(03), 01-31. https://doi.org/10.63125/6a7rpy62

[55]. Md Redwanul, I., & Md. Zafor, I. (2022). Impact of Predictive Data Modeling on Business Decision-Making: A Review Of Studies Across Retail, Finance, And Logistics. *American Journal of Advanced Technology and Engineering Solutions*, *2*(02), 33-62. https://doi.org/10.63125/8hfbkt70

[56]. Md Rezaul, K., & Md Mesbaul, H. (2022). Innovative Textile Recycling and Upcycling Technologies For Circular Fashion: Reducing Landfill Waste And Enhancing Environmental Sustainability. *American Journal of Interdisciplinary Studies*, *3*(03), 01-35. https://doi.org/10.63125/kkmerg16

[57]. Md Sultan, M., Proches Nolasco, M., & Md. Torikul, I. (2023). Multi-Material Additive Manufacturing For Integrated Electromechanical Systems. *American Journal of Interdisciplinary Studies*, *4*(04), 52-79. https://doi.org/10.63125/y2ybrx17

[58]. Md Sultan, M., Proches Nolasco, M., & Vicent Opiyo, N. (2025). A Comprehensive Analysis Of Non-Planar Toolpath Optimization In Multi-Axis 3D Printing: Evaluating The Efficiency Of Curved Layer Slicing Strategies. *Review of Applied Science and Technology*, *4*(02), 274-308. https://doi.org/10.63125/5fdxa722

[59]. Md Takbir Hossen, S., Ishtiaque, A., & Md Atiqur, R. (2023). AI-Based Smart Textile Wearables For Remote Health Surveillance And Critical Emergency Alerts: A Systematic Literature Review. *American Journal of Scholarly Research and Innovation*, *2*(02), 1-29. https://doi.org/10.63125/ceqapd08

[60]. Md Tawfiqul, I. (2023). A Quantitative Assessment Of Secure Neural Network Architectures For Fault Detection In Industrial Control Systems. *Review of Applied Science and Technology*, *2*(04), 01-24. https://doi.org/10.63125/3m7gbs97

[61]. Md. Sakib Hasan, H. (2022). Quantitative Risk Assessment of Rail Infrastructure Projects Using Monte Carlo Simulation And Fuzzy Logic. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 55-87. https://doi.org/10.63125/h24n6z92

[62]. Md. Tarek, H. (2022). Graph Neural Network Models For Detecting Fraudulent Insurance Claims In Healthcare Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(01), 88-109. https://doi.org/10.63125/r5vsmv21

[63]. Md. Zafor, I. (2025). A Meta-Analysis Of AI-Driven Business Analytics: Enhancing Strategic Decision-Making In SMEs. *Review of Applied Science and Technology*, *4*(02), 33-58. https://doi.org/10.63125/wk9fqv56

[64]. Md.Kamrul, K., & Md Omar, F. (2022). Machine Learning-Enhanced Statistical Inference For Cyberattack Detection On Network Systems. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 65-90. https://doi.org/10.63125/sw7jzx60

[65]. Md.Kamrul, K., & Md. Tarek, H. (2022). A Poisson Regression Approach to Modeling Traffic Accident Frequency in Urban Areas. *American Journal of Interdisciplinary Studies*, *3*(04), 117-156. https://doi.org/10.63125/wqh7pd07

[66]. Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill. https://doi.org//10.5555/541177

[67]. Moin Uddin, M. (2025). Impact Of Lean Six Sigma On Manufacturing Efficiency Using A Digital Twin-Based Performance Evaluation Framework. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 343-375. https://doi.org/10.63125/z70nhf26

[68]. Moin Uddin, M., & Rezwanul Ashraf, R. (2023). Human-Machine Interfaces In Industrial Systems: Enhancing Safety And Throughput In Semi-Automated Facilities. *American Journal of Interdisciplinary Studies*, *4*(01), 01-26. https://doi.org/10.63125/s2qa0125

[69]. Momena, A., & Md Nur Hasan, M. (2023). Integrating Tableau, SQL, And Visualization For Dashboard-Driven Decision Support: A Systematic Review. *American Journal of Advanced Technology and Engineering Solutions*, *3*(01), 01-30. https://doi.org/10.63125/4aa43m68

[70]. Mubashir, I., & Abdul, R. (2022). Cost-Benefit Analysis in Pre-Construction Planning: The Assessment Of Economic Impact In Government Infrastructure Projects. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 91-122. https://doi.org/10.63125/kjwd5e33

[71]. Omar Muhammad, F., & Md.Kamrul, K. (2022). Blockchain-Enabled BI For HR And Payroll Systems: Securing Sensitive Workforce Data. *American Journal of Scholarly Research and Innovation*, *1*(02), 30-58. https://doi.org/10.63125/et4bhy15

[72]. Petroutsatou, K., Georgiou, P., Lambropoulos, S., & Pantouvakis, J. P. (2012). Early cost estimating of road construction projects using case-based reasoning. *Journal of Construction Engineering and Management*, *138*(2), 146–157. https://doi.org//10.1061/(asce)co.1943-7862.0000410

[73]. Reduanul, H., & Mohammad Shoeb, A. (2022). Advancing AI in Marketing Through Cross Border Integration Ethical Considerations And Policy Implications. *American Journal of Scholarly Research and Innovation*, *1*(01), 351-379. https://doi.org/10.63125/d1xg3784

[74]. Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach (3rd ed.)*. Prentice Hall. https://doi.org//10.5555/1671231

[75]. Sabuj Kumar, S., & Zobayer, E. (2022). Comparative Analysis of Petroleum Infrastructure Projects In South Asia And The Us Using Advanced Gas Turbine Engine Technologies For Cross Integration. *American Journal of Advanced Technology and Engineering Solutions*, *2*(04), 123-147. https://doi.org/10.63125/wr93s247

[76]. Sadia, T., & Shaiful, M. (2022). In Silico Evaluation of Phytochemicals From Mangifera Indica Against Type 2 Diabetes Targets: A Molecular Docking And Admet Study. *American Journal of Interdisciplinary Studies*, *3*(04), 91-116. https://doi.org/10.63125/anaf6b94

[77]. Sanjai, V., Sanath Kumar, C., Maniruzzaman, B., & Farhana Zaman, R. (2023). Integrating Artificial Intelligence in Strategic Business Decision-Making: A Systematic Review Of Predictive Models. *International Journal of Scientific Interdisciplinary Research*, *4*(1), 01-26. https://doi.org/10.63125/s5skge53

[78]. Sanjai, V., Sanath Kumar, C., Sadia, Z., & Rony, S. (2025). AI And Quantum Computing For Carbon-Neutral Supply Chains: A Systematic Review Of Innovations. *American Journal of Interdisciplinary Studies*, *6*(1), 40-75. https://doi.org/10.63125/nrdx7d32

[79]. Sheratun Noor, J., & Momena, A. (2022). Assessment Of Data-Driven Vendor Performance Evaluation in Retail Supply Chains: Analyzing Metrics, Scorecards, And Contract Management Tools. *American Journal of Interdisciplinary Studies*, *3*(02), 36-61. https://doi.org/10.63125/0s7t1y90

[80]. Succar, B. (2009). Building information modelling framework: A research and delivery foundation for industry stakeholders. *Automation in Construction*, *18*(3), 357–375. https://doi.org//10.1016/j.autcon.2008.10.003

[81]. Tahmina Akter, R., Debashish, G., Md Soyeb, R., & Abdullah Al, M. (2023). A Systematic Review of AI-Enhanced Decision Support Tools in Information Systems: Strategic Applications In Service-Oriented Enterprises And Enterprise Planning. *Review of Applied Science and Technology*, *2*(01), 26-52. https://doi.org/10.63125/73djw422

[82]. Tixier, A. J.-P., Hallowell, M. R., Rajagopalan, B., & Bowman, D. (2016). Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. *Automation in Construction*, *62*, 45–56. https://doi.org//10.1016/j.autcon.2015.11.001

[83]. Turkan, Y., Bosché, F., Haas, C. T., & Haas, R. (2012). Automated progress tracking using 4D schedule and 3D sensing technologies. *Automation in Construction*, *22*, 414–421. https://doi.org//10.1016/j.autcon.2011.10.003

[84]. Volk, R., Stengel, J., & Schultmann, F. (2014). Building Information Modeling (BIM) for existing buildings — Literature review and future needs. *Automation in Construction*, *38*, 109–127. https://doi.org//10.1016/j.autcon.2013.10.023

[85]. Zhang, S., Sulankivi, K., Kiviniemi, M., Romo, I., Eastman, C. M., & Teizer, J. (2015). BIM-based fall hazard identification and prevention in construction safety planning. *Safety Science*, *72*, 31–45. https://doi.org//10.1016/j.ssci.2014.08.001

[86]. Zhang, S., Teizer, J., Lee, J.-K., Eastman, C. M., & Venugopal, M. (2013). Building Information Modeling (BIM) and safety: Automatic safety checking of construction models and schedules. *Automation in Construction*, *29*, 183–195. https://doi.org//10.1016/j.autcon.2012.05.006

[87]. Zhang, S., Teizer, J., Pradhananga, N., & Eastman, C. M. (2015). Proximity hazard indicator for workers-on-foot near miss interactions with construction equipment and geo-referenced hazard areas. *Automation in Construction*, *60*, 58–73. https://doi.org//10.1016/j.autcon.2015.09.003

[88]. Zou, Y., Kiviniemi, A., & Jones, S. W. (2017). Retrieving similar cases for construction project risk management using natural language processing techniques. *Automation in Construction*, *80*, 66–76. https://doi.org//10.1016/j.autcon.2017.04.003