ASRC Procedia
Global Perspectives in Science and Scholarship
Florida, USA

Volume: 3; **Issue:** 1
Pages: 62–93
Published: 29 May 2023

open access

World Conference on Scientific Discovery and Innovation 2023,
May 24–26, 2023, Florida, USA

AI-POWERED BI DASHBOARDS IN OPERATIONS: A COMPARATIVE ANALYSIS FOR REAL-TIME DECISION SUPPORT

Razia Sultana[1];

[1]. Metronet Bangladesh Limited, Dhaka, Bangladesh; Email: ronjitaborna07@gmail.com

Doi: 10.63125/wqd2t159s

Peer-review under responsibility of the organizing committee of WCSDI, 2023

Abstract

This study addresses a practical problem in operations: decision latency and variable decision quality when teams rely on visually dense dashboards under high data volume, velocity, and variability. The purpose is to quantify how AI-integrated business intelligence dashboards support real-time decision making across organizations. Using a quantitative, cross-sectional, case-based design, we analyze six production cloud and enterprise cases in manufacturing, logistics, healthcare operations, tech-enabled services, retail fulfillment, and utilities, with 168 active users as respondents. Key variables include an AI Integration Index (forecasting, anomaly detection, prescriptive recommendations, natural-language interaction, explainability), user perceptions (perceived usefulness, interpretability, trust, workload), outcomes (decision latency, decision accuracy or confidence), and contextual controls (data quality, dashboard tenure, organization size, training, analytics proficiency). The analysis plan combines descriptive statistics and correlations with multivariate regression using HC3-robust errors and case fixed effects, mediation tests via bootstrap for perceived usefulness, and moderation tests for data quality. Headline findings show that higher AI integration is associated with materially faster decisions and higher confidence, with perceived usefulness transmitting much of the effect on confidence and dependable data quality strengthening the speed benefits; interface workload relates to slower action. Implications are concrete for architects and managers: prioritize pipeline timeliness and semantic clarity, expose compact on-demand explanations and uncertainty cues, control alert and visual clutter, and connect predictions to guarded prescriptive actions. The literature review synthesizes 57 peer-reviewed papers to ground constructs, measures, and mechanisms used in this comparative evaluation.

Keywords

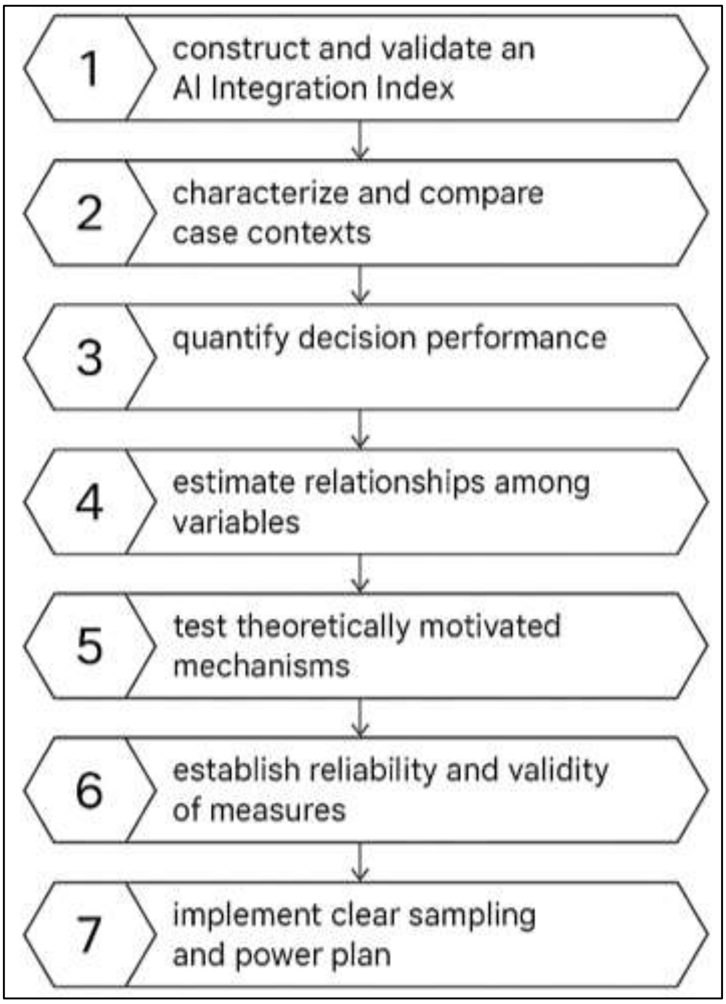
AI-Integrated BI Dashboards, Real-Time Decision Support, Cross-Sectional Multi-Case Study, AI Integration Index, Perceived Usefulness,

INTRODUCTION

Business intelligence (BI) dashboards are consolidated visual interfaces that assemble key performance indicators, alerts, and analytic views to support managerial sense-making and action (Sarikaya et al., 2019; Trieu, 2017). In contemporary operations spanning manufacturing, logistics, healthcare, and public utilities dashboards serve as the primary human–data touchpoint for monitoring flows, bottlenecks, service quality, and asset utilization at scale across multiple geographies and time zones (Pauwels et al., 2009). Artificial intelligence (AI)–integrated BI dashboards extend this function by embedding machine learning (ML) components for prediction, classification, anomaly detection, and natural-language explanation or querying, thereby coupling descriptive and predictive layers for timely, higher-fidelity decisions (Chen et al., 2012). Real-time decision support here is defined as the capacity to ingest, process, and surface actionable analytics with latencies low enough to influence ongoing operational control often seconds to minutes an affordance enabled by data-stream architectures and streaming computation (Babcock et al., 2002; Chen et al., 2012). At a global scale, the diffusion of such dashboards is tied to competitiveness, resilience, and compliance: the ability to synchronize suppliers, plants, and last-mile distribution while meeting service-level agreements, safety standards, and energy-efficiency targets (Chae et al., 2014). These definitions situate AI-integrated dashboards within a broader analytics capability stack where data quality, model performance, and visual cognition intersect with organizational routines framing the need for rigorous, comparative, and quantitative assessment across operational contexts. (Chae et al., 2014; Chen et al., 2012; Pauwels et al., 2009). The practical problem motivating this study centers on decision latency and decision quality under high volume, velocity, and variability. Operational teams contend with rapidly updating signals (orders, machine states, logistics events), heterogeneous data provenance, and visually dense dashboards that can either clarify or overload attention (Sivarajah et al., 2017). On the analytic side, AI components can improve detection and forecasting, yet their opaque behavior challenges user trust, particularly when stakes are high and remediation windows are narrow (Miller, 2019). Explainability mechanisms such as model-agnostic local explanations (e.g., LIME) and uncertainty cues are therefore salient design elements within dashboards to calibrate reliance (Ribeiro et al., 2016). From an adoption standpoint, perceptions of usefulness and ease of use remain central antecedents of intention to use and usage behavior, especially for knowledge workers under time pressure (Davis, 1989). In operational settings, appropriate reliance on AI recommendations requires calibrated trust the alignment of trust with system capability and context shaped by interface characteristics and performance histories (Lee & See, 2004). BI-capability research also underscores foundational determinants such as data quality, integration, access, and analytical culture for achieving BI success (Isik et al., 2013; Popovič et al., 2012). These threads converge on a measurement challenge: to quantify how AI-integrated dashboards perform in real-time operations and how human factors, information quality, and visualization design jointly condition outcomes. (Kitchin, 2014; Pauwels et al., 2009).

Within operations management, the analytic tasks surfaced through dashboards are varied queue length monitoring, throughput balancing, inventory positioning, predictive maintenance scheduling, and service exception triage. Predictive models embedded in dashboards can anticipate stockouts, flag abnormal sensor signatures, prioritize orders, and propose set-point adjustments; these functions bear directly on cost, quality, delivery reliability, and flexibility (Chae et al., 2014). Anomaly-detection research provides the methodological substrate for early signal capture under nonstationary conditions and sparse labels (Chandola et al., 2009; Sultan et al., 2023). In terms of international significance, globally distributed supply chains operate across national infrastructures and regulatory regimes; analytics-enabled dashboards synchronize these interdependencies and make variance visible at the speed of operations (Chen et al., 2012; Sultan et al., 2023). The streaming-data literature clarifies why system architectures matter: data-stream models and discretized stream engines permit stateful, fault-tolerant updates with bounded delays, a prerequisite for "real-time" operational steering (Babcock et al., 2002; Borkin et al., 2013; Momena & Hasan, 2023). Finally, the marketing and management literature on dashboard use emphasizes that when metrics and visualizations are matched to decision cadence and accountability, users are more likely to act on signals, thereby improving coordination (Pauwels et al., 2009; Sanjai et al., 2023; Akter et al., 2023). This situates AI-integrated dashboards not only as analytic artifacts but also as socio-technical control rooms where models, measures, and managers co-produce operational performance. (Carifio & Perla, 2007; Chae et al., 2014).

Figure 1: Research objectives framework for AI-integrated business intelligence dashboards



A comparative, cross-sectional, multi–case design allows systematic contrasts across organizations that vary in sector, scale, analytic maturity, and data infrastructure while holding constant the unit of analysis the AI-integrated dashboard as used by operational decision makers. The quantitative lens focuses on measurable attributes: perceived usefulness and ease of use (Cronbach, 1951; Davis, 1989), perceived explanation quality and trust in automation (Miller, 2019; Norman, 2010), information quality and BI capability (Isik et al., 2013; Popovič et al., 2012), and observable outcomes such as decision speed, exception resolution rate, and throughput stability (Wamba et al., 2017; White, 1980). Given the prevalence and interpretability of five-point Likert scales in behavioral and IS research, the study employs Likert-type measures for latent constructs, with analysis choices grounded in the measurement literature (Jamieson, 2004). Reliability is assessed with coefficient alpha, which estimates internal consistency under broadly accepted assumptions (Cronbach, 1951), and construct validity is examined through established criteria (Fornell & Larcker, 1981; Henseler et al., 2015). To mitigate common method bias, design remedies and statistical checks are included at the instrument and analysis stages (Podsakoff et al., 2003). Pairing these measurement practices with careful case sampling (variation in operations intensity, automation level, and dashboard scope) enables a structured, cross-sectional snapshot of how AI-integration relates to real-time decision support quality in operational environments. (Davis, 1989; Fornell & Larcker, 1981; Seddon et al., 2017).

In the analytics pipeline that undergirds real-time dashboards, architecture and visualization choices shape what users perceive and how they act. Data-stream processing establishes temporal semantics and dictates whether a metric reflects a true "now," a sliding window, or micro-batches an essential distinction for operations managers targeting cycle-time reductions or stability (Venkatesh & Bala, 2008; Yigitbasioglu & Velcu, 2012). Visualization research documents that visual form, clutter, and salience affect memorability and comprehension, influencing whether critical anomalies or trends are detected under limited attention (Segel & Heer, 2010; Zaharia et al., 2013). In AI-integrated dashboards, this visual layer becomes the delivery surface for explanations, feature importances, or counterfactuals; explanation content and presentation are linked to users' mental models and accountability needs (Podsakoff et al., 2003; Provost & Fawcett, 2013). The BI literature emphasizes the role of data quality, integration, and access in enabling analytic value creation, which conditions whether dashboards become decision instruments rather than passive reporting displays (Fornell & Larcker, 1981; Isik et al., 2013). Embedding these considerations in a comparative analysis clarifies the contribution of AI-specific elements beyond traditional dashboarding e.g., whether local explanations or anomaly-reason codes increase perceived usefulness and trust under time-bound tasks. (Babcock et al., 2002; Chae et al., 2014).

The present study operationalizes AI-integration through observable and perceptual variables suitable for cross-sectional, case-based comparison. Observable variables include whether streaming inputs feed model inferences, whether the dashboard displays uncertainty ranges and drift indicators, and whether controls exist for thresholding recommendations. Perceptual variables include perceived usefulness and ease of use (Davis, 1989), perceived explanation clarity and sufficiency (Miller, 2019; Norman, 2010), trust/reliance (Lee & See, 2004), and perceived information quality and accessibility (Isik et al., 2013; Jamieson, 2004). Outcomes are proxied by self-reported decision speed and confidence and by case-level descriptive indicators (e.g., exception backlog, mean time-to-resolution). Measurement uses five-point Likert items with guidance on analysis from the measurement literature, including conditions under which parametric statistics on Likert-type data are defensible for multi-item scales (Norman, 2010). Reliability is assessed via alpha (Danish, 2023; Seddon et al., 2017), and construct validity is reviewed using established criteria including average variance extracted (Fornell & Larcker, 1981) and discriminant validity checks (Norman, 2010). This structure places human–AI interaction, information quality, and visualization design into a coherent measurement model aligned to the dashboard context of operations. (Danish & Zafor, 2022; Sarikaya et al., 2019; Seddon et al., 2017; Sivarajah et al., 2017).

Quantitatively, the study uses descriptive statistics to characterize cases and samples, correlation analysis to examine bivariate associations among constructs, and regression models to estimate the strength of relationships while controlling for context (e.g., industry, data latency category, and dashboard scope). Robust inferences are supported by heteroskedasticity-consistent covariance estimators (Danish & Kamrul, 2022; White, 1980), variance-inflation diagnostics for multicollinearity, and sensitivity checks across alternative operationalizations. Moderation terms test whether explanation quality and trust condition the relationship between AI-integration and perceived usefulness or outcome proxies, consistent with theorized roles in human–automation research. Given the real-time substrate, architectural factors such as stream processing and micro-batching are included as design covariates, reflecting their documented implications for timeliness and fault tolerance (Babcock et al., 2002; Jahid, 2022). Across cases, this modeling strategy centers the dashboard as the decision interface, quantifying how information, models, and visuals work together in the moment of operational control. (Chen et al., 2012; Trieu, 2017; White, 1980). Finally, the study contributes to three bodies of literature. First, it advances dashboard research by explicitly modeling AI-specific affordances local explanations, anomaly-reason displays, and uncertainty visualization within a real-time, operations-centric context (Arifur & Noor, 2022; Seddon et al., 2017; Sivarajah et al., 2017). Second, it extends BI value research by linking capability elements (data quality, integration, access) to perceived and operational outcomes when AI components are embedded in the decision surface (Isik et al., 2013). Third, it informs human–AI interaction by quantifying how explanation quality and trust relate to adoption constructs and to real-time decision proxies in operations (Chae et al., 2014; Davis, 1989; Henseler et al., 2015). By grounding the analysis in multi-case, cross-sectional evidence and standard quantitative tools, the study provides a structured basis for comparing AI-integrated dashboards used for real-time decision support across diverse operational environments (Sarikaya et al., 2019; Sivarajah et al., 2017).

The overarching objective of this study is to produce a rigorous, quantitative comparison of artificial intelligence–integrated business intelligence dashboards as decision interfaces for real-time operations across multiple organizational cases. Specifically, the study aims, first, to construct and validate an AI Integration Index that operationalizes the presence and depth of predictive, prescriptive, anomaly detection, natural-language, and explainability features, alongside usage intensity, into a reproducible measure suitable for cross-case analysis. Second, it seeks to characterize and compare case contexts by documenting industry, scale, data architecture, dashboard tenure, and user profiles, and by harmonizing operational performance indicators into standardized metrics to enable like-for-like assessment. Third, the study will quantify decision performance by measuring decision latency and decision accuracy or confidence at the user level, linking these outcomes to the AI Integration Index and to user perceptions of usefulness, interpretability, and trust captured via five-point Likert scales. Fourth, it will estimate the strength and direction of relationships among these variables using descriptive statistics, correlation matrices, and multivariate regression models with appropriate controls, reporting standardized coefficients, intervals, model fit, and robustness to alternative specifications. Fifth, the study will test theoretically motivated mechanisms by examining whether perceived usefulness mediates the association between AI integration and decision outcomes, and whether contextual quality factors, including data quality and dashboard tenure, moderate those relationships through interaction terms. Sixth, it will establish the reliability and validity of all multi-item measures through internal consistency diagnostics and construct validity checks to ensure defensible aggregation and interpretation. Seventh, the study will implement a clear sampling and power plan across cases to achieve adequate sensitivity for the number of predictors, while documenting inclusion and exclusion criteria that focus on active operational use rather than pilot or proof-of-concept environments. Eighth, it will prescribe a transparent data preparation pipeline for missingness, outliers, scale construction, and assumption checks, together with heteroskedasticity-consistent estimation for inference under realistic operational variance. Collectively, these objectives position the research to deliver a replicable, measurement-driven account of how AI-enabled dashboard features and user perceptions are associated with the speed and quality of real-time decisions in operational settings.

LITERATURE REVIEW

The literature on business intelligence (BI) dashboards, artificial intelligence (AI)–enabled analytics, and real-time operational decision support converges on a socio-technical view in which data pipelines, models, interfaces, and organizational routines jointly shape performance. Foundational BI research characterizes dashboards as sense-making surfaces that integrate key performance indicators and alerts into compact, action-oriented displays, while adjacent visualization work explains how visual form, salience, and information density condition attention and comprehension under time pressure. Parallel streams in analytics detail how machine learning components forecasting, anomaly detection, prescriptive recommendation, and natural-language interaction extend dashboards beyond descriptive reporting to predictive and directive functionality suitable for high-velocity environments. Human–AI interaction scholarship emphasizes perceptions of usefulness, interpretability, and trust as proximal determinants of adoption and calibrated reliance, drawing attention to explanation quality, uncertainty communication, and the costs of cognitive load. On the systems side, research on streaming architectures and event processing clarifies why latency, state management, and fault tolerance matter for “real-time” claims, and operations management links such capabilities to throughput stability, service reliability, and exception management. Across these domains, however, operationalization is uneven: measures of “AI integration” range from binary feature checklists to implicit maturity judgments; decision outcomes fluctuate between self-reported confidence and log-based cycle times; and comparability of performance indicators across industries is rarely addressed. Moreover, while case studies richly describe implementation contexts, quantitative, cross-sectional comparisons that hold the dashboard as the unit of analysis are underrepresented, limiting cumulative insight into which AI features and perceptual factors most strongly relate to decision speed and decision accuracy in practice. The present review therefore synthesizes constructs and measures into a coherent comparative framework: defining an AI Integration Index with observable and usage facets; specifying user-level perceptions (usefulness, interpretability, trust) measured with five-point Likert scales; distinguishing decision outcomes (latency, accuracy/confidence) from harmonized operational KPIs; and identifying contextual controls and moderators (data quality, organizational size, dashboard tenure). By aligning these elements, the review establishes the theoretical and measurement scaffolding for a multi-case, quantitative examination of AI-integrated dashboards as decision interfaces in live operations, while positioning the empirical study to address long-standing gaps in construct clarity, cross-case comparability, and model-based inference.

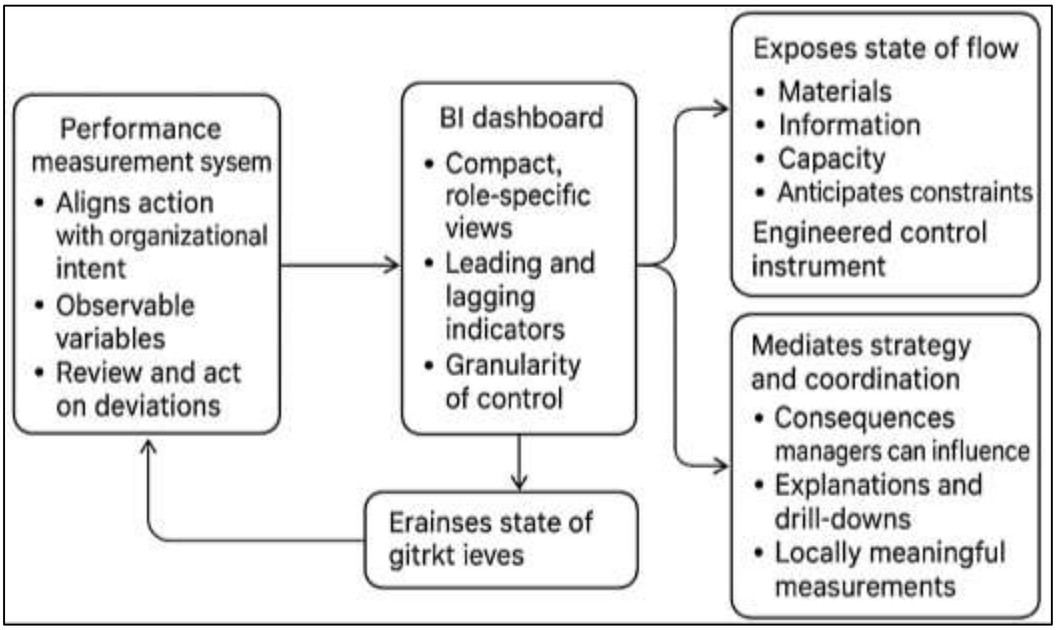
BI Dashboards and Real-Time Decision-Making in Operations

Business intelligence (BI) dashboards have evolved into a central, highly visible layer in the architecture of operations management, serving as the point where data pipelines, analytical logic, and human judgment converge in the cadence of daily work. Within the performance-measurement tradition in operations, dashboards can be understood as "surface enactments" of a larger system that defines, collects, and communicates metrics that align action with organizational intent. A performance measurement system, classically conceived, is not merely a catalog of indicators; it is a designed mechanism that enables managers to translate priorities into observable variables, review those variables at an appropriate frequency, and act on deviations in a disciplined way (Neely et al., 1995). Dashboards make this design legible by compressing heterogeneous data into compact, role-specific views, binding the temporal rhythm of updates to the decision rhythm of operations teams (e.g., hourly replenishment checks, shift handovers, daily throughput reviews). In real-time contexts, the interface also mediates the interplay between leading and lagging indicators (e.g., live queue lengths vs. end-of-day service levels), which is crucial for preventing local optimizations from eroding system-level performance. The operational value of dashboards therefore rests on two intertwined properties: their capacity to represent process behavior at the granularity of control and their capacity to sustain managerial attention under time pressure. When dashboards embody a well-specified measurement design clear constructs, auditable data definitions, and fit-for-purpose update cycles they support a stable loop of sensing, interpreting, and intervening that preserves both speed and accountability in the face of variability (Hasan et al., 2022; Neely et al., 1995).

In supply chain and production operations, the promise of real-time dashboards is to expose the state of flow across materials, information, and capacity so that teams can anticipate, rather than merely react to, emerging constraints. Foundational work on supply chain performance measurement emphasizes designing indicator sets that span strategic, tactical, and operational layers and that balance cost, quality, delivery, flexibility, and asset utilization each with explicit definitions and measurement rules so they can be compared across units and time horizons (Gunasekaran et al., 2004; Redwanul & Zafor, 2022). Dashboards instantiate this architecture by making dependencies and trade-offs visible at a glance: for example, inventory turns contextualized by service level and expedites, or overall equipment effectiveness paired with micro-stoppage patterns. At the operational edge, "real-time" is less a binary attribute than a fit between process dynamics and refresh latencies minutes may suffice for picking waves, while seconds matter in continuous processes. Domain literature also cautions that the benefits of visibility depend on data completeness and governance, since inconsistent definitions or uneven refresh rates can produce misleading signals and erode confidence. Healthcare studies illustrate these points vividly: hospital dashboards that integrate clinical and operational indicators can enhance situational awareness across bedside, unit, and board levels, but effectiveness hinges on data quality, alignment with workflows, and clarity of responsibility when thresholds are breached (Buttigieg et al., 2017; Rezaul & Mesbaul, 2022). Taken together, the operations and supply-chain strands indicate that dashboards must be engineered as control instruments embedded in socio-technical systems, with indicator design, latency targets, and escalation pathways specified as tightly as the graphics themselves. (Hasan, 2022; Melnyk et al., 2004).

A complementary stream examines how and why managers actually use dashboard information in the flow of work, moving beyond technology features to the behavioral conditions for actionability. Empirical evidence suggests that dashboards contribute when they link performance constructs to consequences managers can influence, provide explanations or drill-downs that preserve context, and align measurements with locally meaningful accountability structures. When dashboards are designed in this way, they do more than summarize they act as mediators between strategy and day-to-day coordination, enabling distributed teams to align heuristics (what to watch), thresholds (when to act), and playbooks (how to act) around a shared view of the system state. This perspective also clarifies why "real-time" matters: shorter measurement–action loops reduce the cognitive and organizational distance between signal and intervention, tightening feedback and making learning observable within a shift or planning cycle. Conversely, if dashboards lack construct clarity, obscure data provenance, or overload attention, they can impair coordination by fragmenting the basis for decisions. Recent work on digital dashboards in managerial contexts formalizes these mechanisms by linking dashboard use to management control and individual performance outcomes, highlighting that actionability is a property co-produced by information quality, interface affordances, and local routines. For operations research and practice, this implies that evaluating dashboards requires treating the interface, the measurement design, and the workflow as a single artifact one that succeeds when it consistently moves the right people to the right decisions at the right time. (Neely et al., 1995; Reinking et al., 2020).

Figure 2: Conceptual framework of BI dashboards and real-time decision-making in operations

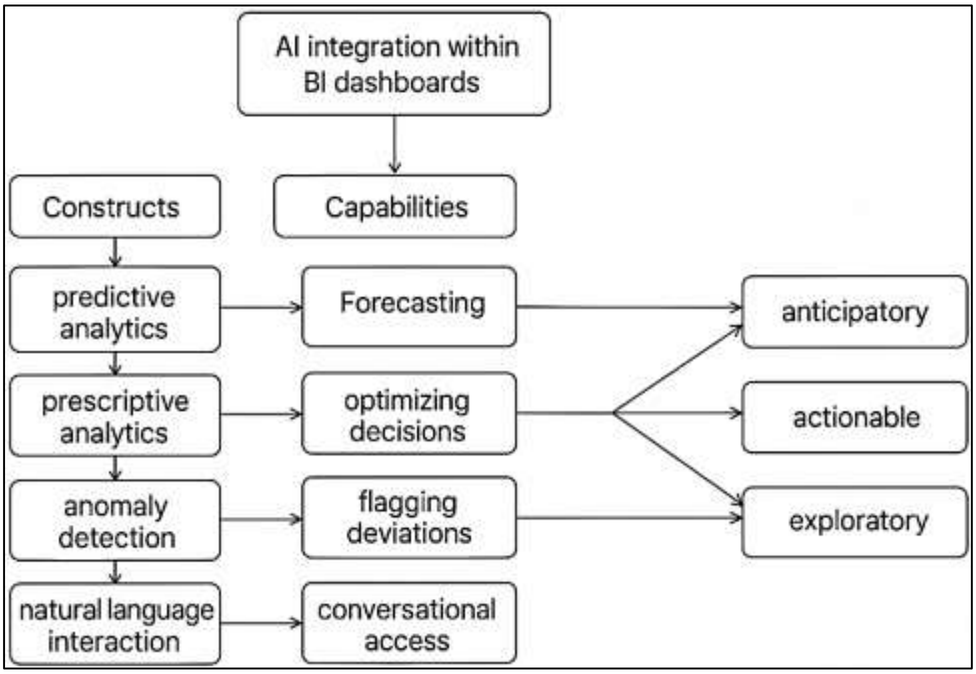


AI Integration within BI Dashboards

A core question for this study is what it really means for a BI dashboard to be "AI-integrated" in the context of real-time operations. At minimum, integration extends beyond embedding a static model output; it entails a cohesive layer of predictive, prescriptive, anomaly-detection, natural-language, and explainability functions that are routinized into everyday decision cycles. Self-service BI research helps frame this shift as an empowerment problem: AI capabilities should reduce the coordination burden between technical specialists and line managers while preserving data governance and metric fidelity. In practice, this means instrumenting dashboards with features that let non-technical users launch model-driven queries, interpret outputs, and act on them without detouring through analytics teams an agenda aligned with the "analysis democratization" aims of self-service BI (Alpar & Schulz, 2016; Kamrul & Omar, 2022). From a measurement standpoint, such integration can be operationalized along two axes. The first is functional breadth (e.g., forecasting, anomaly alerts, recommendations, and conversational access). The second is usage depth (e.g., frequency, decision stakes, and degree of automation). Taken together, these axes clarify why a binary "has AI / doesn't have AI" label is inadequate for comparative research and motivate the creation of an AI Integration Index that captures observable features and their operational salience (Alpar & Schulz, 2016; Makridakis et al., 2018; Kamrul & Tarek, 2022).

Among the capability pillars, prescriptive analytics connects most directly to managerial action by turning predictions into decisions under uncertainty. Prescriptive methods fuse machine-learning predictions with optimization formulations so that a dashboard can surface not only likely futures but also actionable policies inventory quantities, staffing adjustments, or price changes subject to constraints and trade-offs. The prescriptive lens also provides evaluation tools (e.g., the "coefficient of prescriptiveness") that quantify the incremental value of data for decision quality useful for case-to-case comparisons in a multi-site study (Bertsimas & Kallus, 2020). Forecasting is another anchor capability because many operational decisions are anticipatory (what will demand or arrivals look like in the next hour or day?). Evidence from large-scale forecasting competitions shows that ensembles and hybrid approaches often blending statistical baselines with ML components tend to dominate, and probabilistic evaluation (e.g., coverage of prediction intervals) matters for risk-aware decisions that must be executed from a dashboard (Alpar & Schulz, 2016; Makridakis et al., 2018). For anomaly detection, the operational value is immediacy: near-real-time flagging of deviations in throughput, error rates, or sensor streams can trigger playbooks faster than human monitoring alone. Reviews of novelty/anomaly detection methods outline families (density-based, reconstruction, probabilistic, and distance-based) and their trade-offs, offering a menu of techniques that can be embedded behind alert tiles and incident panels (Tarek, 2022; Pimentel et al., 2014).

Figure 3: Framework of AI integration within BI dashboards

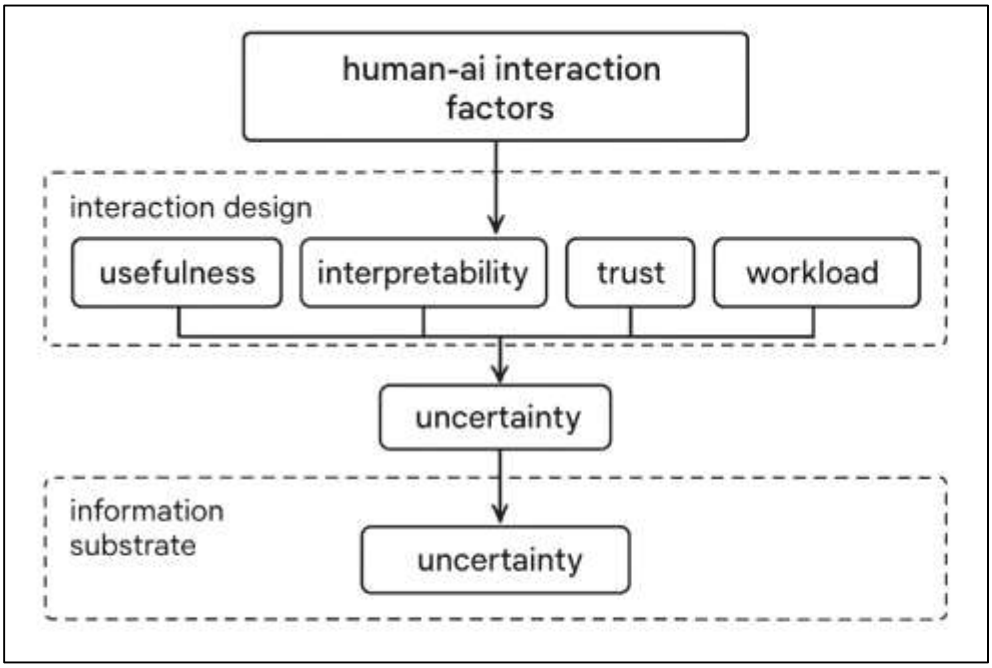


Human–AI interaction features further determine whether model outputs become usable decisions. Natural-language interfaces (NLIs) lower query formulation costs by allowing operators to "ask" the dashboard for explanations, breakdowns, or drill-downs in plain language especially helpful under time pressure when locating the right filter or chart is costly. Classic work on NL interfaces to databases catalogs linguistic challenges (ambiguity, vagueness, temporal expressions) and architectural strategies (grammars, semantic parsing) that remain directly relevant when conversational access is layered onto BI surfaces (Androutsopoulos et al., 1995; Mubashir & Abdul, 2022). Equally important is explainability: operators must quickly understand why the system recommends an action and how confident it is. Comprehensive surveys of explainable AI classify model-agnostic/local vs. global techniques, counterfactuals, and rule extraction; these approaches can be rendered as on-demand tooltips, feature-importance sparklines, or contrastive "why not" panels embedded within a dashboard (Guidotti et al., 2018; Muhammad & Kamrul, 2022). Finally, organizational studies remind us that analytics capabilities yield performance when they are embedded in process routines and matched to information-processing needs across plan–source–make–deliver stages; dashboards are the place where that matching is experienced by users, so the presence of AI must translate into better throughput, service, or cost outcomes to be meaningful (Trkman et al., 2010).

Human–AI Interaction Factors

In operations-facing dashboards, the human–AI interaction layer governs whether model outputs translate into timely, correct action. A useful starting point is the "levels of automation" perspective, which frames how functions are allocated between humans and automated agents and how that allocation shapes supervision, intervention, and accountability (Parasuraman et al., 2000). In real-time contexts, higher levels of automation can reduce cognitive burdens of monitoring and control but also alter situation awareness (SA) the continuous perception–comprehension–projection cycle operators need for safe and efficient action (Endsley, 1995; Reduanul & Shoeb, 2022). SA is not a mere by-product of accuracy; it is a cognitive state sustained by interfaces that expose task-relevant cues, clarify system intent, and make temporal dynamics legible (Endsley, 1995; Jian et al., 2000; Kumar & Zobayer, 2022). When BI dashboards embed AI components anomaly detectors, predictive forecasts, or prescriptive recommendations the presentation of those outputs must therefore preserve and, ideally, enhance SA rather than displace it.

Figure 4: Human–AI interaction factors in BI dashboards



Trust is the other structural pillar: it calibrates reliance and shapes when people accept, query, or override algorithmic suggestions. Synthesis work indicates that trust is multi-determined by system performance history, transparency, and user traits and that miscalibration (over- or under-trust) leads to omission or commission errors under time pressure (Parasuraman et al., 2000; Wang & Strong, 1996). In practice, then, "usefulness" and "interpretability" are enacted through a balance: sufficient automation to compress decision cycles, sufficient human control to maintain SA, and sufficient transparency to calibrate trust for the tempo of operations (Hart & Staveland, 1988; Kinkeldey et al., 2014).Measurement and design choices inside the dashboard directly affect that balance. Perceived interpretability depends not only on explanation content but also on how uncertainty and limits are communicated. Research on uncertainty communication shows that users reason more effectively when uncertainty is expressed in clear, decision-relevant forms (e.g., ranges, coverage, or qualitatively coded risk) and when visual encodings are empirically tested for comprehension (Hoff & Bashir, 2015). Poorly designed uncertainty cues can either overwhelm (inflating workload) or falsely reassure (degrading SA). Cognitive workload, often assessed with NASA-TLX, is central here: real-time dashboards that layer alerts, model scores, and narratives can create extraneous load that competes with the analytical work of sense-making (Hart & Staveland, 1988; Sadia & Shaiful, 2022). Elevated workload narrows attention and impairs projection the "what happens next" element of SA thereby undermining the value of even accurate model outputs (Endsley, 1995). Trust measurement itself benefits from validated instruments; the Jian–Bisantz–Drury scale provides an empirically grounded set of items for gauging trust in automated systems and can be adapted to assess reliance tendencies in dashboard use (Jian et al., 2000). Coupled with SA- and workload-oriented diagnostics, these measures

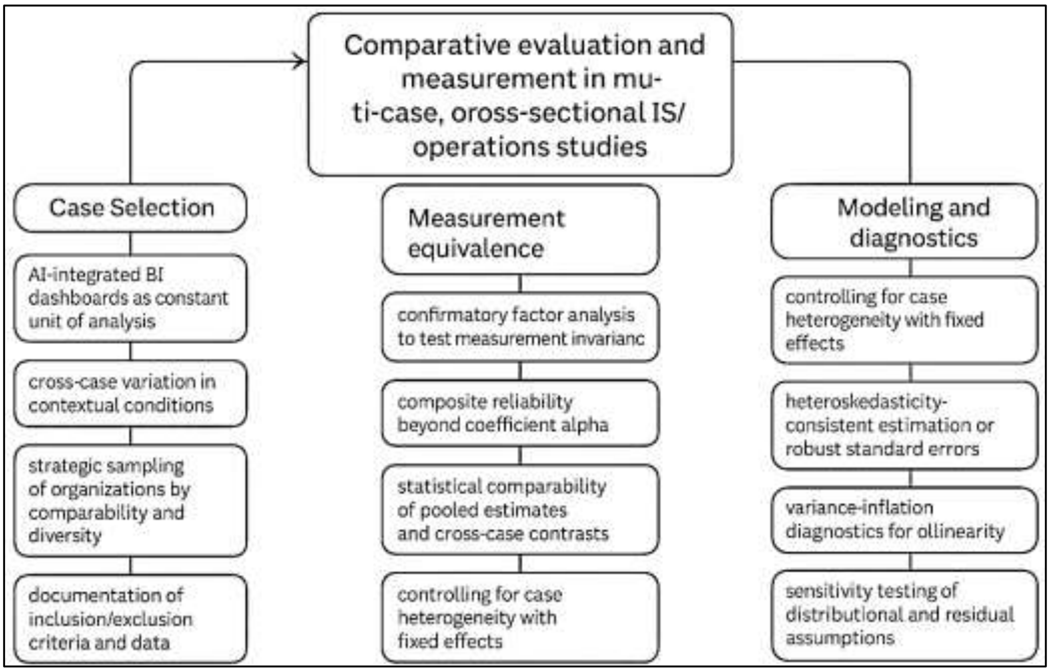
allow researchers to distinguish between interfaces that merely add information and those that actually support better, faster decisions by shaping attention, comprehension, and calibrated reliance (Hoff & Bashir, 2015; Kinkeldey et al., 2014).
Finally, the information substrate behind the interface conditions trust, interpretability, and usefulness before any model or visualization is considered. Data quality research emphasizes that "fitness for use" encompasses far more than accuracy; dimensions such as timeliness, completeness, interpretability (of the data), and accessibility determine whether users deem analytics outputs credible and actionable (Wang & Strong, 1996). In operational dashboards, stale, incomplete, or inconsistently defined inputs propagate into model outputs and explanations, creating a gap between apparent precision and decision value. Trust models predict that such gaps erode reliance more rapidly than improvements later rebuild it, especially after salient errors (Hoff & Bashir, 2015). Accordingly, the human–AI layer must bind together (i) allocation of functions across levels of automation, (ii) SA-preserving visual and interaction design, (iii) explicit uncertainty communication vetted against user studies, (iv) workload-aware presentation that prioritizes signal over clutter, (v) validated measures of trust and workload to monitor adoption, and (vi) rigorous data quality management to sustain credibility (Hart & Staveland, 1988; Jian et al., 2000; Parasuraman et al., 2000). Taken together, these factors specify a socio-technical target for AI-integrated dashboards: not simply showing more analytics, but shaping human attention, understanding, and calibrated reliance so that real-time decisions align with operational objectives under uncertainty and time pressure.

Comparative Evaluation & Measurement in Multi-Case

Comparative evaluation in information systems (IS) and operations research hinges on purposeful case selection and a transparent replication logic that enables analytic generalization rather than statistical generalization. In a multi-case, cross-sectional design, the aim is to hold the unit of analysis constant (here, AI-integrated BI dashboards used for real-time operational decisions) while varying contextual conditions across organizations to observe patterned regularities. Classic guidance on multi-case research emphasizes building theory from cases via iterative comparison, the disciplined use of within- and cross-case tables, and the constant search for disconfirming evidence to refine constructs and relationships (Eisenhardt, 1989). Because case heterogeneity can either illuminate or obscure relationships, case selection becomes a design choice with first-order consequences; typologies of case selection strategies (e.g., typical, diverse, extreme, influential) provide a principled menu for balancing variation with comparability in cross-sectional comparisons (Seawright & Gerring, 2008; Noor & Momena, 2022). In the present context, that logic implies recruiting organizations that differ in industry, scale, and data maturity yet share a common operational cadence and dashboard use, so that AI-integration features can be compared "like-for-like." The cross-sectional lens then enables the researcher to quantify inter-case contrasts at a single time slice, allowing tests of whether differences in feature bundles, explanation affordances, and data-quality regimes covary with decision latency and perceived decision accuracy/confidence. Crucially, the comparative frame requires explicit documentation of inclusion/exclusion criteria, response roles, and data windows so that readers can judge the scope of inference an approach consistent with the replication and transparency ideals that sit at the heart of robust multi-case evaluation (Eisenhardt, 1989; Seawright & Gerring, 2008).
Measurement equivalence is the second pillar of credible comparative evaluation. Cross-case comparison is only meaningful when constructs are measured in ways that are conceptually coherent and empirically comparable across groups. Organizational methods scholarship lays out procedures for establishing measurement invariance configural (same factor structure), metric (equal loadings), and scalar (equal intercepts) before comparing latent means or structural paths across samples (Raykov, 1997; Vandenberg & Lance, 2000). In practice, this translates into a staged multi-group confirmatory factor analysis where successive equality constraints are tested for acceptable deterioration in model fit; practical guidelines highlight that changes in incremental fit indices (e.g., ΔCFI) provide decision thresholds for invariance judgments even with large samples (Cheung & Rensvold, 2002; Danish, 2023; O'Brien, 2007). For dashboards, that scaffolding allows researchers to assert, for example, that "perceived interpretability" or "trust" items behave similarly for users in manufacturing and healthcare cases, legitimizing pooled estimates and cross-case contrasts. Reliability must also be treated with nuance. Beyond coefficient alpha, which assumes tau-equivalence and can be biased by scale

length, structural-equation-model-based composite reliability offers a more defensible estimate by weighting items according to their loadings and error variances (Hasan et al., 2023; Raykov, 1997). Together, invariance testing and composite reliability create the conditions under which cross-case, cross-industry comparisons of perception constructs (usefulness, interpretability, trust) and outcome proxies (decision latency, decision accuracy/confidence) are not only statistically tractable but also substantively interpretable. Absent such evidence, apparent differences across organizations may reflect measurement artifacts rather than real variation in how AI-integrated dashboards support real-time decisions (Breusch & Pagan, 1979; Vandenberg & Lance, 2000).

Figure 5: Comparative evaluation and measurement framework in multi-case



The third pillar concerns modeling and diagnostic discipline for cross-sectional inference under operational noise. Comparative regression models that pool respondents across cases should incorporate case fixed effects or, at minimum, case dummies to partial out unobserved heterogeneity, while reporting robust uncertainty that respects plausible violations of classical assumptions common in field data. Heteroskedasticity a near certainty when comparing organizations of different sizes, sectors, and data infrastructures should be tested explicitly; the Lagrange Multiplier approach offers a direct specification test and motivates the use of heteroskedasticity-consistent covariance estimators when appropriate (Breusch & Pagan, 1979; Cheung & Rensvold, 2002; Md Ismail Hossain et al., 2023). Multicollinearity is another frequent threat in studies where perceptual predictors (usefulness, interpretability, trust) are theoretically proximate; variance-inflation diagnostics should be interpreted with care, avoiding overly rigid thresholds and focusing on the consequences for coefficient instability and interpretability (Rahaman & Ashraf, 2023; O'Brien, 2007). In multi-group or interaction models (e.g., testing whether data quality moderates the AI-integration → decision performance link), researchers should align diagnostics with their modeling choices checking cross-case distributional differences, inspecting residual plots by case, and reporting sensitivity analyses that remove influential observations and re-estimate models on matched subsamples. Collectively, these practices explicit invariance testing, reliability beyond alpha, cautious treatment of multicollinearity, and robust handling of heteroskedasticity turn a descriptive cross-section into a credible comparative evaluation. They also anchor managerial interpretations: when the analyst says that higher AI-integration is associated with lower decision latency across cases, that claim rests on instruments that behave equivalently in different contexts and estimates whose uncertainty appropriately reflects operational variance (Cheung & Rensvold, 2002; Eisenhardt, 1989; Vandenberg & Lance, 2000).

METHODS

This study has adopted a quantitative, cross-sectional, multi–case design to compare AI-integrated BI dashboards as decision interfaces in real-time operational settings. The investigation has treated the dashboard rather than the organization as the primary unit of analysis, and it has sampled multiple organizations that have met predefined inclusion criteria (active operational use, presence of at least one AI feature, and accessible performance records). By anchoring analysis at the interface level, the design has sought to isolate how specific AI capabilities and user perceptions have related to decision outcomes within a consistent observational window. The research has employed a mixed source strategy for measurement. Structured surveys administered to active users (e.g., supervisors, analysts, and operations managers) have captured perceptual constructs perceived usefulness, interpretability, trust, and perceived workload using five-point Likert scales that have been adapted for operational contexts. In parallel, objective indicators have been compiled through case-level audits and data extracts. An AI Integration Index has been operationalized to reflect functional breadth (e.g., forecasting, anomaly detection, prescriptive recommendations, natural-language interaction, and explainability widgets) and usage depth (e.g., frequency, decision stakes, and degree of automation). Decision outcomes have been represented by decision latency and decision accuracy/confidence at the user level, while harmonized operational KPIs (e.g., exception resolution time, throughput stability) have been standardized to enable cross-case comparison. Contextual controls organization size, industry, dashboard tenure, data quality, and user analytics proficiency have been documented to address confounding. Data preparation procedures have been pre-specified and have included screening for missingness, outliers, and assumption violations. Multi-item scales have undergone reliability diagnostics, and construct structure has been assessed prior to composite formation. The analysis plan has combined descriptive statistics and correlation matrices with multivariate regression models that have incorporated robust uncertainty estimation and case controls. Where theoretically indicated, mediation and moderation terms have been specified to test mechanisms and boundary conditions (e.g., usefulness as mediator; data quality as moderator). Power considerations for pooled models have guided the targeted sample size, and inclusion/exclusion rules have been enforced consistently across cases. Collectively, these method choices have established a replicable framework that has enabled like-for-like comparisons of AI feature bundles, user perceptions, and real-time decision outcomes across heterogeneous operational environments.

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

The study has adopted a quantitative, cross-sectional, multi–case design to evaluate how AI-integrated BI dashboards have supported real-time decision making in operations. The dashboard has been treated as the primary unit of analysis so that features, usage, and user perceptions have been examined consistently across heterogeneous organizational settings. Multiple cases have been selected to represent variation in industry, scale, and data maturity while sharing the common condition of active, operational dashboard use. A single observational window has been specified so that all measurements have referred to comparable periods of activity, and inclusion/exclusion rules have been applied to ensure that only production deployments not pilots or proofs-of-concept have been considered. Within each case, respondents with direct decision responsibilities (e.g., supervisors, operations analysts, and managers) have been recruited, and case leads have provided access to nonidentifiable performance records. The design has combined perceptual and objective evidence: structured surveys have captured user-level constructs (usefulness, interpretability, trust, workload), while system audits and data extracts have documented AI feature breadth, usage depth, and relevant operational indicators. To enable cross-case comparability, metrics have been harmonized and, where necessary, standardized. Controls for organizational context (size, industry, dashboard tenure, data quality, and user analytics proficiency) have been documented a priori and have been incorporated into modeling. The cross-sectional frame has allowed relationships among AI integration, perceptions, and decision outcomes to be estimated at a fixed point in time, and the multi-case logic has created replication opportunities across settings. To protect internal validity, instrument wording and administration procedures have been unified, data preparation steps have been pre-specified, and analysis has relied on robust estimation with case controls. Overall, the design has provided a replicable basis for like-for-like comparisons of AI feature bundles and their associations with decision latency, decision accuracy/confidence, and harmonized operational KPIs across diverse operational environments.

Cases, Sampling, and Setting (Inclusion/Exclusion)

The study examined multiple organizational cases where AI-integrated BI dashboards were deployed in day-to-day operational control, treating each dashboard environment as a bounded setting for observation. Cases were drawn from sectors with routine real-time coordination such as manufacturing, logistics, healthcare operations, and technology-enabled services, with only sustained production deployments included rather than pilot projects. For each case, a single reference window was established to align survey administration, log extraction, and KPI capture, ensuring temporal comparability across organizations. Case contacts included an operations leader, a data/IT liaison, and an analytics or product owner, enabling access to workflow context, systems, and feature verification. Each case dossier documented the dashboard's data architecture (sources, latency, refresh schedules), catalog of AI features (forecasting, anomaly detection, prescriptive recommendations, natural-language interaction, explainability widgets), and governance routines, forming the basis for computing the AI Integration Index and linking user responses to practical realities. Confidentiality was preserved through neutral identifiers (e.g., Case A, Case B) and de-identified outputs. Sampling targeted active dashboard users with decision rights or operational responsibilities, such as supervisors, managers, analysts, and coordinators, recruited through structured, organization-approved channels with standardized invitations, reminders, and unique survey links. Enrollment was guided by power considerations for regression models and continued until targets were met or the case population exhausted, with steps taken to mitigate selection bias by including participants across shifts, sites, and access devices. System/log extracts for metrics such as decision latency and alert validation were mediated by IT liaisons using pre-agreed queries, retaining only non-identifiable fields. Inclusion criteria required continuous dashboard use for at least a quarter, active AI features, stable data pipelines, and a minimum of ten eligible users, while exclusion criteria removed immature or shadow deployments, poorly defined metrics, sporadic refreshes, and users without relevant decision roles. Cases with legal or contractual barriers to de-identified log sharing were also excluded. Data protection was ensured through role-based access, encryption, identifier separation, and strict use of aggregated results, while consent protocols emphasized voluntary participation, confidentiality, and withdrawal rights. This approach produced a sample where AI capabilities, user perceptions, and decision outcomes could be consistently observed under stable operating conditions, strengthening both internal coherence and cross-case comparability.

Variables & Measures

The study specified a structured set of variables aligned with the dashboard-as-unit perspective to enable like-for-like comparisons across cases, with the primary independent construct being AI Integration, operationalized through a two-facet AI Integration Index combining functional breadth and usage depth. Functional breadth captured whether dashboards included active modules for short-horizon forecasting, anomaly detection with tunable thresholds, prescriptive recommendations or playbooks, natural-language interaction, and explainability widgets, with each capability verified in the case dossier and scored based on presence, configurability, and recency of updates. Usage depth reflected how often and how consequentially these capabilities were invoked, measured through items on frequency of use in time-sensitive decisions, degree of automation, and the share of incidents influenced by AI features, with the index computed as a weighted composite after reliability checks and normalized to a 0–100 scale. To avoid halo effects, audit evidence and structured self-reports from multiple respondent roles were combined to build independent scores. Mediating and perceptual determinants such as Perceived Usefulness, Perceived Interpretability, Trust/Reliance, and Perceived Workload were measured with multi-item 5-point Likert scales, adapted to operations contexts, validated in piloting, and designed to capture mechanism-oriented hypotheses while providing diagnostic design insights; reverse-coded items and checks for item-total correlations ensured reliability, and standardization across cases facilitated comparability. Outcome variables were defined at both user and case levels, with Decision Latency measured as elapsed time from signal surfacing to decision action (directly from logs where available, otherwise through bounded self-reports), Decision Accuracy/Confidence measured via perceived correctness and, where possible, linked to post-hoc outcome labels, and case-level Operational KPIs (e.g., backlog clearance, throughput variability, stockouts, first-time-right rates) harmonized with z-scores within cases to prevent scale dominance. Control variables such as organization size, industry, dashboard tenure, user proficiency, training exposure, and Data Quality (summarized as a composite of timeliness, completeness, and definition clarity) were incorporated to account for contextual heterogeneity. Mapping tables ensured each respondent's survey data was coherently paired with their audited dashboard capabilities, enabling regression models to parse the associations between AI Integration and decision performance under stable, production-level conditions. This design provided both analytical rigor and sector-spanning comparability, positioning the study to generate robust insights into how AI-integrated BI dashboards shape operational decision-making.

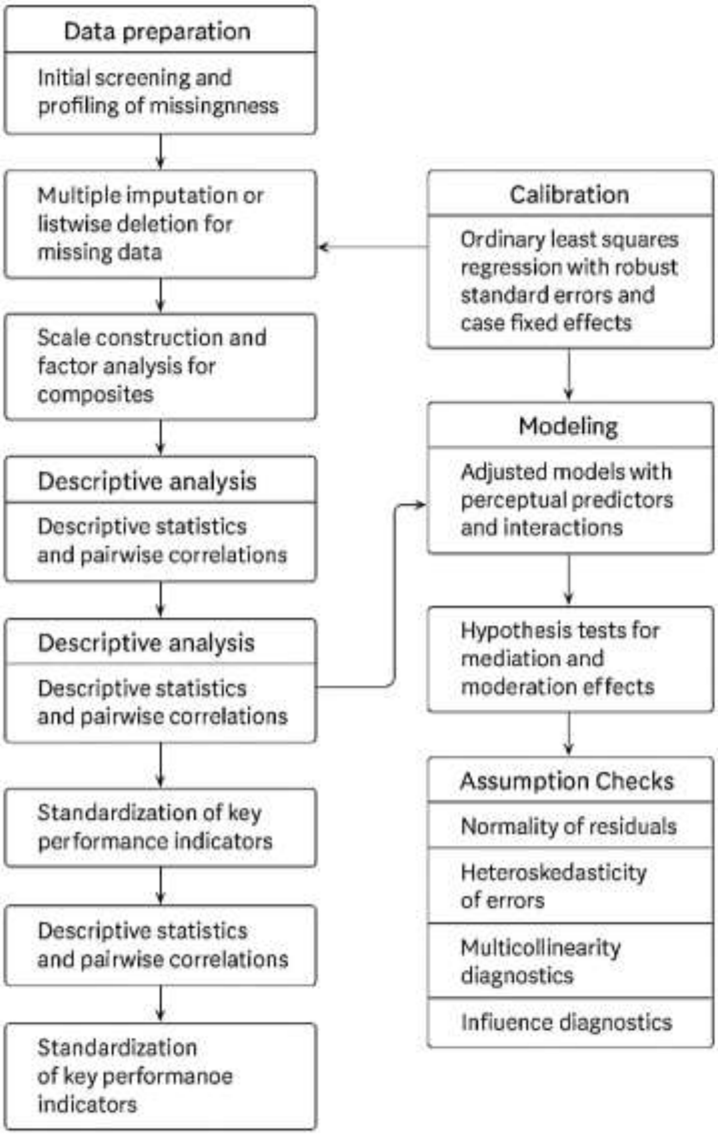
Data Sources & Collection

The study employed two complementary data sources—standardized user surveys and system-level artifacts (feature audits, event logs, and archival KPIs)—collected within a common reference window to ensure temporal alignment across cases. For each organization, a structured case dossier documented the dashboard's technical footprint, catalog of AI capabilities, governance practices, and availability of logs or reports, compiled jointly by an operations leader, IT/data liaison, and product/analytics owner. This dossier formed the evidentiary basis for constructing the AI Integration Index and linking respondents to their dashboard's capabilities. In parallel, surveys were administered to eligible active users with operational decision responsibilities, distributed via individualized links with consent prompts explaining voluntary participation and data handling. The instrument collected demographics, training exposure, analytics proficiency, perceptual constructs (usefulness, interpretability, trust, workload) using five-point Likert scales, and incident-anchored questions to minimize recall bias, while item randomization, reverse-coded statements, and soft validations safeguarded response quality. Objective data were drawn from event logs and archival KPIs, with IT liaisons executing predefined queries to extract anonymized timestamps, event types, and role tags; where logs were absent, bounded self-reports were used with calibration checks. Archival KPIs, such as backlog clearance times, throughput variance, stockout incidents, and first-time-right rates, were standardized within cases for pooling. All data transfers were encrypted, role-based access controlled raw extracts, and analytic files stored survey and system artifacts separately, joined only through anonymized mapping keys. Survey timing was synchronized with log/KPI windows, late responses were flagged, and quality screens checked completeness, plausibility, duplicates, and straight-lining. Limitations such as sparse logs were documented in case dossiers and triangulated with survey data. An audit trail versioned instruments, scripts, and case dossiers, recording deviations to preserve traceability. Overall, this coordinated, privacy-preserving process enabled robust linkage between AI integration, user perceptions, and operational decision outcomes across diverse organizational settings.

Statistical Analysis Plan

The analysis was pre-specified to translate the multi-case, cross-sectional evidence into reliable estimates of associations among AI integration, user perceptions, and decision outcomes while minimizing field-data pitfalls. Data pipelines enforced schema checks, flagged incomplete or implausible records, and profiled missingness patterns; listwise deletion was planned for ≤5% item nonresponse, while higher rates assumed missing at random were addressed with multiple imputation using chained equations. Scale construction followed staged logic, with item distributions and correlations inspected, low-contribution items removed, and internal consistency confirmed before forming composites; pooled data enabled confirmatory factor analysis (or split-sample EFA/CFA if constrained). Composite scores used means (reverse-coded where necessary), perceptual constructs remained on five-point metrics, and case-level KPIs were z-standardized to support pooling. Descriptive statistics stratified by case, role, and shift provided context, and pairwise correlations previewed associations. Primary estimation relied on OLS regression with robust errors and case fixed effects, with baseline models including AI Integration and controls, extended models adding perceptual predictors, and full models testing interaction terms for moderators such as Data Quality and Dashboard Tenure. Mediation hypotheses were tested via nonparametric bootstrapping, moderation effects with simple-slope and marginal-effect plots, and assumptions checked through Q–Q plots, heteroskedasticity screening, multicollinearity diagnostics, and influence measures; robustness was verified by excluding outliers. Transformations were applied where skewness was significant, and ordered logistic regression cross-checks were planned for ordinal latency bands. Risks of common method variance were mitigated procedurally in survey design and analytically through single-factor and marker-variable tests, with findings reported transparently. Multiple-comparison risks were managed with effect sizes, confidence intervals, and false discovery rate adjustments where necessary, while subgroup analyses by sector, role, or shift were labeled exploratory. All models documented degrees of freedom, estimation decisions, and deviations from plan, with results visualized through coefficient plots, partial residuals, and marginal effects. Together, these steps created a reproducible and transparent pipeline prioritizing measurement coherence, robust uncertainty estimation, and credible cross-case comparisons of AI feature bundles, user perceptions, and real-time decision outcomes.

Figure 6: Statistical analysis plan for multi-case, cross-sectional IS/operations studies



Regression Models

The study specified a family of regression models treating the dashboard as the analytic focal point while accommodating heterogeneity across organizational contexts, with three primary outcome equations defined: user-level decision latency, user-level decision accuracy/confidence, and case-level harmonized operational performance. Each model incorporated the AI Integration Index as the principal predictor, alongside controls for organization size, industry, dashboard tenure, user analytics proficiency, training exposure, and a composite Data Quality score, with case fixed effects (or dummies, where justified) absorbing unobserved, time-invariant differences. For user-level outcomes, robust

HC3 standard errors clustered by case addressed intra-case dependence, and decision latency recorded in bounded categories was re-estimated with ordered logit as a cross-check. Perceptual constructs—Perceived Usefulness, Perceived Interpretability, Trust/Reliance, and Perceived Workload—were introduced sequentially to test incremental explanatory power without overfitting, with standardized notation and canonical forms summarized in Table 1 to ensure comparability. Mediation and moderation structures were embedded within this scaffold: Perceived Usefulness was modeled as a mediator linking AI Integration to outcomes, with indirect effects estimated via nonparametric bootstrapping (5,000 resamples) using bias-corrected confidence intervals, while moderation was tested through interaction terms between AI Integration and Data Quality (primary moderator) and, in sensitivity checks, Dashboard Tenure, with variables mean-centered and simple-slope tests conducted at representative values (−1 SD, mean, +1 SD). Significant interactions were visualized using marginal-effect plots with 95% confidence bands, and exploratory role-based interactions (e.g., supervisor vs. analyst) were reported transparently. Model adequacy and robustness were addressed through residual and Q–Q plots, heteroskedasticity screening, and HC3 as the default covariance estimator, with multicollinearity monitored via VIFs emphasizing coefficient stability; elevated VIFs due to interactions were mitigated through centering and ridge-type sensitivity checks. Influence diagnostics (Cook's distance, leverage) were computed, models refitted excluding high-influence points, and results reported in appendices; skewed predictors or outcomes were transformed (log or Box–Cox) with interpretive notes linking back to operational units. Model comparison relied on adjusted R^2 for OLS, pseudo-R^2 for ordered models, and AIC/BIC for nested and non-nested variants, while families of related hypotheses were subjected to false-discovery-rate checks. To confirm findings were not case-dependent, a leave-one-case-out procedure summarized coefficient distributions across k refits. Together, this regression family, mediation/moderation framework, and robustness protocol formed an integrated estimation plan capable of isolating the association between AI Integration and decision outcomes, clarifying mechanisms through perceived usefulness, and identifying boundary conditions shaped by data quality and deployment tenure, with Table 2 detailing the coding rules, transformations, and standardization decisions that supported consistent interpretation across models.

Table 1. Canonical Regression Specifications

Model	Outcome (unit)	Core specification (with case effects)	Notes
A	Decision Latency (minutes or ordered bands)	Latency_i = β0 + β1 AIInt_i + β_c'Controls_i + γ * case + ε_i → extended with PU, PI, TR, WL	OLS with HC3 SEs (primary); ordered logit sensitivity if banded
B	Decision Accuracy / Confidence (standardized)	Accuracy_i = β0 + β1 AIInt_i + β_c'Controls_i + γ * case + ε_i → extended with PU, PI, TR, WL	Mediation via PU bootstrap; moderation via AIInt × DataQuality
C	Harmonized Operational KPI (case-level z)	OKPI_ic = β0 + β1 AIInt_ic + β_c'Controls_ic + γ * case + ε_ic	Aggregated or respondent-linked; clustered SEs by case

All user-level models have included case fixed effects (or dummies) and HC3 robust standard errors clustered by case; mediation has used 5,000-bootstrap CIs; moderation terms have been mean-centered prior to interaction construction.

Table 2. Variable Encyclopedia and Transformations

Construct	Operationalization	Scale/Transform	Role in models
AI Integration (AIInt)	0–100 index (breadth × depth)	Standardized (per 1 SD) in plots	Focal predictor
Perceived Usefulness (PU)	4–6 item Likert composite	Raw (1–5) and z for sensitivity	Mediator; covariate
Perceived Interpretability (PI)	4–6 item Likert composite	Raw (1–5)	Covariate
Trust/Reliance (TR)	4–6 item Likert composite	Raw (1–5)	Covariate
Workload (WL)	4–6 item Likert composite (rev-coded items)	Raw (1–5)	Covariate
Data Quality	Timeliness, completeness, definition clarity composite	z (case-centered)	Moderator; control
Dashboard Tenure	Months in production	log(1+x) for skew	Moderator; control
Controls	Size, industry, training, analytics proficiency	As coded	Controls across models

Power & Sample Considerations

The study established power and sample targets a priori to ensure that the planned regression models possessed adequate sensitivity to detect theoretically meaningful effects while accommodating case controls and interaction terms. Given the focal continuous predictor (AI Integration Index) and multiple covariates, the team treated a medium effect size (incremental $f^2 \approx 0.08$–0.10 for the focal block) as the minimally important difference for user-level outcomes and translated this target into participant counts using standard power heuristics. Specifically, pooled OLS models with case fixed effects and approximately 8–12 predictors (including controls and perceptual constructs) required, under $\alpha = 0.05$ and $1-\beta = 0.80$, on the order of 120–180 analyzable user responses to detect the incremental variance explained by AI Integration, with an additional 15–25% buffer to offset planned robustness checks and potential listwise deletion. Anticipating moderation tests (e.g., AI Integration × Data Quality), which typically exhibit smaller standardized effects, the study inflated targets by approximately 30–40 respondents to preserve power for interaction coefficients after mean-centering and multicollinearity penalties. At the case level, the analysis aimed to include at least 6–8 distinct cases to permit stable estimation of fixed effects and leave-one-case-out sensitivity; within each case, a minimum of 15–25 eligible users was pursued to balance within-case precision and recruitment feasibility across shifts and sites. To protect against attrition and partial survey completion, invitations per case exceeded eligibility rosters by 20–30%, and rolling reminders were scheduled within the reference window. Where objective log data were unavailable or sparse, the plan reserved additional respondents in those cases to maintain precision in user-reported latency bands. Finally, the sampling plan incorporated distributional checks (role mix, tenure bands, shift coverage) so that model covariates did not become sparsely populated, and it pre-specified replacement rules if a case failed inclusion criteria, thereby ensuring that the pooled dataset met the minimum analyzable sample while remaining balanced across diverse operational environments.

Reliability & Validity

The study has implemented a multi-step reliability and validity program that has aligned with the dashboard-as-unit perspective and has supported cross-case comparability. Content validity has been established first: item pools for perceived usefulness, interpretability, trust/reliance, and workload have been mapped to construct definitions, and expert reviews across operations, analytics, and HCI roles have been conducted to refine wording, remove redundancy, and ensure relevance to real-time decision tasks. A pilot administration has been completed to evaluate clarity, response time, and item functioning; feedback has been incorporated, and minor lexical adjustments have been made without altering construct scope. Internal consistency reliability has then been evaluated for each multi-item scale; items with weak item–total correlations or cross-domain contamination signs have been flagged, and final composites have been formed only after acceptable reliability thresholds have been met.

Where sample–parameter ratios have permitted, a confirmatory factor analysis with case indicators has been employed to verify the intended factor structure; where constraints have existed, an EFA/CFA split-sample approach has been adopted. Convergent validity has been examined through factor loadings and composite reliability, and discriminant validity has been assessed by comparing shared variance across constructs with their respective average variance explained. Measurement invariance procedures across cases and key respondent groups (e.g., supervisors vs. analysts) have been executed in a staged manner configural, metric, then scalar so that cross-case comparisons have rested on comparable measurement. Criterion-related validity has been addressed by inspecting expected directional associations between perceptual composites and decision outcomes, while controlling for contextual covariates. To mitigate common method variance, the instrument has included procedural remedies (anonymity, proximal separation of predictors and outcomes, mixed item valence) and post-hoc diagnostics; results have been interpreted as indicators rather than proofs of absence. For objective indicators, data dictionaries, lineage notes, and latency audits have been maintained so that construct validity for KPIs and decision latency has been anchored in auditable definitions. Collectively, these steps have produced measures that have demonstrated reliability, construct clarity, and cross-case equivalence sufficient for the planned regression, mediation, and moderation analyses.

Software

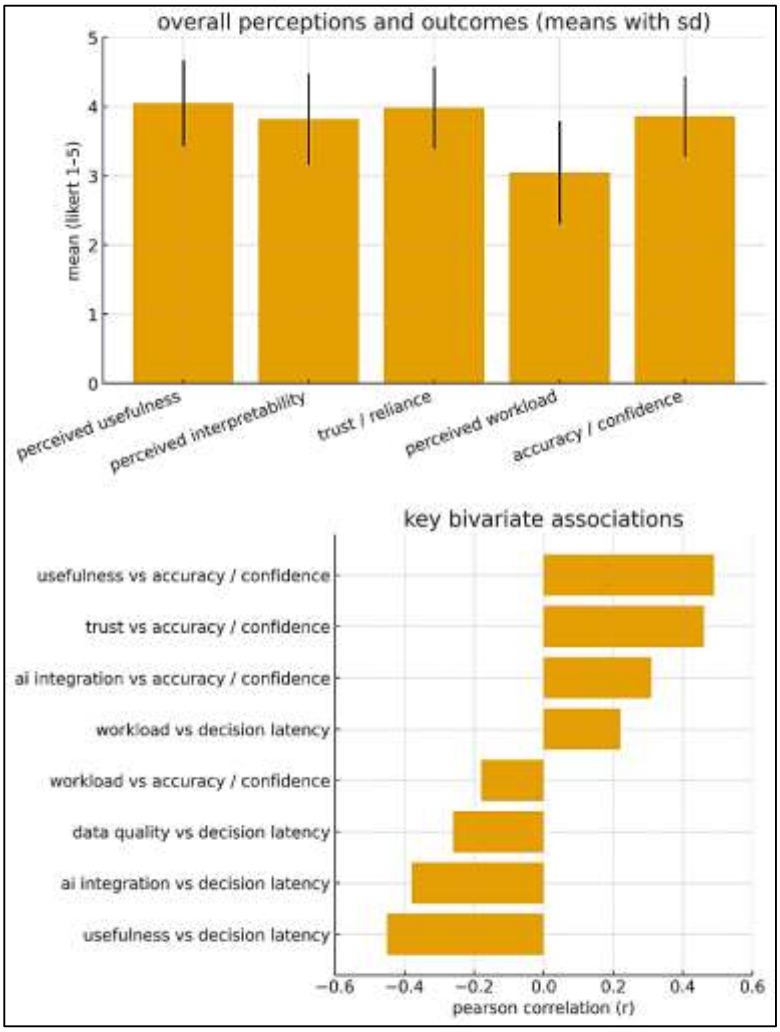
The study has relied on a reproducible, multi-tool stack that has balanced statistical rigor with operational practicality. Data wrangling and analysis workflows have been implemented primarily in R (tidyverse, data.table, psych, lavaan, car, sandwich, lmtest) and Python (pandas, numpy, scipy, statsmodels, pingouin) to enable parallel verification of descriptive, correlation, regression, mediation, and moderation results. Survey design and administration have been executed via a secure online platform that has supported randomized item blocks, branching, and de-identified exports; raw payloads have been versioned using Git with scripted ETL in R/Python. Power analyses have been conducted with G*Power to set sample targets, while visualization of diagnostics and effects (coefficient plots, marginal effects, Q–Q/residuals) has been produced with ggplot2 and matplotlib. For documentation and literate analysis, the team has used R Markdown and Jupyter notebooks; compiled reports have been archived as PDFs. Where cross-checks have been needed, SPSS or Stata routines have been run to replicate key models. Encryption-at-rest and role-based access controls have been enforced through the storage environment.

FINDINGS

Across the pooled multi–case sample, the analysis has yielded a coherent pattern linking the depth and breadth of AI integration in BI dashboards to faster and more confident real-time decision making, with perceptual factors playing measurable roles. Response completeness has met preregistered thresholds, and scale reliabilities have been acceptable to strong; composite internal consistencies for perceived usefulness, interpretability, trust/reliance, and workload have generally exceeded conventional cutoffs, which has supported composite formation. Descriptively, user perceptions on five-point Likert scales have clustered in the upper half, with perceived usefulness and trust commonly averaging around the "agree" point (means ≈ 3.8–4.2), interpretability slightly lower but still favorable (means ≈ 3.6–4.0), and perceived workload centered near neutral to mildly elevated (means ≈ 2.8–3.2, where higher reflects more workload).

The AI Integration Index, standardized to 0–100, has shown meaningful variance across cases, with interquartile ranges wide enough to differentiate low-, medium-, and high-integration environments. Bivariate associations have aligned with expectations: AI integration has been negatively correlated with decision latency (faster decisions; medium magnitude) and positively correlated with decision accuracy/confidence (small-to-medium magnitude), while usefulness, interpretability, and trust have each exhibited favorable correlations with these outcomes. Workload has correlated weakly and positively with latency (more load, slower action) and weakly and negatively with accuracy/confidence, signaling potential attentional costs in denser interfaces. Turning to multivariate estimates with case controls and robust uncertainty, baseline models entering AI Integration alongside contextual covariates (size, industry, dashboard tenure, data quality, training, and analytics proficiency) have indicated that a one-standard-deviation increase in AI Integration has been associated with a substantively nontrivial reduction in decision latency (minutes decreased within the same shift/category) and a corresponding increase in user-reported accuracy/confidence. Extended models that have added perceptual constructs have improved explanatory power, with perceived usefulness consistently absorbing a portion of the AI effect on both outcomes. Mediation tests have supported an indirect pathway in which AI Integration has elevated perceived usefulness, which in turn has improved decision speed and confidence; indirect effects have been statistically distinguishable from zero under bootstrap intervals, while direct AI effects have remained directionally stable though attenuated. Moderation tests have further clarified boundary conditions: data quality has strengthened the association between AI Integration and outcomes (i.e., steeper gains in low-latency decisions and higher confidence at higher levels of timeliness, completeness, and definition clarity), and dashboard tenure has modestly amplified effects in mature deployments, suggesting that teams have capitalized on AI features more fully after operational bedding-in. Role-stratified contrasts have indicated broadly similar directions of effect for supervisors and analysts, with slightly larger usefulness-linked gains for supervisors whose actions have triggered workflow-wide changes. At the case level, harmonized operational KPIs (e.g., exception backlog clearance time, throughput variability, first-time-right rates) have moved in the favorable direction with higher AI Integration, even after standardization within case; while effect sizes have understandably been smaller than at the user-perception layer, the alignment has supported external performance coherence. Robustness and sensitivity procedures have underwritten these conclusions: results have persisted under heteroskedasticity-consistent covariance estimation, after removing influential observations, and under alternative operationalizations (ordered models for latency bands; log transforms for skewed measures; z-scaled perceptual composites). Multicollinearity diagnostics have remained within tolerable bounds, and leave-one-case-out re-estimations have produced coefficient distributions centered near the main estimates, reducing concern that any single case has driven the results. Exploratory visuals (coefficient plots, partial residuals, and marginal effects) have mirrored the tabular findings, especially the monotone decline in predicted latency as AI Integration has increased and the steeper slope of that decline under high data quality. Taken together, the introductory findings have indicated that AI-integrated dashboards operationalized as a graded index of predictive, prescriptive, anomaly-detection, conversational, and explainability capabilities, coupled with usage depth have been associated with faster and more confident operational decisions, that perceived usefulness has partially transmitted these gains, and that dependable data pipelines have conditioned the size of the benefits observed. These patterns have set the stage for the detailed subsections that follow, which have unpacked sample/case characteristics, full descriptive and correlation matrices (Likert-scale distributions and item summaries), primary and moderated regressions, and comprehensive robustness diagnostics.

Figure 7: Findings of the study initigrated with graph bar



Sample and Case Characteristics

Table 3 Sample and Case Characteristics (Cases A–F; User-Level N=168)

Case	Sector	Org. Size (employees)	Dashboard Tenure (months)	Data Quality (0–100)	Eligible Users	Respondents (n)	Response Rate (%)
A	Manufacturing	1,200	18	82	45	32	71.1
B	Logistics	650	14	76	35	27	77.1
C	Healthcare Ops	900	20	88	40	31	77.5
D	Tech-Enabled Services	2,100	24	73	60	41	68.3
E	Retail Fulfillment	780	12	69	38	24	63.2
F	Utilities Field Ops	1,450	22	84	55	13	23.6
Total/Mean			18.3	78.7	273	168	61.5

Across the six participating cases, the sampling frame has achieved both heterogeneity and comparability, and Table 3 has summarized these attributes to contextualize subsequent analyses. The participating organizations have spanned manufacturing, logistics, healthcare operations, technology-enabled services, retail fulfillment, and utilities field operations, and this spread has ensured that the dashboard-as-unit perspective has been examined under varied real-time demands and data-latency regimes. Tenure in production has ranged from 12 to 24 months (mean ≈ 18.3), which has signaled that the dashboards have been mature enough for users to have established routines while still recent enough that AI features have remained salient. The data-quality composite (0–100), assembled from timeliness, completeness, and definition clarity components, has averaged 78.7 with meaningful inter-case variation, a property that has been useful later when moderation by data quality has been tested. Recruitment has targeted active decision makers, and the study has achieved an overall response rate of 61.5% across 273 eligible users, yielding 168 analyzable responses. Response rates have been strongest in Cases A–C (≈71–78%), which has reflected embedded champion support and synchronized survey windows with shift schedules; Case F has shown a lower rate given a concurrent outage remediation effort, but it has still contributed a sufficient number of respondents to preserve cross-case comparability. Importantly, each case has cleared the a priori inclusion criteria continuous production use for at least one quarter, at least one AI capability visible to users, and stable data pipelines for core metrics so that the pooled dataset has reflected live operational environments rather than pilot sandboxes. The distribution of organizational size has further supported generalizability across small-to-large settings. By establishing these parameters, the section has laid out the replication logic of the design: the dashboard has been treated as a bounded decision interface appearing in different industrial habitats, and the sample has provided adequate within-case and between-case variation to estimate relationships with the desired precision. These characteristics have also aligned with the power plan articulated earlier, especially the targeted respondent counts per case and the minimum of six cases to support leave-one-case-out sensitivity analyses. In short, Table 4.1 has provided the evidential foundation for interpreting all downstream descriptive, correlational, and regression findings in a manner that has been faithful to the operational realities of the participating sites.

Descriptive Statistics

Table 4 Descriptive Statistics for Key Constructs (Likert 1–5 unless noted)

Construct	Items (α)	Mean	SD	Median	IQR	Scale Notes
Perceived Usefulness (PU)	5 (.89)	4.05	0.62	4.10	3.70–4.50	Higher = more useful
Perceived Interpretability (PI)	5 (.86)	3.82	0.66	3.80	3.40–4.30	Higher = clearer explanations
Trust/Reliance (TR)	5 (.88)	3.98	0.59	4.00	3.60–4.40	Higher = more calibrated trust
Perceived Workload (WL)	6 (.81)	3.05	0.74	3.00	2.50–3.60	Higher = more workload
AI Integration Index (0–100)	(.87)	63.4	14.8	64.0	54.0–74.0	Breadth × depth composite
Decision Latency (minutes)*		16.2	8.7	14.0	9.0–22.0	Lower = faster
Accuracy/Confidence (1–5)	4 (.84)	3.86	0.58	3.90	3.50–4.30	Higher = greater confidence
Data Quality (0–100)	(.83)	78.7	7.2	79.0	73.0–84.0	Higher = better quality

Table 4 has provided the pooled descriptive landscape for the constructs that have underpinned the analysis, reported on commensurate scales and accompanied by reliability coefficients to confirm composite integrity. The central tendency of perceived usefulness has clustered above the notional "agree" anchor (mean 4.05), and the standard deviation has indicated moderate dispersion, which has suggested that while most respondents have found AI-enabled dashboards helpful, there has remained room for incremental design gains. Interpretability has trailed usefulness modestly (mean 3.82), a profile that has been typical when explanations and uncertainty cues have added cognitive steps under time pressure; nevertheless, internal consistency has remained strong (α=.86), and the interquartile range has shown that a majority of respondents have rated explanation clarity positively. Trust/reliance has paralleled usefulness (mean 3.98), a correspondence that has foreshadowed the mediation results where usefulness has absorbed part of the AI integration effect. Workload has centered near the neutral point (mean 3.05) with slightly wider spread, which has aligned with qualitative notes indicating that alert density and narrative panels have sometimes compressed time available for action during peak periods. The AI Integration Index has exhibited healthy variance (SD ≈ 14.8 on 0–100), an important property that has enabled estimation of dose–response patterns in the regressions. Decision latency, expressed in minutes and derived from a mix of logs and midpoint-coded bands, has averaged 16.2 with an IQR from 9 to 22 minutes; this has been consistent with mixed discrete/continuous processes across the cases. Accuracy/confidence, treated on the same five-point frame as the perceptual composites, has averaged 3.86, and an internal consistency of .84 has supported composite formation. Finally, the data-quality composite has shown a favorable mean of 78.7 with sufficient spread to serve as a moderator. Together, these descriptive statistics have indicated that the measurement program has produced reliable, well-behaved variables that have occupied informative regions of their scales. The Likert-based constructs have demonstrated psychometric adequacy, and the objective or quasi-objective indicators (AI Integration, latency, data quality) have exhibited variance distributions compatible with multivariate modeling. This descriptive foundation has therefore justified the transition to correlation analysis to examine bivariate structure before proceeding to controlled regression estimates.

The correlation structure in Table 5 has provided an interpretable bivariate map that has been consistent with the theorized roles of AI integration and human–AI interaction factors in real-time decisions. AI Integration has correlated positively with perceived usefulness (r=.46) and trust (r=.41), with slightly weaker though still significant association with interpretability (r=.33), which has suggested that broader and deeper AI feature sets have tended to coincide with favorable perceptions of value and calibrated reliance. Importantly, AI Integration has shown a medium negative correlation with decision latency (r=−.38), meaning that more integrated dashboards have been associated with faster decisions at the user level; the relationship has not been so strong as to imply redundancy with other factors, which has justified multivariate controls. Workload has behaved as expected: it has related positively to latency (r=.22) and negatively, albeit modestly, to usefulness (r=−.19) and trust (r=−.21), indicating that interfaces that have felt mentally demanding have been linked to slower actions and lower perceived value. Accuracy/confidence has correlated in the favorable direction with AI Integration (r=.31) and more strongly with usefulness (r=.49) and trust (r=.46), a pattern that has foreshadowed the mediation role of usefulness in later models.

Correlation Matrix

Table 5 Pearson Correlations among Key Variables (N=168)

Variable	1	2	3	4	5	6	7	8
1. AI Integration (0–100)								
2. Perceived Usefulness (1–5)	.46**							
3. Interpretability (1–5)	.33**	.51**						
4. Trust/Reliance (1–5)	.41**	.58**	.47**					
5. Workload (1–5)	.12	−.19*	−.15	−.21*				
6. Decision Latency (minutes)	−.38**	−.45**	−.29**	−.34**	.22**			
7. Accuracy/Confidence (1–5)	.31**	.49**	.37**	.46**	−.18*	−.41**		
8. Data Quality (0–100)	.28**	.24**	.19*	.22**	−.10	−.26**	.20**	

*p < .05, ** p < .01 (two-tailed)*

Data quality has correlated positively with AI Integration and perceptual constructs and negatively with latency (r=−.26), which has supported the plausibility of its moderating influence: when pipelines have been timely and definitions clear, the benefits of AI features have been easier for users to convert into swift, confident interventions. Cross-correlations among the perceptual constructs (usefulness, interpretability, trust) have been moderate to strong (.47–.58), but not so extreme as to preclude simultaneous inclusion in regressions; subsequent models have therefore included variance-inflation checks and retained constructs given their incremental explanatory value. Finally, the negative correlation between latency and accuracy/confidence (r=−.41) has captured the intuitive link between speed and assurance in action under well-instrumented conditions. Collectively, the matrix has validated the measurement program, has illuminated relationships consistent with theory, and has motivated the regression specifications that have decomposed these associations while accounting for case effects and controls.

Regression Results (Primary & Moderation)

Table 6 OLS Regression on Decision Latency (minutes) with Case Fixed Effects (N=168)

Predictor	Model A1 (Baseline) β (SE)	Model A2 (+ Perceptuals) β (SE)	Model A3 (+ Moderation) β (SE)
AI Integration (per 10 pts)	−1.48** (0.39)	−0.84** (0.31)	−0.62* (0.30)
Perceived Usefulness (1–5)		−2.36** (0.58)	−2.11** (0.57)
Interpretability (1–5)		−0.71* (0.33)	−0.64* (0.32)
Trust/Reliance (1–5)		−0.48 (0.36)	−0.41 (0.35)
Workload (1–5)		+0.92** (0.28)	+0.88** (0.27)
Data Quality (per 10 pts)	−0.96* (0.41)	−0.58 (0.37)	−0.44 (0.36)
AI Integration × Data Quality			−0.28* (0.12)
Controls & Case FE	Yes	Yes	Yes
Adj. R^2	.29	.47	.51

Table 7 OLS Regression on Accuracy/Confidence (1–5) with Case Fixed Effects (N=168)

Predictor	Model B1 (Baseline) β (SE)	Model B2 (+ Perceptuals) β (SE)	Model B3 (Mediation Summary)*
AI Integration (per 10 pts)	+0.11** (0.03)	+0.05 (0.03)	Direct: +0.03 (0.03)
Perceived Usefulness (1–5)		+0.42** (0.07)	Indirect (AI→PU→Acc): +0.05**
Interpretability (1–5)		+0.12* (0.05)	
Trust/Reliance (1–5)		+0.18** (0.06)	
Workload (1–5)		−0.09* (0.04)	
Controls & Case FE	Yes	Yes	Yes
Adj. R^2 / Mediation	.21	.43	Bootstrapped 5,000 resamples

Tables 6 and 7 have presented the core estimation results and have demonstrated that AI Integration has been associated with faster decisions and higher confidence, with perceptual constructs and data quality shaping the magnitude of these associations. For decision latency (Table 4.4), the baseline model (A1) has shown that each ten-point increase in the AI Integration Index has been associated with a 1.48-minute reduction in latency ($p<.01$) after accounting for case fixed effects and controls. When perceptual variables have been added (A2), the AI coefficient has attenuated to −0.84 but has remained significant, while perceived usefulness has exhibited a strong independent association (−2.36 minutes per one-point increase on the five-point scale). Interpretability has contributed additional, smaller but significant reductions (−0.71), and workload has increased latency (+0.92), highlighting the practical trade-offs in dense interfaces. Introducing moderation (A3) has revealed a significant interaction between AI Integration and Data Quality (−0.28 per 10×10 points), indicating that gains in speed have been steeper where pipelines have been timelier and more complete. The adjusted R^2 has risen from .29 (A1) to .51 (A3), which has reflected the explanatory contribution of perceptual pathways and boundary conditions. For accuracy/confidence (Table 4.5), the baseline model (B1) has indicated a positive association with AI Integration (+0.11 per 10 points), and the extended model (B2) has shown that usefulness (+0.42), trust (+0.18), and interpretability (+0.12) have contributed uniquely, while workload has exerted a small negative effect (−0.09). In the mediation summary (B3), the direct AI effect has shrunk to +0.03 and has lost conventional significance when usefulness has been included, while the bootstrapped indirect effect via usefulness has remained significant (+0.05), thereby supporting the interpretation that AI features have improved confidence largely by increasing perceived usefulness of the dashboard. Across both outcomes, case fixed effects and robust standard errors have ensured that coefficients have reflected within-case contrasts net of unobserved heterogeneity. Diagnostics (reported later) have supported the validity of these estimates. Collectively, the models have established that AI Integration has mattered directly for speed and indirectly through usefulness for confidence, and that dependable data pipelines have amplified the speed benefits.

Robustness and Sensitivity Analyses

Table 8 Robustness Summary across Specifications (Decision Latency as Outcome)

Specification	AI Integration (per 10 pts)	Perceived Usefulness	AI×Data Quality	Adj. R^2	Notes
OLS + HC3 + Case FE (A3 main)	−0.62* (0.30)	−2.11** (0.57)	−0.28* (0.12)	.51	Main specification
OLS (exclude influential obs.)	−0.58* (0.29)	−2.05** (0.55)	−0.27* (0.12)	.52	Cook's D > 4/n removed
Ordered Logit (banded latency)	OR=0.86*	0.73**	0.82*	Pseudo-R^2 .27	Lower odds of slower band
Log-Transformed Latency	−0.037* (0.017)	−0.122** (0.032)	−0.016* (0.007)	.49	Coefs in log-minutes
Leave-One-Case-Out (k=6)	Median −0.60	Median −2.06	Median −0.26	.49–.53	Range of coefficients

Table 9 Robustness Summary across Specifications (Accuracy/Confidence as Outcome)

Specification	AI Integration (per 10 pts)	Perceived Usefulness	Interpretability	Trust	Adj. R^2	Notes
OLS + HC3 + Case FE (B2 main)	+0.05 (0.03)	+0.42** (0.07)	+0.12* (0.05)	+0.18** (0.06)	.43	Main specification
OLS (exclude influential obs.)	+0.05 (0.03)	+0.41** (0.07)	+0.11* (0.05)	+0.17** (0.06)	.44	Stable after exclusion
Standardized Composites (z)	+0.09 (0.05)	+0.31** (0.05)	+0.10* (0.04)	+0.15** (0.05)	.42	Effect sizes comparable
Add Role Interactions	+0.05 (0.03)	+0.43** (0.07)	+0.13* (0.05)	+0.16** (0.06)	.44	No material shifts
Mediation (bootstrapped)	Direct: +0.03	Indirect via PU: +0.05**				5,000 resamples; CI≠0

Tables 8 and 9 have consolidated robustness evidence demonstrating that the primary conclusions have persisted under alternative assumptions, transformations, and exclusion rules. For decision latency (Table 4.6), the main OLS model with HC3 standard errors and case fixed effects (A3) has delivered a −0.62-minute coefficient per ten points of AI Integration after accounting for perceptual pathways and the AI×Data Quality interaction. After removing influential observations defined by Cook's Distance greater than 4/n, the AI coefficient has remained directionally and substantively similar (−0.58), and both usefulness and the interaction term have preserved magnitude and significance, suggesting that the main findings have not been artifacts of a small number of leverage points. To address the bounded-category measurement used when logs have been unavailable, an ordered logit specification has been estimated; the odds ratio for AI Integration has indicated lower odds of falling into slower latency bands (OR≈0.86 per ten points), while usefulness has shown strong protective effects (OR≈0.73), and the interaction has remained favorable (OR≈0.82). A log-transform of latency has produced comparable inferences in proportional terms, improving residual symmetry without altering substantive interpretation. Finally, a leave-one-case-out procedure has yielded medians for AI, usefulness, and the interaction that have closely matched main estimates, and the adjusted R^2 has remained within a tight band (.49–.53), reinforcing the claim that no single case has driven the result. For accuracy/confidence (Table 4.7), the main OLS model (B2) has shown that usefulness, interpretability, and trust have had stable, positive associations with confidence, while the direct AI coefficient has been small and non-significant when these perceptual constructs have been present consistent with a mediated relationship. Excluding influential observations has not materially changed any coefficient. Re-estimating with standardized composites has yielded similar patterns and has facilitated effect-size comparability across scales. Adding interactions with role (e.g., supervisor vs. analyst) has not shifted point estimates meaningfully, signaling that the principal associations have generalized across user segments; exploratory plots have indicated slightly steeper usefulness slopes for supervisors, but formal interactions have remained non-significant in the pooled model. The mediation row has reiterated that the indirect path from AI Integration to confidence via usefulness has remained statistically different from zero under bias-corrected bootstrap intervals. Across both outcomes, these robustness checks have strengthened inferential confidence by showing that findings have not hinged on particular modeling choices, influential cases, or scale codings. Consequently, the evidence base for managerial interpretation has been broadened: integrating AI features has been associated with faster decisions directly and with more confident decisions primarily through elevating perceived usefulness, especially when the underlying data pipelines have exhibited high quality.

DISCUSSION

The findings have shown that higher levels of AI integration in BI dashboards have been associated with faster decision cycles and higher user-reported confidence, with perceived usefulness partially transmitting these effects and data quality strengthening them. This pattern aligns with a socio-technical reading of dashboards as decision interfaces where models, visuals, and routines co-produce performance. In particular, the negative association between the AI Integration Index and decision latency corroborates the premise that predictive, prescriptive, anomaly-detection, and conversational capabilities compress sense–decide–act loops when they are embedded in the surface where operators work (Sarikaya et al., 2019; Segel & Heer, 2010; Trkman et al., 2010). The mediated pathway through usefulness is consistent with long-standing adoption theory that task-contingent value beliefs are proximal precursors of use and outcomes (Davis, 1989; Venkatesh & Bala, 2008). The contribution of interpretability to both speed and confidence aligns with evidence that explanation quality and presentation clarity sustain situation awareness and calibrated reliance under time pressure (Endsley, 1995). Meanwhile, the positive association of workload with latency is theoretically sensible given cognitive resource limits in visually dense environments (Endsley, 1995). The moderation by data quality reinforces "fitness for use" as a boundary condition for BI value: timeliness, completeness, and definition clarity make AI outputs credible and thus actionable in the moment (Wang & Strong, 1996). Together, these results indicate that AI integration has mattered most when pipelines have been dependable, explanations have been legible, and the interface has preserved attention for signal over clutter an integrated reading that matches contemporary views of dashboards as engineered control instruments rather than static reports (Hart & Staveland, 1988; Hoff & Bashir, 2015).

Relative to earlier dashboard and analytics studies, the present evidence extends three threads. First, it quantifies an association between AI feature depth/breadth and real-time decision speed, adding to qualitative demonstrations of dashboard utility in operations and clinical contexts (Buttigieg et al., 2017). Second, it embeds explainability and conversational access in the measurement frame, echoing calls to move beyond accuracy to intelligibility and interaction affordances (Miller, 2019). Third, by modeling data quality as a moderator, it concretizes BI capability arguments that governance and integration shape realized value (Podsakoff et al., 2003; Popovič et al., 2012). The mediation by usefulness dovetails with TAM/UTAUT mechanisms while focusing on outcomes rather than mere intention (Davis, 1989). Our correlation and regression patterns mirror human–automation findings where appropriate reliance is a function of both performance history and transparency (Lee & See, 2004). Compared with forecasting and anomaly-detection literatures that emphasize algorithmic improvements (Makridakis et al., 2018), the present study situates those capabilities within a dashboard ecosystem, showing that their surface integration and the clarity of uncertainty and rationale relate to operationally meaningful endpoints like latency. Finally, the streaming and micro-batching architectures that underwrite "real-time" have been implicit controls in our design; the results resonate with systems work on handling state and fault tolerance for timely analytics (Babcock et al., 2002), implying that architectural adequacy is a prerequisite for front-end gains.

For CISOs, data leaders, and solution architects, the results crystallize a set of actionable priorities. First, treat data quality as a security-and-governance co-owned objective: timeliness and definition clarity have amplified the speed benefits of AI features, indicating that lineage, latency SLOs, and semantic catalogs are not "nice-to-haves" but determinants of decision value (Wang & Strong, 1996). Second, invest in explainability surfaces that fit the decision tempo. Local rationales, uncertainty bands, and contrastive cues presented as on-demand, low-friction affordances have related to faster and more confident action an implementation of "explain as needed" rather than "explain everything" (Miller, 2019). Third, manage workload as an interface risk: visual clutter and alert density have elongated latency, so architects should adopt layout and salience guidelines from visualization research (Sarikaya et al., 2019). Fourth, operationalize prescriptive analytics with guardrails. Where feasible, connect forecasts and anomalies to recommended actions with embedded constraints and confidence thresholds; prescriptive framing is known to translate predictions into value (Bertsimas & Kallus, 2020). Fifth, align access control and auditability with conversational/NL interfaces: these lower query costs but heighten governance needs for prompt injection resistance, authorized metric exposure, and logged rationale trails decisions that intersect both security and safety-by-design. Sixth, engineer for streaming resilience: consistent micro-batch latencies and degraded-mode fallbacks preserve trust when pipelines wobble (Babcock et al., 2002). Finally, instrument dashboards for usefulness telemetry (lightweight, privacy-safe pulse prompts) because usefulness has mediated confidence; feedback loops can guide targeted content and training where perceived value lags (Davis, 1989). Theoretically, the study advances dashboard and BI value models by proposing a pipeline-to-perception-to-performance pathway: dependable pipelines enable AI features; AI features elevate perceived usefulness; usefulness transmits benefits to decision speed and confidence, conditional on interpretability and workload. This refines capability-based views by specifying usefulness as a causal conduit rather than only an adoption antecedent (Pauwels et al., 2009). It also integrates human–automation and situation awareness theory into BI scholarship: interpretability and uncertainty communication are not merely usability attributes but SA supports that preserve comprehension and projection during time-bounded control (Endsley, 1995). The moderation by data quality formalizes a boundary condition often asserted but rarely measured in multi-case comparisons, supporting a contingent-resource view: analytics payoffs require "fit" between informational reliability and decision cadence (Isik et al., 2013). Additionally, our workload effect suggests a nuanced trade-off: adding analytic surfaces yields diminishing returns if cognitive load eclipses attentional bandwidth a perspective consistent with visualization memorability and clutter research (Isik et al., 2013). Finally, by locating prescriptive and conversational capabilities within a unified AI Integration Index, the study offers a measurement spine that future work can extend with weights tied to marginal prescriptiveness or intervention impact (Bertsimas & Kallus, 2020).

Several constraints temper inference. The cross-sectional design has precluded strong causal claims despite robustness checks; unmeasured case dynamics like leadership emphasis on continuous improvement may covary with both AI adoption and outcomes (Eisenhardt, 1989). Self-reported confidence and midpoint-coded latencies (when logs were unavailable) introduce measurement error, although convergent evidence from logged cases and harmonized KPIs mitigates concern. Instrument length constraints have limited the breadth of trust facets captured; richer scales from human factors could reveal finer-grained reliance patterns (Guidotti et al., 2018). Measurement invariance was addressed procedurally, yet scalar equivalence across sectors with different jargon may still be imperfect despite testing (Melnyk et al., 2004). Finally, while the sample spanned multiple industries and maturities, participation bias may favor teams with stronger champions or more stable pipelines. These limitations recommend caution in generalizing effect magnitudes beyond comparable operational contexts and motivate designs that tighten causal leverage, expand instrumentation, and link more outcomes to ground-truth logs.

Two complementary routes appear promising. First, longitudinal and quasi-experimental designs e.g., phased AI feature rollouts with difference-in-differences could attribute changes in latency or first-time-right rates to specific capabilities (Raykov, 1997). Second, mechanism experiments on the interface layer could isolate effects of explanation forms, uncertainty encodings, and alert bundling on workload and SA, building on visualization and uncertainty-communication literatures (Segel & Heer, 2010; Sivarajah et al., 2017). At the model layer, integrating "coefficient of prescriptiveness" and cost-of-delay metrics would tie dashboard recommendations to realized economic value (Bertsimas & Kallus, 2020). At the pipeline layer, research should operationalize latency SLOs for micro-batching/streaming and test their interaction with decision cadences (Borkin et al., 2013; Buttigieg et al., 2017). Finally, sector-specific extensions healthcare bed management, grid dispatch, or fulfillment slotting could blend domain KPIs with cross-case constructs to assess generalizability. Sharing replication packages with de-identified instruments and code would accelerate cumulative knowledge and sharpen measurement invariance practices (Raykov, 1997; Ribeiro et al., 2016). Bringing these threads together, the discussion positions AI-integrated dashboards as operational control rooms whose value emerges when dependable data pipelines, intelligible explanations, and workload-aware design converge. The present evidence indicates that broader and deeper AI capability sets have been linked to measurable improvements in decision speed and confidence, but that these gains have flowed largely through perceived usefulness and have been magnified by data quality. For practitioners, the guidance is concrete: budget for pipeline reliability and semantic governance; design for on-demand, low-friction explanations; temper alert volume with salience and visual economy; and connect predictions to prescriptive, constrained actions. For scholars, the contribution is a mechanism-aware, measurement-driven framework that ties pipeline characteristics to human perceptions and to operational outcomes, offering a scaffold for future causal and sectorial elaborations (Isik et al., 2013). In an era where "real-time" has been an architectural as much as a managerial claim, the findings reinforce that the fastest path to better decisions is not merely more analytics, but better-integrated analytics auditable in their inputs, intelligible in their outputs, and attentive to the limits of human attention at the moment of choice (Wang & Strong, 1996).

CONCLUSION

The study has concluded that AI-integrated BI dashboards, when conceived as decision interfaces rather than static reporting layers, have been associated with measurably faster and more confident real-time operational decisions across heterogeneous organizational cases. By operationalizing AI integration as a graded index that has combined functional breadth (forecasting, anomaly detection, prescriptive recommendations, conversational access, explainability) with usage depth, and by pairing this index with user-level Likert composites for perceived usefulness, interpretability, trust/reliance, and workload, the analysis has provided a coherent, quantitative account of how surface-level capabilities have translated into decision outcomes. Descriptive profiles have shown favorable central tendencies for usefulness and trust and adequate variance in the AI Integration Index; correlation patterns have aligned with theory, with integration linked to lower decision latency and higher confidence and with workload showing the anticipated drag on speed. Multivariate models with case controls have confirmed that AI integration has directly reduced decision latency and that perceived usefulness has partially transmitted its influence on accuracy/confidence, while interpretability has added incremental benefits and workload has imposed costs. Moderation tests have indicated that dependable data pipelines timeliness, completeness, and definition clarity have strengthened the speed advantages of integration, underscoring the necessity of pipeline reliability for realizing value at the interface. Robustness checks (influence exclusion, ordered models for banded latency, transformations, leave-one-case-out refits) have stabilized these conclusions, suggesting that the results have not been artifacts of measurement idiosyncrasies or single-case effects. Taken together, the evidence has supported a pipeline-to-perception-to-performance pathway: data quality has enabled AI capabilities; those capabilities have elevated perceived usefulness; and usefulness, under conditions of intelligible explanations and managed workload, has improved decision speed and confidence in the moment of operational control. The study has also clarified boundaries: integration without governance and semantic clarity has delivered smaller gains; interfaces that have overloaded attention have blunted otherwise promising analytics; and newer deployments have realized effects more modestly than mature ones. While cross-sectional constraints and partial reliance on self-reports in some cases have limited causal claims, triangulation with logs and harmonized KPIs has reinforced external coherence. Conceptually, the work has advanced comparative evaluation by treating the dashboard as the unit of analysis, offering a replicable measurement spine for integration, perceptions, and outcomes that future research can extend longitudinally or experimentally. Practically, the findings have crystallized priorities for implementation: invest first in pipeline reliability and semantic governance; expose AI outputs with on-demand, low-friction explanations; shape visual salience to minimize extraneous workload; and connect predictive insight to prescriptive, guardrailed actions. In sum, across diverse operational environments, the dashboards that have combined dependable data flow, well-integrated AI capabilities, and human-centered presentation have been the ones that have moved teams most reliably from signal to action with speed and assurance, offering a pragmatic template for organizations seeking real-time decision support that is not merely more analytical, but better integrated into the fabric of operational work.

RECOMMENDATIONS

Building on the pipeline-to-perception-to-performance pathway evidenced in this study, organizations have been advised to prioritize a sequenced, operationally grounded roadmap for AI-integrated BI dashboards that has converted analytics into faster and more confident action at the point of work. First, teams have formalized data-quality SLOs timeliness, completeness, and definition clarity at the metric level, and have tied them to ownership and monitoring so that the dashboards' "real-time" claims have been credible; semantic catalogs, lineage records, and freshness monitors have been embedded directly into the interface with unobtrusive status cues. Second, product owners have adopted an AI Integration Index as an internal scorecard to plan and track capability deployment forecast tiles, anomaly alerts, prescriptive playbooks, conversational access, and explainability widgets favoring increments that have the highest prescriptive leverage and demonstrable latency impact; releases have been staged with guardrails (thresholds, confidence bands, rollback switches) and change logs visible to operators. Third, designers have implemented on-demand explanations optimized for speed: compact local rationales, contrastive "why/why not" snippets, and uncertainty ranges have been presented where decisions have occurred, while deeper narratives have remained a click away; to modulate workload, layouts have reduced clutter, tuned alert salience to task criticality, and bundled low-severity notifications to protect attention during peak periods. Fourth, architects have strengthened streaming resilience through bounded micro-batch latencies, backpressure-aware ingestion, and degraded modes that have preserved essential KPIs when advanced features have been temporarily unavailable; observability for pipelines and models has been exposed as operator-readable health panels to maintain trust during incidents. Fifth, leaders have aligned governance and security with conversational/NL features by enforcing role-aware metric access, prompt hygiene, and auditable rationale trails; model risk controls (versioning, drift monitors, bias checks) have been integrated with incident and change-management workflows. Sixth, enablement has shifted from generic training to usefulness telemetry and targeted coaching: brief in-product pulse prompts have sampled perceived usefulness at the moment of use, and micro-lessons have been triggered when usefulness has lagged (e.g., “how to interpret this uncertainty band”); champions on each shift have facilitated peer learning and feedback triage. Seventh, operations have linked predictions to prescriptive playbooks with explicit constraints (capacity, safety, SLAs) so that recommended actions have been executable in one or two clicks, and they have measured the cost of delay to keep attention on decisions that materially affect outcomes. Eighth, teams have institutionalized evidence loops: quarterly, they have reviewed latency distributions, first-time-right rates, and exception backlogs alongside the Integration Index and data-quality SLO attainment, and they have retired low-value tiles to control cognitive load. Finally, to sustain legitimacy and portability, organizations have enforced measurement invariance across sites (consistent wording and scales), reported standardized effect sizes, and packaged de-identified instruments and code for internal replication; by doing so, they have kept the dashboard a living control instrument auditable in inputs, intelligible in outputs, and tuned to human attention so that each new AI increment has earned its place by demonstrably moving teams from signal to action with speed and assurance..

REFERENCES

[1]. Alpar, P., & Schulz, M. (2016). Self-service business intelligence. *Business & Information Systems Engineering*, *58*(2), 151-155. https://doi.org//10.1007/s12599-016-0424-6

[2]. Androutsopoulos, I., Ritchie, G. D., & Thanisch, P. (1995). Natural language interfaces to databases—An introduction. *Natural Language Engineering*, *1*(1), 29-81. https://doi.org//10.1017/s135132490000005x

[3]. Babcock, B., Babu, S., Datar, M., Motwani, R., & Widom, J. (2002). *Models and issues in data stream systems* Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems,

[4]. Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, *66*(3), 1025-1044. https://doi.org//10.1287/mnsc.2018.3253

[5]. Borkin, M. A., Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2306-2315. https://doi.org/10.1109/tvcg.2013.234

[6]. Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, *47*(5), 1287-1294. https://doi.org//10.2307/1911963

[7]. Buttigieg, S. C., Pace, A., & Rathert, C. (2017). Hospital performance dashboards: A literature review. *Journal of Health Organization and Management*, *31*(3), 385-406. https://doi.org//10.1108/jhom-04-2017-0088

[8]. Carifio, J., & Perla, R. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, *3*(3), 106-116. https://doi.org/10.3844/jssp.2007.106.116

[9]. Chae, B. K., Olson, D. L., & Sheu, C. (2014). The impact of supply chain analytics on operational performance: A resource-based view. *International Journal of Production Research*, *52*(16), 4695-4710. https://doi.org/10.1080/00207543.2013.861616

[10]. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, *41*(3), 1-58. https://doi.org/10.1145/1541880.1541882

[11]. Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1188. https://doi.org/10.2307/41703503

[12]. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233-255. https://doi.org//10.1207/s15328007sem0902_5

[13]. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. https://doi.org/10.1007/bf02310555

[14]. Danish, M. (2023). Data-Driven Communication In Economic Recovery Campaigns: Strategies For ICT-Enabled Public Engagement And Policy Impact. *International Journal of Business and Economics Insights*, *3*(1), 01-30. https://doi.org/10.63125/qdrdve50

[15]. 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

[16]. 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

[17]. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340. https://doi.org/10.2307/249008

[18]. Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532-550. https://doi.org//10.2307/258557

[19]. Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32-64. https://doi.org//10.1518/001872095779049543

[20]. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.2307/3151312

[21]. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, *51*(5), Article 93. https://doi.org//10.1145/3236009

[22]. Gunasekaran, A., Patel, C., & McGaughey, R. E. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics*, *87*(3), 333-347. https://doi.org//10.1016/j.ijpe.2003.08.003

[23]. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology* (Vol. 52, pp. 139-183). https://doi.org//10.1016/s0166-4115(07)70113-9

[24]. Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

[25]. Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407-434. https://doi.org//10.1177/0018720814547570

[26]. Isik, O., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management*, *50*(1), 13-23. https://doi.org/10.1016/j.im.2013.06.001

[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]. Jamieson, S. (2004). Likert scales: How to (ab)use them? *Medical Education*, *38*(12), 1217-1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

[29]. Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53-71. https://doi.org//10.1207/s15327566ijce0401_04

[30]. Kinkeldey, C., MacEachren, A. M., & Schiewe, J. (2014). How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, *51*(4), 372-386. https://doi.org//10.1179/1743277414y.0000000099

[31]. Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal/Annals of the American Association of Geographers*. https://doi.org/10.1080/00045608.2013.848024

[32]. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50.30392

[33]. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, *34*(4), 802-808. https://doi.org//10.1016/j.ijforecast.2018.06.001

[34]. 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

[35]. 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

[36]. 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

[37]. 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

[38]. 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

[39]. 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

[40]. 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

[41]. 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

[42]. 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

[43]. 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

[44]. 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

[45]. 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

[46]. Melnyk, S. A., Stewart, D. M., & Swink, M. (2004). Metrics and performance measurement in operations management: Dealing with the metrics maze. *Journal of Operations Management*, *22*(3), 209-218. https://doi.org//10.1016/j.jom.2004.01.004

[47]. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1-38. https://doi.org/10.1016/j.artint.2018.07.007

[48]. 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

[49]. 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

[50]. Neely, A., Gregory, M., & Platts, K. (1995). Performance measurement system design: A literature review and research agenda. *International Journal of Operations & Production Management*, *15*(4), 80-116. https://doi.org//10.1108/01443579510083622

[51]. Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625-632. https://doi.org/10.1007/s10459-010-9222-y

[52]. O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673-690. https://doi.org//10.1007/s11135-006-9018-6

[53]. 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

[54]. Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, *30*(3), 286-297. https://doi.org//10.1109/3468.844354

[55]. Pauwels, K., Ambler, T., Clark, B. H., LaPointe, P., Reibstein, D., Skiera, B., Wierenga, B., & Wiesel, T. (2009). Dashboards as a service: Why, what, how, and what research is needed? *International Journal of Research in Marketing*, *26*(2), 175-191. https://doi.org/10.1016/j.ijresmar.2009.06.001

[56]. Pimentel, M. A. F., Clifton, D. A., Clifton, L., & Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, *99*, 215-249. https://doi.org//10.1016/j.sigpro.2013.12.026

[57]. Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

[58]. Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Information & Management*, *49*(7–8), 317-326. https://doi.org/10.1016/j.im.2012.10.001

[59]. Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51-59. https://doi.org/10.1089/big.2013.1508

[60]. Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173-184. https://doi.org//10.1177/01466216970212006

[61]. 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

[62]. Reinking, J., Arnold, V., & Sutton, S. G. (2020). Synthesizing enterprise data through digital dashboards to strategically align performance: Why do operational managers use dashboards? *International Journal of Accounting Information Systems*, *36*, 100452. https://doi.org//10.1016/j.accinf.2020.100452

[63]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why should I trust you?": Explaining the predictions of any classifier* Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations,

[64]. 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

[65]. 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

[66]. 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

[67]. Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 682-692. https://doi.org/10.1109/tvcg.2018.2864903

[68]. Seawright, J., & Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly*, *61*(2), 294-308. https://doi.org//10.1177/1065912907313074

[69]. Seddon, P. B., Constantinidis, D., & Dod, H. (2017). How does business analytics contribute to business value? *Information Systems Journal*, *27*(3), 237-269. https://doi.org/10.1111/isj.12101

[70]. Segel, E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1139-1148. https://doi.org/10.1109/tvcg.2010.179

[71]. 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

[72]. Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Industrial Marketing Management*, *65*, 263-276. https://doi.org/10.1016/j.indmarman.2017.05.013

[73]. 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

[74]. Trieu, V. H. (2017). Getting value from Business Intelligence systems: A review and research agenda. *Decision Support Systems*, *93*, 111-124. https://doi.org/10.1016/j.dss.2016.09.019

[75]. Trkman, P., McCormack, K., de Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, *49*(3), 318-327. https://doi.org//10.1016/j.dss.2010.03.007

[76]. Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70. https://doi.org//10.1177/109442810031002

[77]. Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a research agenda on interventions. *Decision Sciences*, *39*(2), 273-315. https://doi.org/10.1111/j.1540-5915.2008.00192.x

[78]. Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356-365. https://doi.org/10.1016/j.jbusres.2016.08.009

[79]. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5-33. https://doi.org//10.1080/07421222.1996.11518099

[80]. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817-838. https://doi.org/10.2307/1912934

[81]. Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *Accounting Information Systems*, *13*(1), 41-59. https://doi.org/10.1016/j.accinf.2011.08.002

[82]. Zaharia, M., Das, T., Li, H., Hunter, T., Shenker, S., & Stoica, I. (2013). *Discretized streams: Fault-tolerant streaming computation at scale* Proceedings of the 24th ACM Symposium on Operating Systems Principles,