From AI Explanations to Human Decisions: A Structured Literature Review and Governance Taxonomy for Enterprise and Supply Chain Decision Support

Bismit Pratapsingh

ABSTRACT

Artificial intelligence (AI) is increasingly used in enterprise and supply chain decision support, yet explanation availability alone does not guarantee informed human decisions. This structured literature review critically connects research on explainable AI, human–AI collaboration, organizational adoption, and governance. It proposes an Explanation-to-Decision Governance (EDG) Taxonomy comprising five analytical dimensions: AI explanation, human interpretation, decision authority, cross-cutting governance controls, and operational outcomes. Drawing on selected scholarly literature and governance guidance, the paper distinguishes explanation comprehensibility from verifiability, appropriate reliance, and authorized intervention. It illustrates how the taxonomy can frame decisions in forecasting, inventory planning, and exception management and proposes evaluation questions for future research. The synthesis is narrative and selective rather than exhaustive: no complete database screening, reproducible corpus, or PRISMA flow counts are claimed. The EDG Taxonomy remains a conceptual contribution requiring independent empirical validation.

Keywords

Explainable Artificial Intelligence; Human–AI Collaboration; Enterprise Systems; Supply Chain Management; AI Governance; Decision Support; Structured Literature Review; Human Oversight

1. INTRODUCTION

Artificial intelligence (AI) is increasingly incorporated into enterprise information systems and supply chain management processes to support forecasting, inventory planning, procurement, logistics coordination, and operational decision-making. As organizations adopt machine learning, predictive analytics, and intelligent decision-support capabilities, attention is shifting from algorithmic performance alone toward the transparency, interpretability, and governance of AI-assisted decisions [1–3].

Explainable artificial intelligence (XAI) has emerged as an important research area for addressing the limited interpretability of complex AI models. Techniques such as feature attribution, local explanations, and interpretable modeling can provide insights into how algorithmic outputs are generated. However, the availability of an explanation does not necessarily establish whether decision-makers understand its implications, can challenge its recommendations, or possess the authority to intervene when operational circumstances require human judgment [6, 9–12, 18]. Foundational work on local explanations, feature attribution, and XAI taxonomies provides additional methodological context [9–11].

These concerns are particularly relevant in enterprise resource planning (ERP) and supply chain environments, where AI recommendations may influence interconnected business processes. Decisions involving inventory allocation, supplier selection, demand forecasting, and exception management frequently involve competing objectives, organizational constraints, and consequences that extend beyond the immediate prediction. Consequently, the effectiveness of AI-supported decision-making depends not only on model accuracy and explainability but also on the relationship between human interpretation, decision authority, accountability, and operational governance.

Existing research has examined AI adoption, explainability, human–AI collaboration, and intelligent supply chain applications from multiple disciplinary perspectives [7–9]. Nevertheless, it remains necessary to critically examine how these research streams address the transition from understanding AI recommendations to governing consequential enterprise decisions. In particular, the relationships among explanation quality, human oversight mechanisms, organizational accountability, and measurable decision outcomes warrant closer examination.

This paper conducts a structured, selective review of literature on explainable and human-centered AI in enterprise and supply chain decision support. Relevant peer-reviewed research published principally between 2020 and October 2026 is considered, together with earlier foundational studies and clearly identified governance standards. The analysis organizes recurring concepts and compares explanations, human decision rights, accountability mechanisms, and operational outcomes. It does not claim comprehensive coverage or reproducible systematic screening.

The intended contributions are threefold. First, the study synthesizes existing research on explainability and human involvement in enterprise decision-support systems. Second, it develops an evidence-informed taxonomy connecting AI explanations, human interpretation, decision authority, governance controls, and operational outcomes. Third, it identifies research gaps and future directions for the design and evaluation of accountable AI-assisted decision processes.

The remainder of this paper is organized as follows. Section 2 examines related research and identifies the motivating research gap. Section 3 describes the structured review approach. Section 4 presents the evidence synthesis. Section 5 develops the governance taxonomy. Section 6 discusses research gaps and implications, followed by the conclusion in Section 7.

2. RELATED WORK AND RESEARCH GAP

2.1 Artificial Intelligence in Enterprise and Supply Chain Systems

Artificial intelligence (AI) has become an important research area in supply chain management (SCM), supporting demand forecasting, inventory optimization, operational planning, and risk analysis. A systematic review of empirical AI applications in SCM identified five major research themes and emphasized the need to connect technological developments with established supply chain theory and practice [1]. Similarly, a review of 109 journal articles demonstrated the expanding application of

machine learning across forecasting, inventory management, and supply chain risk analysis [2].

Research on organizational AI adoption has examined the technological, organizational, environmental, and human factors influencing implementation. A systematic review of 68 studies applied an extended Technology–Organization–Environment (TOE) perspective to explain adoption drivers, barriers, and operational outcomes [3]. Although these studies establish the importance of organizational context, adoption-oriented frameworks do not necessarily specify how individual AI recommendations should be interpreted, challenged, or authorized within enterprise workflows.

2.2 Explainability and Human–AI Collaboration

Explainable artificial intelligence (XAI) seeks to improve the interpretability of algorithmic outputs and support informed human interaction. A systematic review of 117 publications examined XAI applications in SCM, including research themes, adoption drivers, and barriers [4]. Another review of 206 studies developed a human–AI collaboration framework spanning pre-adoption, adoption, and post-adoption activities in operations and supply chain management [5].

However, explainability and effective human decision-making are not equivalent. Fok and Weld [6] argue that explanations may fail to improve joint human–AI performance when users cannot verify whether an AI recommendation is correct. Evidence from another consequential decision-making domain also demonstrates that explanations may increase, decrease, or leave trust unchanged, depending on their quality and context [7]. These findings indicate the importance of distinguishing technical interpretability from appropriate reliance and meaningful human oversight. Experimental and design-oriented research further examines overreliance, decision support, and the limits of explanations [12, 13].

An important distinction across these streams concerns the unit of analysis. Supply chain reviews often examine adoption patterns, application categories, or aggregate performance, whereas human–AI experiments investigate individual responses to model advice. Neither level alone establishes whether a recommendation can be legitimately executed within an enterprise workflow. For example, an explanation may be intelligible to a planner while a procurement policy prohibits the proposed purchase. Conversely, a policy-compliant automated recommendation may be difficult for the responsible employee to verify. This mismatch motivates an analytical bridge from model-level explanation to workflow-level authorization [1, 3, 5, 6].

Experimental evidence also cautions against equating explanation exposure with decision improvement. Alufaisan et al. [12] found no conclusive improvement attributable to explanations in the evaluated decision tasks, while Bansal et al. [21] showed that explanations could increase acceptance of AI advice without improving complementary team performance. These results should not be generalized as proof that explanations are ineffective in supply chains; rather, they establish the need to measure error detection, reliance calibration, and decision consequences separately.

2.3 Comparative Analysis of Existing Research

Table 1. Comparison of representative literature reviews and conceptual contributions

Study	Research focus	EDG relevance
[1]	Empirical SCM AI	Operational setting
[2]	SCM AI/ML review	Application trends
[3]	AI adoption / TOE	Organization context
[4]	SCM explainability	Explanation evidence
[5]	Human–AI collaboration	Human participation
[6]	Verifiable explanations	Reliance and checking
[7]	XAI and trust	Trust calibration

2.4 Research Gap and Study Positioning

The reviewed literature provides substantial coverage of AI adoption, supply chain applications, explainability, and human–AI collaboration. Nevertheless, an important question for further investigation is how these research streams collectively address the transition from AI-generated explanations to governed operational decisions.

This study therefore examines five interconnected dimensions: (i) explanation availability and quality, (ii) human interpretation and verification, (iii) decision authority and intervention, (iv) governance and accountability mechanisms, and (v) operational decision outcomes.

Rather than proposing another AI adoption model, the review aims to develop an evidence-informed taxonomy that identifies how these dimensions are represented and connected in existing research. The taxonomy will distinguish established relationships from areas requiring further empirical validation, particularly within enterprise resource planning and supply chain decision-support environments.

3. STRUCTURED LITERATURE REVIEW APPROACH

3.1 Review Design and Research Questions

This structured narrative review uses four guiding questions to organize a critical comparison of selected scholarly literature. PRISMA 2020 [8] provides a useful reference for transparent reporting, but its systematic-review flow reporting is not applicable here because a comprehensive database search and formal screening log have not been completed. Accordingly, no PRISMA counts or claims of exhaustive study identification are reported.

RQ1: How is XAI used to support human decisions in enterprise and supply chain environments?

RQ2: What mechanisms support human approval, intervention, escalation, and accountability?

RQ3: How are trust, decision quality, and operational outcomes evaluated?

RQ4: What research gaps exist between technical explainability and governed decision-making?

3.2 Literature Identification and Scope

Relevant publications were identified selectively through accessible publisher and scholarly records and the bibliography assembled during manuscript development. The emphasis is on 2020–October 2026 literature concerning XAI, human–AI collaboration, governance, and supply chain decision support, with foundational works included where useful. Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and ScienceDirect were considered as possible sources for a future comprehensive

search; this paper does not claim that exhaustive searches were run across them.

Search expressions combine three conceptual groups: (i) explainability and human–AI collaboration, (ii) enterprise systems and supply chain management, and (iii) decision support and governance.

Illustrative search expression for future replication (not a documented completed database query):

("explainable AI" OR "interpretable AI" OR "human-AI collaboration") AND ("enterprise systems" OR "ERP" OR "supply chain") AND ("decision support" OR "human oversight" OR "AI governance")

This expression is provided for transparency and future systematic extension. It does not establish a PRISMA-eligible sampling frame.

3.3 Literature Relevance Criteria

Table 2. Relevance criteria used to guide the narrative selection (not formal screening counts)

Criterion	Included	Excluded
Period	2020–2026; key earlier work	Unrelated older work
Type	Peer-reviewed research	Promotional material
Topic	XAI; decisions; governance	Unrelated AI
Context	Enterprise, SCM, transferable	No decision relevance
Language	English	Other languages
Evidence	Methods or clear concepts	Unsupported assertions

3.4 Source Appraisal and Limitations

Sources were considered for conceptual relevance, publication venue, clarity of methods, and applicability to the four guiding questions. Empirical experiments, systematic reviews conducted by other authors, conceptual articles, and governance standards are distinguished in interpretation. No independent full-text quality scoring, inter-rater agreement, or exhaustive eligibility screening was performed; therefore, the synthesis cannot estimate study prevalence or the completeness of coverage.

The selection may favor accessible and frequently cited studies. Any future systematic review should prospectively register search dates, database queries, deduplication decisions, full-text exclusion reasons, and a reproducible evidence-extraction log.

Relevant studies can be coded across five dimensions: AI explanation mechanisms, human interpretation, decision authority, governance controls, and operational outcomes. Additional attributes will include application domain, research design, evaluation method, and reported limitations.

The structured narrative synthesis compares recurring patterns and differences across studies. The proposed governance taxonomy could be refined using the extracted evidence, with explicit distinctions between empirically supported relationships, conceptual proposals, and unresolved research questions. No meta-analysis is planned because substantial methodological and outcome heterogeneity is anticipated.

3.5 Synthesis Procedure and Evidence Boundaries

The narrative synthesis groups sources by explanation design (D1), user comprehension and appropriate reliance (D2), human control and intervention (D3), organizational policy and auditability (D4), and measured outcomes (D5). For instance, Bansal et al. [21] report that explanations increased acceptance of AI advice without increasing complementary performance in their experimental settings. The NIST AI Risk Management Framework [16] treats governance as cross-cutting, supporting the proposed placement of D4 across the decision process. Neither finding establishes causal effects in ERP or supply chain deployments; such transfer remains a research question.

Evidence status: the bibliography mixes empirical studies, existing reviews, conceptual contributions, and policy or standards documents. Inclusion in the bibliography does not imply a formally screened systematic-review sample.

4. EVIDENCE SYNTHESIS AND COMPARATIVE ANALYSIS

4.1 Explainability in Enterprise Decision Support

Explainable artificial intelligence (XAI) aims to make algorithmic predictions and recommendations more interpretable to human decision-makers. In enterprise and supply chain environments, AI applications support demand forecasting, inventory planning, supplier evaluation, and operational risk management [1, 2]. However, the usefulness of an explanation depends on whether it enables decision-makers to understand the recommendation and evaluate its relevance to operational circumstances.

Research on explanation verifiability suggests that providing explanations does not automatically improve combined human–AI decision performance [6]. This distinction is important in enterprise systems, where a technically interpretable prediction may still be insufficient to justify an operational decision. For example, a demand forecast may identify influential variables without adequately explaining whether a proposed inventory adjustment is appropriate under existing supply, contractual, or service-level constraints.

Consequently, this review distinguishes explanation availability from explanation usefulness and decision verification. These concepts provide a basis for examining how technical transparency contributes to operational decision-making.

4.2 Human Interpretation and Decision Authority

Human–AI collaboration research emphasizes the interaction between technological capabilities, human expertise, and organizational processes. Kumar and Kumar [5] examine collaboration across different stages of AI adoption in operations and supply chain management. Their findings provide a basis for considering human involvement as a continuing organizational activity rather than a single approval event.

For enterprise decision support, human involvement can be examined through four proposed decision responses: acceptance, modification, rejection, and escalation. These responses represent analytical categories for the present review, not empirically established frequencies or a validated classification.

A distinction is also necessary between human participation and meaningful decision authority. A user may receive an explanation without possessing the organizational permission, information, or technical capability to override the recommendation. The taxonomy therefore prompts investigation of whether published studies explicitly define human decision rights and intervention mechanisms.

An experimental counterpoint is provided by Bansal et al. [21], who found that explanations increased acceptance of AI recommendations regardless of correctness, without improving complementary team performance in their studied tasks. This cautions against treating user acceptance as evidence of calibrated trust or better operational outcomes. Transferability to ERP and supply chain decisions remains untested in this paper.

4.3 Governance and Accountability Mechanisms

AI adoption research identifies organizational readiness, technological characteristics, and environmental conditions as important implementation considerations [3]. Explainability research further highlights the relationship between explanations and trust, including circumstances in which explanations may fail to produce appropriate reliance [6, 7].

In enterprise applications, governance may involve approval rules, escalation procedures, access controls, accountability assignments, and decision traceability. These mechanisms are especially relevant when AI recommendations affect procurement commitments, inventory allocation, supplier relationships, or operational continuity. Human-centered AI guidance, risk-management standards, and regulatory developments offer complementary governance perspectives [14, 16, 17, 20].

The analysis seeks to distinguish governance mechanisms that are empirically evaluated from those discussed only as conceptual requirements. This distinction is essential because the existence of a proposed governance framework does not establish its operational effectiveness.

For governance analysis, the distinction between an explanation and an actionable control is especially consequential. The NIST AI Risk Management Framework emphasizes governance throughout AI risk management [16], while the EU AI Act provides a regulatory example of obligations associated with certain AI uses [17]. Neither document demonstrates that a specific enterprise deployment achieves better decisions. Their relevance here is normative: they motivate explicit responsibilities, documentation, monitoring, and escalation arrangements that can be evaluated empirically.

The proposed decision-control perspective therefore asks four operational questions: who can authorize an AI-supported action, what evidence must accompany the recommendation, what circumstances require escalation, and how can a later reviewer reconstruct the decision? These questions can be operationalized using approval records, override logs, uncertainty disclosures, and post-decision outcome measures. Their usefulness in a particular enterprise remains to be tested rather than assumed.

4.4 Comparative Evidence Analysis

Table 3. Preliminary Evidence-Dimension Matrix

Dimension	Source basis	Review question
Explanation	[4, 6, 11]	Can users verify?
Interpretation	[5, 7, 12]	Is reliance calibrated?
Authority	[3, 5, 21]	Can users intervene?
Governance	[14, 16, 17]	Are controls defined?
Outcomes	[1, 2, 6]	Are results evaluated?

Note: The matrix maps selected sources to analytical questions rather than counting studies. The sources differ in method and context: supply chain reviews [1–5] characterize applications and organizational conditions, human–AI studies [6, 7, 12, 21] examine interpretation and reliance, and governance guidance [14, 16, 17] specifies normative controls. These evidence types cannot be treated as interchangeable proof of operational effectiveness.

4.5 Cross-Dimensional Interpretation

The comparison reveals three important discontinuities. First, explanation methods and XAI studies address what a model communicates, whereas human–AI studies test whether people can assess or appropriately rely on that information [6, 7, 9–12]. An explanation that identifies influential demand variables may be understandable but insufficient to verify an inventory recommendation against current constraints. Thus D1 (explanation) and D2 (interpretation) should be evaluated separately.

Second, studies of human participation do not necessarily specify who may authorize an enterprise transaction [3, 5]. A planner may recognize a flawed forecast yet lack permission to change a purchase order; conversely, an authorized approver may lack evidence needed to challenge the recommendation. This separates D2 from D3 (decision authority). Third, acceptance of AI advice is not a reliable proxy for decision improvement: experimental findings on reliance and complementary performance [6, 12, 21] motivate measuring the final authorized action and its consequences (D5), rather than treating trust or acceptance as outcomes in themselves.

Across these discontinuities, D4 (governance controls) provides a cross-cutting analytical layer: policy thresholds, escalation routes, role permissions, and audit records may determine whether an understood recommendation becomes a permissible and traceable action [14, 16, 17]. This is a conceptual synthesis, not evidence that such controls cause better supply chain performance. The resulting research gap is the lack of directly comparable evidence connecting explanation quality, verification, authority, controls, and post-decision outcomes within the same enterprise workflow. Section 5 organizes these questions through the proposed EDG Taxonomy; prospective empirical testing remains necessary.

4.6 Evidence Transfer and Boundary Conditions

The selected literature differs in both empirical context and outcome definition. Reviews of AI in supply chain management characterize broad categories of use, adoption, and organizational performance [1–3], while controlled human–AI experiments frequently focus on individual classification or judgment tasks [6, 12, 21]. These bodies of evidence are complementary but not interchangeable. A finding about an individual's willingness to accept a prediction cannot, by itself, establish that a purchasing or inventory workflow is safe, accountable, or operationally efficient. This review therefore uses experimental studies to motivate hypotheses about decision behavior and uses supply chain reviews to identify plausible operational settings.

Three boundary conditions are especially important. First, the stakes and reversibility of an action vary: a replenishment suggestion may be adjustable before release, whereas a committed purchase order can have contractual consequences. Second, the informational context differs: a model may use historical demand data, while an experienced planner knows about a supplier disruption or upcoming promotion not yet represented in the data. Third, authority is distributed: an employee who understands the explanation may not be the person permitted to approve the resulting transaction. These considerations support separating the human interpretation dimension (D2) from decision authority (D3) and organizational controls (D4).

The literature also raises the question of explanation format. Feature-attribution methods such as LIME and SHAP can communicate influential variables [9, 10], but such attributions do not automatically communicate whether a recommendation respects a contractual restriction, operational threshold, or policy exception. For high-stakes settings, Rudin [18] argues for interpretable models rather than relying exclusively on post-hoc explanations of black boxes. The appropriate choice depends on the application and should be evaluated against decision tasks, stakeholder needs, and available alternatives [11, 19].

Accordingly, the synthesis does not infer a universal causal chain from explanations to improved enterprise outcomes. It instead identifies the information that future studies would need to collect to test that chain: what the system recommended, what explanation was shown, how the user interpreted it, which action was authorized, which governance rules applied, and what happened after execution. This evidence boundary is essential to distinguish the proposed EDG Taxonomy from a validated model of organizational performance.

4.7 Results of the Literature Synthesis

The narrative synthesis yields four qualitative findings aligned with the research questions. For RQ1, supply chain reviews document AI applications and XAI research describes explanation methods [1, 2, 4, 9–11]; however, these sources do not establish that explanation availability enables a planner to verify a proposed action. For RQ2, human–AI collaboration studies address participation and reliance [5, 6, 12, 21], whereas approval rights, escalation thresholds, and audit records are principally articulated as organizational or governance requirements [3, 14, 16, 17]. Evidence of their combined operation in an ERP workflow is not demonstrated by this review. Figure 2 summarizes these qualitative contrasts without implying measured effect sizes or study frequencies.

For RQ3, experimental studies caution that greater acceptance of AI advice need not improve complementary decision performance [6, 12, 21]; therefore, trust, acceptance, error detection, authorized action, and downstream outcomes must be assessed separately. For RQ4, the principal synthesis result is a gap between model-level explanation evidence, individual decision behavior, and workflow-level accountability. The five EDG dimensions organize this gap into testable questions rather than validated causal relationships. These findings reflect a selective narrative comparison, not measured frequencies, effect sizes, or results from an implemented enterprise system.

Literature synthesis: evidence and unresolved links

RQ	Literature addresses	Further evidence needed
RQ1	Explanation methods	Decision verification
RQ2	Human participation	Authorized intervention
RQ3	Advice acceptance	Operational decision quality
RQ4	Separate research streams	End-to-end workflow evidence

Qualitative synthesis only; no study counts or effect sizes inferred.

Figure 2. Qualitative comparison of literature coverage and remaining evidence gaps (RQ1–RQ4).

5. PROPOSED GOVERNANCE TAXONOMY

5.1 Taxonomy Development and Conceptual Foundation

Building upon the preliminary literature synthesis, this study proposes an Explanation-to-Decision Governance (EDG) Taxonomy for analyzing AI-assisted decisions in enterprise and supply chain environments. Existing research addresses explainability, human–AI collaboration, organizational adoption, and operational applications [1–7]. The proposed taxonomy brings these perspectives together to examine how algorithmic explanations may contribute to accountable organizational decisions.

The EDG Taxonomy comprises five interconnected dimensions: (D1) AI Explanation, (D2) Human Interpretation, (D3) Decision Authority, (D4) Governance Controls, and (D5) Operational Outcomes. These dimensions are intended as analytical categories for classifying existing research rather than as a validated operational architecture. Their definitions and relationships will be refined through future systematic evidence extraction [16, 17].

5.2 Explanation-to-Decision Governance Structure

The taxonomy conceptualizes AI-assisted decision-making as a sequence of interpretation and authorization activities, supported by governance controls and outcome evaluation.



Figure 1. Proposed Explanation-to-Decision Governance (EDG) Taxonomy.

The first dimension, AI Explanation (D1), concerns the information provided to communicate model predictions, recommendations, assumptions, and uncertainty. Human Interpretation (D2) examines whether decision-makers can understand, contextualize, and verify these outputs. Stakeholder-sensitive explanation design and the distinction between inherently interpretable and post-hoc explanations are relevant considerations [18, 19].

Decision Authority (D3) concerns the organizational and technical permissions available to human decision-makers, including accepting, modifying, rejecting, or escalating AI recommendations. Governance Controls (D4) encompass policies, accountability assignments, approval requirements, and decision traceability. Operational Outcomes (D5) represent the consequences of AI-assisted decisions, including decision quality, timeliness, reliability, and operational performance.

5.3 Taxonomy Dimensions and Analytical Indicators

Table 4. Proposed EDG Taxonomy Dimensions and Indicators

EDG	Analytical focus	Proposed indicator
D1	Explanation	Drivers; uncertainty
D2	Interpretation	Verification; reliance
D3	Decision authority	Approve; reject; escalate
D4	Governance	Policy; audit trail
D5	Operational outcome	Quality; time; reliability

Note: The indicators are proposed coding categories and have not yet been empirically validated.

5.4 Application to Enterprise and Supply Chain Decisions

The taxonomy can be applied to AI-supported processes such as inventory replenishment, demand forecasting, supplier evaluation, and supply chain exception management.

For example, an AI system may recommend an inventory adjustment based on projected demand. Under the proposed taxonomy, the explanation would be examined for interpretability and uncertainty disclosure (D1), while the user's ability to evaluate the recommendation would be classified under human interpretation (D2). The authority to approve, modify, or reject the adjustment would fall under D3. Applicable approval rules and accountability requirements would be classified under

D4, and resulting service levels, inventory performance, or decision delays would be considered under D5.

This illustrative example demonstrates how the taxonomy could organize evidence across technical, human, and organizational dimensions. It does not represent an implemented or empirically evaluated system.

Consider a hypothetical replenishment exception in which a predictive model recommends increasing an order quantity. Under D1, the reviewer receives the forecast horizon, key demand drivers, data recency, and uncertainty range. Under D2, the planner checks whether the demand signal reflects a one-off promotion and whether a known supplier disruption invalidates the assumptions. Under D3, the planner may approve, reduce, reject, or escalate the proposal according to delegated authority. D4 applies across those steps through purchase thresholds, segregation of duties, documented reasons for overrides, and retained audit records. D5 assesses whether the final action improved service levels without unacceptable excess stock or delay.

This example suggests measurable constructs for future evaluations. Explanation utility could be tested through the proportion of erroneous recommendations correctly challenged; decision authority through the proportion of exceptions handled by an authorized role; governance quality through audit completeness and policy deviations; and operational effectiveness through inventory turns, service levels, and resolution time. The metrics are proposed indicators, not findings from a deployed system.

5.5 Taxonomy Contribution and Limitations

The intended contribution of the EDG Taxonomy is to provide a structured basis for comparing research that addresses different stages of AI-assisted enterprise decision-making. Rather than treating explainability as sufficient evidence of trustworthy decision support, the taxonomy explicitly distinguishes explanation, interpretation, authority, governance, and outcomes.

Its principal limitation is that the current structure remains conceptual. Further validation requires systematic coding of eligible studies, assessment of whether the proposed dimensions adequately represent the literature, and empirical evaluation across enterprise and supply chain settings. The taxonomy should therefore be interpreted as a proposed analytical synthesis rather than an established causal model.

5.6 Illustrative Evaluation Protocol

An empirical assessment of the EDG Taxonomy could use a controlled enterprise decision-support simulation. Each participant would receive a set of supply chain exceptions with a recommended action, known operational constraints, and a reference outcome established by domain experts. The study would vary whether users receive an AI prediction alone, a prediction with an explanation, or a prediction with an explanation plus explicit approval and escalation controls. To avoid confounding, cases should be balanced for difficulty, model correctness, uncertainty, and the consequences of a wrong action.

The primary outcome should be the quality of the final human-authorized decision, not merely whether the participant accepted the AI recommendation. Secondary measures could include correct rejection of flawed advice, unnecessary override of correct advice, time to resolution, confidence calibration, and documented reasons for escalation. Where relevant, operational proxies such as projected stockouts, expedited freight, or inventory carrying cost could be assessed. User satisfaction and perceived trust should be reported separately from objective performance because favorable perceptions need not imply better decisions [6, 12, 21].

Governance should be operationalized rather than treated as a binary attribute. Candidate indicators include whether the acting user had the required authority, whether an exception exceeded an approval threshold, whether a second reviewer was required, and whether the final action left an auditable record. D4 should be modeled as applying across explanation delivery, interpretation, authorization, and subsequent monitoring. A system may have high explanation quality yet fail governance requirements if its escalation mechanism is bypassed or its action log is incomplete [16].

To assess robustness, future studies should recruit participants with different levels of ERP and supply chain expertise and evaluate both routine and unusual scenarios. Expert planners may benefit from concise explanations that foreground uncertainty, whereas less experienced users may need additional context and safeguards. Researchers should predefine success criteria and examine interaction effects rather than assume that the most elaborate explanation is always optimal. The design described here is a proposal for future validation; no participants were recruited and no experimental results are claimed in this article.

6. RESEARCH GAPS AND FUTURE DIRECTIONS

6.1 Emerging Research Gaps

The preliminary literature synthesis highlights several areas requiring further investigation. Although existing studies examine explainability, human–AI collaboration, organizational adoption, and supply chain applications [1–7], additional research is needed to establish how these dimensions interact within operational decision processes.

First, explanation effectiveness requires evaluation beyond technical interpretability. Research on explanation verifiability indicates that explanations do not necessarily improve joint human–AI performance [6]. Future studies should investigate whether explanations enable enterprise users to identify incorrect recommendations, evaluate uncertainty, and make appropriately calibrated decisions [12, 13, 15].

Second, human decision authority warrants closer examination. Human–AI collaboration frameworks address organizational and behavioral aspects of AI adoption [5], but the implementation and effectiveness of approval, override, rejection, and escalation mechanisms require further empirical investigation in enterprise workflows.

Third, governance integration presents an important research direction. Organizational adoption factors and explainability mechanisms should be examined alongside accountability assignments, policy enforcement, and decision traceability. This would help determine whether governance requirements translate into observable operational controls.

Finally, outcome-oriented evaluation remains essential. Future research should examine whether explainability and human oversight improve decision accuracy, timeliness, operational resilience, and accountability rather than relying exclusively on user perceptions or technical model performance.

6.2 Future Research Agenda

Table 5. Proposed Research Gaps and Future Research Directions

EDG	Research question	Potential measure
D1	Can errors be detected?	Error detection rate
D2	Does expertise matter?	Calibrated reliance
D3	When should AI be overruled?	Override accuracy
D4	Are controls embedded?	Policy / audit compliance
D5	Does performance improve?	Service level; lead time

Note: The research directions are proposed from the preliminary synthesis and require validation against the completed review corpus.

A practical research design would compare at least three workflow conditions: AI advice without an explanation, AI advice with an explanation, and AI advice with both an explanation and explicit human decision controls. Participants could complete realistic exception-management tasks under randomized conditions. Researchers could then measure error detection, decision time, unnecessary overrides, policy compliance, and final decision quality. Stratifying by user expertise and task uncertainty would help identify when explanations and governance controls are complementary rather than redundant [6, 12, 13, 21].

Longitudinal enterprise studies would provide a second form of evidence. They could trace how human override patterns, accountability assignments, and operational results evolve after deployment. Such studies should distinguish model prediction accuracy from the quality of the final authorized business action. They should also document potential costs, including approval delays and excessive escalation, so that stronger oversight is not incorrectly assumed to be cost-free.

6.3 Implications for Research and Practice

For researchers, the EDG Taxonomy offers a proposed structure for investigating relationships between AI explanations, human interpretation, decision authority, governance mechanisms, and operational outcomes. Future empirical studies could evaluate these dimensions individually and examine their interactions across different organizational contexts.

For practitioners, the taxonomy provides an analytical checklist for assessing AI-supported decisions in ERP and supply chain environments. Organizations considering AI-assisted procurement, inventory planning, or exception management can use these dimensions to examine whether explanations are understandable, decision rights are clearly assigned, governance controls are enforceable, and outcomes can be evaluated.

However, the taxonomy should not yet be interpreted as a validated implementation framework. Its practical effectiveness remains a subject for empirical investigation.

The proposed framework also raises implementation trade-offs that should be evaluated explicitly. Requiring manual approval for every low-risk recommendation may introduce delays without improving outcomes, while allowing unrestricted execution of high-impact recommendations may undermine accountability. Future research could therefore compare tiered decision-rights policies that route cases according to uncertainty, operational impact, and reversibility. Such policies should specify the conditions under which a user can proceed, when additional evidence is required, and when a decision must be escalated. These rules are potential governance design variables, not recommendations validated by the present review.

Another research opportunity concerns feedback after execution. Some decision-support evaluations emphasize whether a human accepts or rejects a recommendation, even though the downstream operational effect may only become visible days or weeks later. For supply chain settings, this creates a gap between immediate decision behavior and outcomes such as order fulfillment, inventory availability, supplier performance, and exception recurrence. A future longitudinal evaluation could link the initial explanation and approval record to later outcomes, enabling researchers to distinguish a defensible decision made under uncertainty from an outcome that happened to be favorable.

Finally, researchers should examine how the EDG dimensions interact rather than treating them as independent checklist items. Clear explanations may improve error detection only when users have sufficient expertise and time to assess them; effective escalation may depend on policy clarity and accessible evidence; and stronger audit controls may support learning from past mistakes while also increasing process overhead. Evaluating these interactions would clarify when governance supports human agency and when it becomes procedural friction. Such evidence would provide a stronger foundation for refining the taxonomy into an empirically tested decision-support framework.

A further consideration is the role of organizational evidence in explaining decisions. A technically accurate explanation may omit business information that is not available to the model, including supplier commitments, contractual service levels, or the operational consequences of a delayed shipment. A governed decision-support process should make those missing inputs visible where feasible and allow responsible users to document their assessment. This approach also creates a record that can later be evaluated for consistency, rather than treating the human override as an unexplained deviation from the model. The EDG dimensions provide a way to categorize such records without prescribing a single technology architecture.

From a research-method perspective, a follow-up systematic review would require a prespecified search protocol, recorded search dates, database-specific query adaptations, duplicate removal, full-text screening, and documented exclusion reasons. At least two independent reviewers would improve the reliability of study selection and evidence coding. A structured extraction sheet should distinguish study context, task type, AI capability, explanation format, human authority, governance mechanism, outcome measure, and limitations. The resulting corpus could test whether the proposed five dimensions are sufficiently comprehensive or whether additional constructs are necessary. The current selective review is intended to motivate that work, not to substitute for it.

In practice, the taxonomy may be most useful as a diagnostic lens during the design of AI-assisted workflows. A project team could ask whether an explanation supports verification, whether the intended user has the right to intervene, whether approval thresholds are enforceable, and whether post-action outcomes are captured. Gaps identified through this exercise would become hypotheses for pilot evaluation. Any organization applying the taxonomy should still conduct its own risk assessment and comply with applicable policies and legal requirements; the taxonomy itself is not a compliance certification or a substitute for operational testing.

7. CONCLUSION

The increasing integration of artificial intelligence into enterprise systems and supply chain management creates a need to examine how algorithmic explanations support accountable human decision-making. This paper presents a preliminary synthesis of research on explainable AI, human–AI collaboration,

organizational governance, and enterprise decision support. It proposes an Explanation-to-Decision Governance (EDG) Taxonomy comprising five interconnected dimensions: AI explanation, human interpretation, decision authority, governance controls, and operational outcomes.

The taxonomy distinguishes technical explainability from meaningful human oversight and organizational accountability. It offers a conceptual structure for investigating how AI recommendations are interpreted, authorized, governed, and evaluated in enterprise and supply chain environments.

The principal limitation is that comprehensive systematic searching and evidence-based validation were not undertaken. Accordingly, the taxonomy represents a proposed analytical synthesis rather than an empirically validated framework. Future research should examine its applicability through systematic evidence analysis and empirical evaluation across enterprise decision-making contexts.

By connecting explainability with human authority, governance, and operational outcomes, this study aims to support further research toward transparent, accountable, and human-centered enterprise AI systems.

8. REFERENCES

[1] Culot, G., Podrecca, M., and Nassimbeni, G. Artificial intelligence in supply chain management: A systematic literature review of empirical studies and research directions. Computers in Industry 162 (2024), 104132. https://doi.org/10.1016/j.compind.2024.104132

[2] Ferreira, A. C. A., Francisco, M. B., and Pinho, A. F. de (2025). The Use of Artificial Intelligence in Supply Chain Management: Systematic Literature Review and Future Research Directions. IEEE Access 13, 157828–157841. https://doi.org/10.1109/ACCESS.2025.3603866

[3] Shahzadi, G., Jia, F., Chen, L., and John, A. (2024). AI adoption in supply chain management: a systematic literature review. Journal of Manufacturing Technology Management 35(6), 1125–1150. https://doi.org/10.1108/JMTM-09-2023-0431

[4] El Jaouhari, A., and Driouiche, S. (2026). Towards an Explainable Artificial Intelligence in the Realm of Supply Chain Management: A Systematic Literature Review. In Transformative Impact of AI in Supply Chain Management, 161–184. https://doi.org/10.4018/979-8-3373-0923-1.ch007

[5] Kumar, N., and Kumar, R. R. (2025). Human–AI collaboration in operations and supply chain management: a systematic literature review. Management Review Quarterly. https://doi.org/10.1007/s11301-025-00575-9

[6] Fok, R., and Weld, D. S. (2024). In search of verifiability: Explanations rarely enable complementary performance in AI-advised decision making. AI Magazine 45(3), 317–332. https://doi.org/10.1002/aaai.12182

[7] Rosenbacke, R., Melhus, Å., McKee, M., and Stuckler, D. (2024). How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Health Care: Systematic Review. JMIR AI 3, e53207. https://doi.org/10.2196/53207

[8] Page, M. J., et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372 (2021), n71. https://doi.org/10.1136/bmj.n71

[9] Ribeiro, M. T., Singh, S. and Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD (2016), 1135–1144. https://doi.org/10.1145/2939672.2939778

[10] Lundberg, S. M. and Lee, S.-I. A Unified Approach to Interpreting Model Predictions. NeurIPS (2017). https://arxiv.org/abs/1705.07874

[11] Arrieta, A. B., et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58 (2020), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[12] Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y., and Kantarcioglu, M. (2021). Does Explainable Artificial Intelligence Improve Human Decision-Making? Proceedings of AAAI 35(8), 6618–6626. https://doi.org/10.1609/aaai.v35i8.16819

[13] Buçinca, Z., Malaya, M. B., and Gajos, K. Z. (2021). To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. Proceedings of the ACM on Human-Computer Interaction 5(CSCW1), Article 188, 1–21. https://doi.org/10.1145/3449287

[14] Amershi, S., et al. Guidelines for Human-AI Interaction. CHI (2019). https://doi.org/10.1145/3290605.3300233

[15] Doshi-Velez, F. and Kim, B. Towards A Rigorous Science of Interpretable Machine Learning (2017). https://arxiv.org/abs/1702.08608

[16] Tabassi, E. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1 (2023). https://doi.org/10.6028/NIST.AI.100-1

[17] European Union. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act). Official Journal of the European Union (2024). https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[18] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1 (2019), 206–215. https://doi.org/10.1038/s42256-019-0048-x

[19] Suresh, H., et al. Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and their Needs. CHI (2021). https://doi.org/10.1145/3411764.3445088

[20] Shneiderman, B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. International Journal of Human–Computer Interaction 36(6) (2020), 495–504. https://doi.org/10.1080/10447318.2020.1741118

[21] Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., and Weld, D. S. Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. Proceedings of CHI 2021, article 81, pp. 1–16. https://doi.org/10.1145/3411764.3445717