



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

Doctor of Business Administration (DBA)

LCDB7007 – Research Principles and Philosophy

**“AI-Enabled Forecasting in Retail Supply Chains: Forecast Accuracy,
Replenishment and Human–AI Decision Integration”**

Student Name: Bitto Xavier

Student ID: 2528392

Submitted To: Dr. Ebad Banissi

Word Count: 7,508 words (excluding references, tables and appendices)

Abstract

Purpose: This report critically examines how research philosophies shape knowledge claims about artificial intelligence (AI)-enabled demand forecasting, replenishment performance and human–AI decision integration in retail supply chains. **Design/methodology/approach:** A PRISMA informed critical review was undertaken, centered on peer reviewed research published predominantly between 2021 and 2026 and supplemented by justified seminal philosophy, forecasting and human algorithm literature. The review evaluates ontological, epistemological and axiological assumptions, methodological quality and ethical implications, followed by a detailed analysis of Wang et al. (2024). **Findings:** The literature is weighted towards measurable predictive and operational outcomes, giving many studies positivist or post-positivist characteristics. Yet improved forecast accuracy does not automatically generate superior inventory or organisational outcomes. Value realisation is conditioned by inventory policy, demand uncertainty, data quality, organisational capability, governance and the way human decision-makers interpret, trust, modify or reject algorithmic advice. **Originality/value:** The report reframes AI forecasting as a socio-technical and philosophically conditioned value-realisation problem. It argues for a problem-centred DBA orientation that retains rigorous performance evidence while examining meanings, mechanisms, context and responsible human oversight.

Keywords: artificial intelligence; demand forecasting; human–AI collaboration; research philosophy; retail supply chains; replenishment

Table of Contents

Abstract.....	2
1. Introduction.....	5
2. Method for Conducting the Literature Review	7
3. Critical Literature Review	10
3.1 Philosophical and theoretical foundations	10
3.2 From forecast accuracy to replenishment and operational value.....	13
3.3 Human–AI decision integration, organisational context and ethics.....	16
3.4 Synthesis, contradictions and research gaps	19
4. Philosophical and Methodological Analysis of Wang et al. (2024).....	20
4.1 Study rationale, problem and contribution	20
4.2 Ontological assumptions	21
4.3 Epistemology	21
4.4 Axiology and researcher positionality	22
4.5 Ethics	23
4.6 Methodological coherence and design quality.....	23
4.7 Reliability, validity and trustworthiness	24
4.8 Theoretical framing and alignment.....	24
4.9 Alternative philosophical approaches	25
4.10 Overall evaluation.....	26
5. Discussion and Contribution	27
6. Conclusion.....	29
References	31
Appendices.....	34
Appendix A. Search concept families	34
Appendix B. Extended evidence-extraction fields	34

List of Tables

Table 1: Search and review protocol.....	8
Table 2: Paradigm comparison for AI-enabled forecasting research	12
Table 3: Critical synthesis of key contemporary studies	16
Table 4: Research gaps, philosophical significance and DBA implications.....	20
Table 5: Philosophical and methodological audit of Wang et al. (2024).....	26
Table 6: Alternative philosophical redesigns of the focal problem	27

List of Figures

Figure 1: PRISMA-informed critical review workflow. Source: Author-developed review protocol informed by Page et al. (2021).	9
Figure 2: Philosophical architecture of AI-enabled forecasting and replenishment research. Source: Author-developed synthesis from the reviewed literature.	13
Figure 3: Evidence architecture for the future DBA inquiry. Source: Author-developed synthesis.	29

1. Introduction

Retail forecasting increasingly sits at the intersection of data science and managerial judgement. Machine-learning systems can process large, heterogeneous data and reveal demand patterns that are difficult to capture through conventional models, but the managerial problem begins rather than ends when a forecast is produced. Predictions must be translated into order quantities, safety-stock settings and exception decisions within supply chains constrained by lead times, capacity, cost structures and organisational routines. Recent reviews expose a paradox: AI-based demand planning is expanding rapidly, yet the evidence base remains strongly modelling-oriented and organisational questions about adoption, coordination and value realisation remain underdeveloped (Walter, Ahsan and Rahman, 2025; Toorajipour et al., 2021).

Predictive accuracy and operational value are not equivalent constructs. Forecasting research has often evaluated prediction as an endpoint, while inventory-control research has treated demand information as an input to a separate decision problem. Goltsos et al. (2022) identify this separation as a persistent forecasting inventory gap. Theodorou, Spiliotis and Assimakopoulos (2025) show directly that the forecasting method with the best statistical accuracy need not minimise inventory-related costs across all conditions: the relationship changes with holding and lost-sales costs, review periods, lead times and demand characteristics. Lower forecast error is therefore evidence of technical performance, but not by itself evidence of higher availability, lower working capital or greater resilience.

Human decision-makers make this translation problem more complex. Planners interpret algorithmic recommendations, apply contextual knowledge and sometimes override them. Revilla et al. (2023), in a retail field experiment, show that the value of human intervention depends on uncertainty and forecast horizon rather than being universally beneficial or harmful. Algorithm-aversion research is similarly non-linear: people may reject algorithms after observing errors, accept them when they retain some control, or prefer algorithmic advice depending on task framing and expertise (Dietvorst, Simmons and Massey, 2015; Dietvorst, Simmons and Massey, 2018; Logg, Minson and Moore, 2019). Fügener et al. (2022) further show that failed delegation can arise from weak knowledge about relative human and AI capability rather than aversion alone. Transparency is also contingent: Vössing et al. (2022) found that information

about AI reasoning increased trust while information about uncertainty reduced it. Human–AI integration is therefore an explanatory mechanism, not merely an adoption issue.

Organisational conditions add another layer. Cultural enablers influence supply-chain AI integration (Cadden et al., 2022); managers' willingness to use AI reflects benefits, facilitating conditions and concerns (Cao et al., 2021); and coordination can mediate relationships between AI use and supply-chain performance (Li et al., 2024). Data quality, system integration, capability, authority and governance can therefore determine whether sound predictions are enacted. A socio-technical view is appropriate because forecasting systems comprise technology, people, tasks, routines and structures rather than a standalone algorithm (Makarius et al., 2020).

These empirical tensions create a research-philosophy problem. Studies operationalising success through forecast error, inventory cost or service outcomes often exhibit positivist or post-positivist characteristics because reality is approached through observable variables and statistical inference. This is valuable for determining whether a model performs, but can under-represent meanings such as trust, legitimacy and accountability. Interpretivism can illuminate how decision-makers understand algorithmic advice, while critical realism directs attention to mechanisms and contextual conditions that generate different outcomes from similar technologies (Frederiksen and Kringelum, 2021). Pragmatism may offer an applied orientation when the problem requires different forms of warranted evidence; however, Morgan (2014) cautions against treating pragmatism merely as a synonym for mixed methods.

Accordingly, this report critically reviews recent research on AI/ML demand forecasting, replenishment and human–AI decision integration in retail supply chains. It asks how philosophical assumptions shape valid evidence, whether greater accuracy translates into operational performance, how human and organisational conditions affect value realisation, and which philosophical positioning can support a coherent future DBA design. Particular attention is paid to ontology, epistemology, axiology, ethics, reliability and validity. Following a transparent review method, the report develops a thematic synthesis and then examines Wang et al. (2024) in depth at the intersection of algorithmic performance, human judgement and disruption. The discussion integrates implications for theory, methodology and practice before returning to the central issue: under what conditions does predictive capability become responsible organisational value?

2. Method for Conducting the Literature Review

This review was designed as a PRISMA-informed critical literature review rather than as a claim of exhaustive or registered systematic-review coverage. PRISMA 2020 was used as a transparency discipline for identification, screening and reporting (Page et al., 2021), while the analytical purpose remained interpretive and problematising. This follows a problematisation logic that questions field assumptions rather than merely locating omissions (Alvesson and Sandberg, 2011), and a conception of criticality that requires scrutiny of the grounds on which knowledge claims are made (Mingers, 2000). The search protocol targeted Scopus, Web of Science Core Collection, ScienceDirect and Emerald Insight, with Google Scholar used supplementarily for discovery and backward/forward citation chaining. Topical searches focused predominantly on 1 January 2021 to 23 August 2026, reflecting the assignment requirement for current research. Older work was retained only where it was foundational to research philosophy, socio-technical theory, forecasting or human–algorithm judgement.

Search families combined terms for artificial intelligence, machine learning and predictive analytics with demand forecasting/planning, retail or supply chain, replenishment, inventory, service level and stock-outs. Complementary strings captured human–AI collaboration, algorithm aversion/appreciation, judgemental adjustment, trust and explainability; organisational terms such as culture, readiness, capability, integration and governance; and philosophical terms including positivism, interpretivism, pragmatism, critical realism, ontology, epistemology, axiology and reflexivity. The full concept strings are provided in Appendix A. Search decisions and source verification were recorded in an evidence matrix so that inclusion, philosophical coding and later citation checks could be traced consistently.

Peer-reviewed empirical studies and rigorous reviews were prioritised where they addressed the technical, human or organisational dimensions of the problem. Generic AI engineering papers without managerial relevance, unrelated digital-transformation studies, promotional sources and weakly reported evidence were excluded from the analytical core. Authoritative governance documents were used only to support ethical discussion. Screening proceeded through duplicate removal, title/abstract relevance, full-text eligibility and critical appraisal (Figure 1). Because this is a critical coursework review rather than a reproducible systematic review, numerical database yields are deliberately not presented; no PRISMA counts have been invented.

Each core study was mapped for context, theory, data/sample, method, performance constructs, human–AI factors, limitations, quality and DBA relevance. Philosophical positions were recorded

when explicitly stated and otherwise inferred cautiously from the design and claims. Analysis followed the lecturer's critical sequence of scanning, mapping, comparing, synthesising and positioning. Quality appraisal considered methodological transparency, reliability or trustworthiness, construct/internal/external validity, ecological relevance, ethics and contribution rather than applying a mechanical exclusion score. Limitations include English-language restriction, database and publication bias, rapidly changing AI terminology, heterogeneous performance measures and interpretive judgement when authors do not declare a paradigm. Cross-database verification, citation chaining, explicit distinction between stated and inferred philosophy, and an auditable source matrix were used to mitigate these risks.

Element	Protocol
Review design	PRISMA-informed critical literature review; transparent but not presented as an exhaustive registered systematic review
Primary sources	Scopus; Web of Science Core Collection; ScienceDirect; Emerald Insight
Supplementary discovery	Google Scholar; backward and forward citation chaining; publisher/DOI verification
Primary period	1 January 2021–23 August 2026; older works only when foundational or methodologically necessary
Core topics	AI/ML forecasting; demand planning; replenishment/inventory; human–AI integration; trust/overrides; organisational context; research philosophy
Inclusion	Peer-reviewed empirical studies and rigorous reviews with direct or transferable managerial relevance; English language
Exclusion	Generic engineering without managerial relevance; unrelated digital transformation; promotional evidence; weak reporting
Analysis	Scanning → mapping → comparing → synthesising → positioning; philosophical and methodological coding
Quality criteria	Relevance; methodological transparency; reliability/trustworthiness; construct/internal/external validity; ecological relevance; ethics/reflexivity
Reporting safeguard	No invented PRISMA counts; stated philosophy distinguished from philosophy inferred from research design

Table 1: Search and review protocol

PRISMA-informed critical review workflow

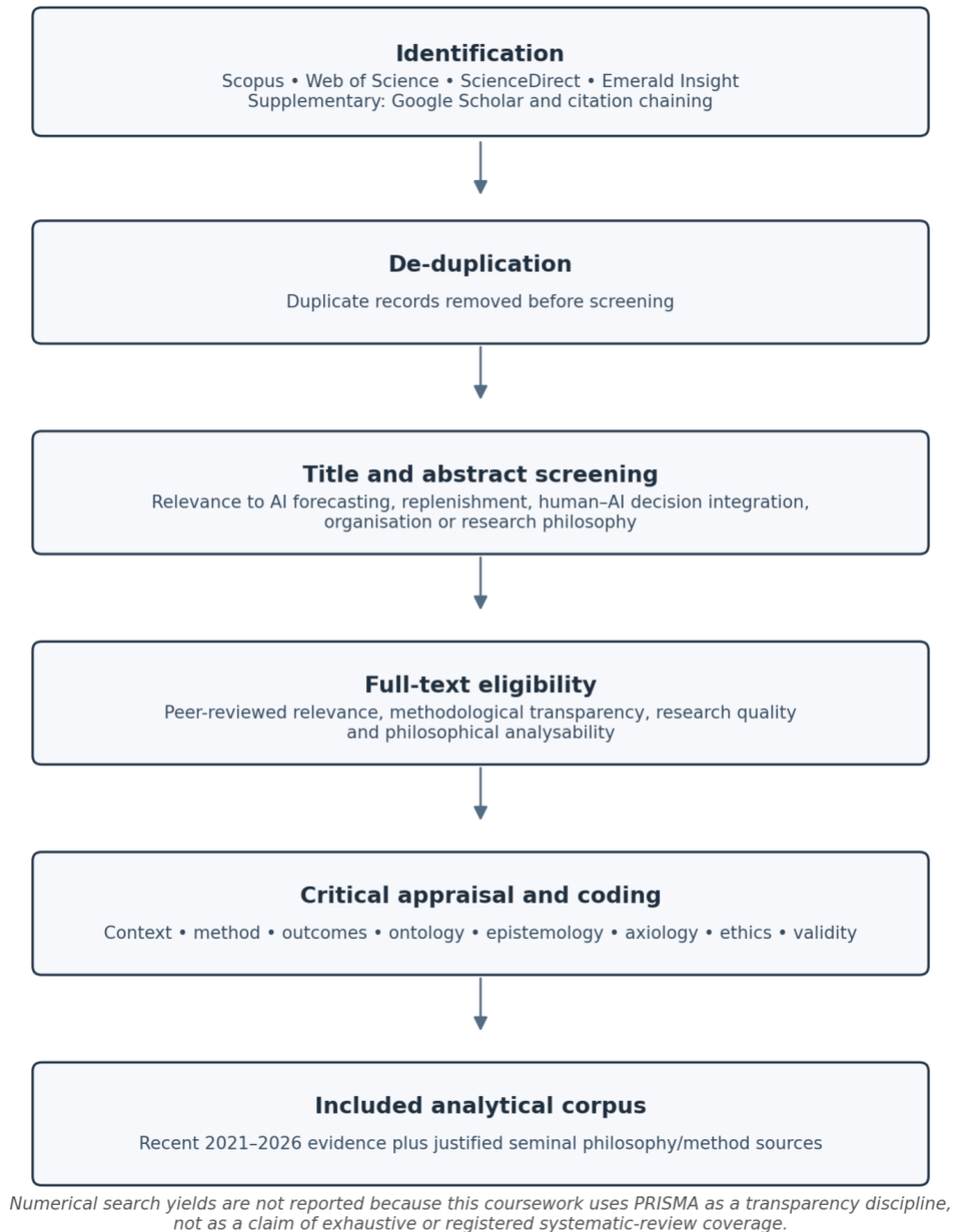


Figure 1: PRISMA-informed critical review workflow. Source: Author-developed review protocol informed by Page et al. (2021).

3. Critical Literature Review

3.1 Philosophical and theoretical foundations

Research philosophy matters in AI-enabled forecasting because it determines what is treated as real, what counts as credible knowledge and which outcomes are worth optimising. A forecasting system can be studied as an objectively measurable predictor, as an artefact interpreted by planners, as one component within deeper organisational mechanisms, or as a practical intervention whose value depends on consequences. Each framing makes some dimensions visible and pushes others to the margins.

A positivist or, more commonly, post-positivist orientation is readily compatible with forecasting because demand, error, inventory and service can be operationalised numerically. Benchmarking models and testing relationships assumes that observable indicators can provide fallible but increasingly reliable knowledge of an external reality. This enables replicable comparison, disciplined measurement and causal or associational inference. Yet the strength can become epistemic narrowing. When forecast error is the principal dependent variable, prediction can silently become the purpose of forecasting. Goltsos et al. (2022) show why this is inadequate, while Theodorou, Spiliotis and Assimakopoulos (2025) demonstrate that accuracy and inventory rankings can diverge. Quantitative evidence can therefore be internally persuasive while remaining incomplete as an explanation of organisational value.

Post-positivism is a particularly useful qualification because contemporary empirical research rarely assumes perfect observation or absolute verification. Forecasts are probabilistic, datasets are incomplete and models remain exposed to structural change. Reliability, robustness checks and out-of-sample testing are therefore not philosophical trivia: they are practical devices for acknowledging fallibility. The limitation arises when methodological rigour within a narrow construct is confused with completeness of the phenomenon. A highly reliable estimate of MAPE can still tell little about why an organisation failed to change its reorder point or why a planner ignored an exception.

Interpretivism starts from a different problem. Managers encounter recommendations within settings in which expertise, accountability and experience shape meaning. Trust is not independent of context: Glikson and Woolley (2020) identify reliability and transparency as important antecedents, while Cao et al. (2021) show that organisational conditions and personal concerns influence managers' intentions to use AI. Interpretive inquiry can reveal why apparently identical recommendations are understood differently or why an override is considered

legitimate. Its limitation is the reverse of the positivist problem: rich accounts of meaning cannot establish that a forecast is more accurate or that behaviour improved inventory performance.

Critical realism offers a useful bridge by distinguishing observable events from mechanisms and structures that generate them. In management research it asks why the same intervention can produce different outcomes under different conditions (Frederiksen and Kringelum, 2021). Forecast error and stock-outs are observable, but their relationship may depend on data quality, replenishment policy, planner authority, organisational slack or learning routines. An override may reflect bias, contextual expertise, weak model fit or institutional accountability. Critical realism therefore accommodates material demand and inventory consequences while recognising contingent causation through organisational structures.

Pragmatism organises inquiry around a consequential problem. For an applied DBA, the managerial question is not exhausted by either 'Does the model reduce error?' or 'How do planners experience it?'; it concerns what evidence is required to determine whether, how and when the system improves decisions. Morgan (2014) warns, however, against a shallow pragmatism in which any method mixture is justified because it appears useful. Coherence still requires clarity about the status of different knowledge claims and how they are integrated. Mixed methods can be used within several traditions; their presence does not itself establish pragmatism.

The distinction also prevents a common methodological shortcut in DBA research. A survey followed by interviews may be technically 'mixed', yet if the qualitative strand is used only to illustrate a pre-existing quantitative conclusion, the design may remain epistemologically hierarchical. Conversely, a critical-realist study may legitimately combine numerical outcomes with interviews because both are used to infer mechanisms. Philosophical positioning should therefore be judged from assumptions about reality, knowledge and explanation rather than from the surface form of the methods (Saunders, Lewis and Thornhill, 2019).

Axiology sharpens the debate. 'Better' can mean lower error to a data scientist, lower working capital to finance, higher availability to customers or an accountable decision process to planners. These ends can conflict. Teleologically, the purpose of forecasting is therefore contestable: statistical loss minimisation, inventory economics, resilience and responsible judgement are not identical objectives. Choices about data, loss functions, override rights and accountability encode values before a model reaches deployment.

Socio-Technical Systems thinking provides the primary organisational lens. AI forecasting joins a technical subsystem (data, models and interfaces), a social subsystem (planners, managers,

expertise and trust), tasks (forecasting, exception handling and replenishment) and organisational arrangements (authority, governance and coordination). Makarius et al. (2020) similarly frame AI integration as a socio-technical challenge. Dynamic capabilities and supply-chain integration are useful secondary lenses for explaining reconfiguration and cross-functional coordination, but only where they illuminate the translation mechanism. Figure 2 consolidates this architecture by placing measurable predictive performance upstream of human–AI integration, replenishment and organisational value, all embedded in contextual conditions.

The four paradigms in Table 2 should therefore be read as competing lenses with different quality criteria, not as a menu from which methods can be selected without consequence. Positivist/post-positivist studies emphasise reliability, measurement and internal/external validity; interpretive work places greater weight on credibility, reflexivity and transferability; critical realism evaluates explanatory adequacy and the plausibility of mechanisms; pragmatism asks whether integrated evidence resolves the practical problem without sacrificing warranted inference. This comparison creates the evaluative frame used in the remainder of the review.

Paradigm	Ontology	Typical evidence	Strength	Blind spot	DBA relevance
Positivism / post-positivism	Reality treated as external and measurable, while post-positivism accepts fallibility	Statistical/experimental evidence; forecast error; inventory and service KPIs	Strong comparison, reliability, causal/associational testing	Can reduce value to measurable proxies; context and meaning underexplained	Essential for performance evidence, but insufficient alone
Interpretivism	Organisational reality is understood through situated meanings and social interaction	Interviews, observation, narratives of trust, expertise, legitimacy	Reveals sense-making, tacit knowledge, identity and accountability	Weak basis for estimating predictive or operational superiority	Explains why recommendations are accepted, modified or rejected
Critical realism	A real world exists, but observed events arise from deeper mechanisms operating contingently	Multi-level cases; mixed evidence; retrodution; context–mechanism reasoning	Explains why similar technologies produce different outcomes	Mechanisms can be difficult to identify and test conclusively	Strong lens for data quality, authority, culture, slack and uncertainty
Pragmatism	Reality/knowledge approached through inquiry into consequences and problem-solving	Methods selected and integrated according to the question and consequences	Fits applied DBA problems requiring several warranted evidence forms	Can become philosophically thin if reduced to “whatever works” or mixed methods	Defensible overarching orientation when integration is explicit and disciplined

Table 2: Paradigm comparison for AI-enabled forecasting research

Philosophical architecture of AI-enabled forecasting and replenishment research



Figure 2: Philosophical architecture of AI-enabled forecasting and replenishment research. Source: Author-developed synthesis from the reviewed literature.

3.2 From forecast accuracy to replenishment and operational value

The technical case for AI forecasting rests on its capacity to detect nonlinearities and interactions in large data sets, particularly where demand is volatile or influenced by many variables. Systematic reviews confirm rapid growth in AI-based demand planning, but also an evidence structure dominated by modelling and prediction studies (Walter, Ahsan and Rahman, 2025). Philosophically, this matters because dominant dependent variables shape the field's conception of success: if comparative error metrics are primary, the research object becomes predictive performance rather than the decision process forecasting is intended to support.

Accuracy remains important. Lower uncertainty can allow tighter inventory parameters, reduce protective stock and improve service. The problem is assuming a stable, direct mapping from statistical accuracy to downstream value. Goltsos et al. (2022) show that forecasting and inventory-control research have often developed separately. Theodorou, Spiliotis and Assimakopoulos (2025), using M5 competition data and inventory simulations, show that accuracy utility changes with policy parameters and economic asymmetries; when lost-sales costs dominate holding costs, the most accurate method need not deliver the best inventory

performance. Review periods, lead times and intermittent demand further alter the relationship. The finding refines rather than rejects accuracy: its value is conditional.

The distinction is especially important in retail because forecasting errors are asymmetric in consequence. Under-forecasting a high-margin or substitution-sensitive item can produce lost sales and customer defection, whereas over-forecasting a perishable item can create waste. Two forecasting methods with similar average accuracy can therefore create very different managerial consequences if their error distributions occur at different points in the assortment. A decision-oriented evaluation must ask not only how large the errors are, but where they occur and how the replenishment policy responds to them.

This exposes a construct-validity issue. Error metrics measure correspondence between prediction and realised demand, while replenishment performance concerns consequences such as availability, holding cost, ordering cost and lost sales. Treating the first as a proxy for the second collapses related but non-identical constructs. The error metric itself embodies values because operational settings often price over- and under-forecasting asymmetrically. A model optimised to minimise an abstract loss function can therefore be 'accurate' under one convention while misaligned with the organisation's service or economic objective.

Operating context creates additional contingencies. Lead time determines exposure to forecast error; review frequency determines when a decision can be corrected; intermittent or promotional demand changes the value of history. Data quality is equally consequential. Missing sales, censored demand, inconsistent product hierarchies or delayed updates alter the empirical reality available to the model. AI can process more data, but volume does not guarantee that those data represent the phenomenon management intends to know. Post-positivist caution is therefore preferable to assuming objective machine knowledge: measurements are fallible and conditioned by data-generating processes.

This point also challenges the claim that AI is intrinsically objective. Feature selection, training windows, treatment of promotions, service-level targets and loss functions are human choices. Historical data can encode prior stock-outs or allocation practices, so a model may learn patterns produced partly by earlier decisions. The epistemic status of an AI forecast is consequently better understood as a model-based estimate generated under assumptions than as an unmediated reading of demand.

Benchmark and simulation studies provide reproducibility and controlled comparison, but their weakness is ecological. A simulated decision does not experience organisational politics, limited attention, supplier unreliability or accountability for exceptions. Field studies capture those

realities while sacrificing some control. Revilla et al. (2023) is important because its retail field experiment connects forecasting performance to automation, adjustable automation and augmentation across approximately 1,888 SKUs. Human intervention adds value under some combinations of uncertainty and horizon but not others, challenging both blanket automation and blanket augmentation.

The comparison between experimental and field evidence also illustrates a validity trade-off. Controlled studies can isolate a treatment such as explanation format or decision authority and support stronger internal inference, while field studies carry noise but reveal whether the effect survives live incentives, workload and process constraints. For this topic, ecological validity is not a secondary concern because the object being explained is organisational value realisation. A result that exists only when the replenishment process, supplier constraints and accountability structure are removed has limited managerial reach.

The literature also cautions against technological determinism. Toorajipour et al. (2021) map substantial AI opportunities across supply-chain functions, but the diversity of applications makes universal superiority claims inappropriate. Walter, Ahsan and Rahman (2025) likewise show that organisational evidence is still developing. Complex models are valuable when signal, data volume, maintenance and decision use justify them; simpler methods may remain competitive in stable or sparse settings, while complex systems can create governance and explainability costs absent from accuracy tables.

Three knowledge claims should therefore remain distinct. Model comparison asks which specification predicts held-out observations most accurately. Inventory evaluation asks which forecast supports the policy that best achieves operational objectives. Management research asks whether an organisation can enact and sustain that policy. The first requires predictive evidence, the second explicit decision and cost assumptions, and the third organisational evidence. Conflating them can make causal claims stronger than the design permits.

The same caution applies to survey-based adoption research. Perceptions that AI improves forecasting or replenishment are evidence about beliefs, not direct measurements of forecast error, stock-outs or inventory cost. Such perceptions are important because they influence use, yet using them as substitutes for operational outcomes creates a different construct-validity problem. A distinction-level research design must preserve the difference between perceived efficacy and realised efficacy.

For DBA research, forecast-error metrics should consequently be analysed alongside inventory, service and cost outcomes, with policy parameters connecting them made explicit. Field

evidence is especially valuable because it tests whether theoretical gains survive organisational use. Where causal identification is unavailable, conclusions should remain proportionate. Philosophically, 'successful AI' is not self-evident: it is defined through choices about measures, horizons, objectives and stakeholders. Once forecasting is treated as one stage in a replenishment system, organisational conditions and human judgement become central explanatory components rather than nuisance variables.

Study	Focus	Design	Critical finding	Role in this review
Goltsos et al. (2022)	Forecasting–inventory integration	Structured review	Forecasting and inventory literatures are highly segregated	Establishes the accuracy-to-decision translation problem
Walter, Ahsan and Rahman (2025)	AI in supply-chain demand planning	Systematic review; 141 studies	Field remains early and modelling-dominated	Supports technical–organisational evidence gap
Theodorou, Spiliotis and Assimakopoulos (2025)	Forecast accuracy and inventory performance	M5 data + inventory simulations	Accuracy ranking can diverge from inventory-cost ranking	Challenges accuracy as proxy for value
Revilla et al. (2023)	Human intervention in retail forecasts	Field experiment; ~1,888 SKUs	Value of intervention depends on uncertainty and horizon	Supports contingent automation/augmentation
Wang et al. (2024)	Algorithm aversion in safety stock	Case data + simulation + interviews/theory	Algorithmic decisions outperform human judgement during disruption; perceived fit/slack explain aversion	Focal study linking performance, behaviour and context
Fügener et al. (2022)	Human–AI delegation	Experimental/analytical research	Poor delegation can reflect weak metaknowledge, not only aversion	Provides alternative human–AI mechanism
Vössing et al. (2022)	Transparency and trust	Design science + organisational cases	Reasoning transparency and uncertainty information affect trust differently	Shows explainability is not a monotonic intervention
Cadden et al. (2022)	Culture and AI supply-chain integration	Quantitative organisational study	Cultural enablers influence AI integration	Moves analysis beyond technical capability
Cao et al. (2021)	Managerial AI acceptance/avoidance	Survey of 269 UK managers	Benefits, concerns and facilitating conditions shape intention	Highlights human-centred adoption conditions
Li et al. (2024)	AI use, coordination and performance	Survey of 236 Chinese firms	Supplier/buyer coordination mediates performance relationship	Supports organisational mechanism and integration logic
Kumar and Kumar (2025)	Human–AI collaboration in O&SCM	Systematic review; 206 studies	HAIC spans pre-adoption, adoption and post-adoption performance	Confirms process view of collaboration
Kumar, Kumar and Raj (2026)	Managerial beliefs and human–AI collaboration	fsQCA; 223 supply-chain managers	Multiple efficacy/legitimacy configurations lead to collaboration orientations	Rejects one universal adoption pathway

Table 3: Critical synthesis of key contemporary studies

3.3 Human–AI decision integration, organisational context and ethics

Human–AI integration is often framed as automation versus augmentation, but recent evidence points to conditional allocation of authority. Revilla et al. (2023) show that human intervention is useful under particular combinations of uncertainty and horizon, not across all forecasting tasks. The relevant question is therefore what the human knows, what the algorithm knows and how the decision system allocates each capability.

Algorithm-aversion research illustrates the ambiguity. Dietvorst, Simmons and Massey (2015) found that observing algorithmic errors can reduce use, while later work showed that limited user control can increase acceptance (Dietvorst, Simmons and Massey, 2018). Logg, Minson and Moore (2019) found algorithm appreciation in several judgement tasks, showing that reluctance is not universal. Fügener et al. (2022) add that weak delegation can reflect poor metaknowledge about relative capability rather than dislike of algorithms. Wang et al. (2024), analysed in Section 4, finds aversion in safety-stock decisions during disruption and proposes reduced perceived fitness and inadequate slack resources as mechanisms. Non-use can therefore reflect control preferences, model fit, expertise or structural constraints, not merely irrationality.

This is consistent with Alon-Barkat and Busuioc's (2023) experimental finding outside retail that users do not display a uniform automation bias but can adhere selectively to algorithmic advice. Although public-sector decisions differ from replenishment, the transferable lesson is methodological: binary assumptions about aversion versus overreliance can obscure the conditions under which reliance changes. The relevant construct is calibrated use rather than maximum use.

Trust is better conceptualised as calibrated reliance than as an outcome to maximise. Glikson and Woolley (2020) identify reliability and transparency as important antecedents, but more trust is undesirable when a model is outside its domain. Vössing et al. (2022) show that reasoning information can increase trust while uncertainty information reduces it. Both responses may improve calibration. Explainability therefore has an epistemic function: it should help users judge when a recommendation deserves reliance rather than simply make them more comfortable with AI.

Organisational evidence further challenges technology-centred accounts. Cao et al. (2021), using 269 UK managers, show that intentions to use AI reflect expected benefits, facilitating conditions and concerns. Cadden et al. (2022) identify cultural enablers of supply-chain AI integration. Kumar, Kumar and Raj (2026), using fsQCA with 223 supply-chain managers, identify multiple efficacy- and legitimacy-related belief configurations rather than one universal adoption pathway. Li et al. (2024) find that supplier and buyer coordination mediate the relationship between generative-AI use and supply-chain performance, while Kumar and Kumar's (2025) review frames human–AI collaboration as a process extending from pre-adoption to post-adoption outcomes. Collectively, these studies support a socio-technical interpretation in which technology, people, tasks and organisational structures are interdependent.

They also reveal a level-of-analysis problem. Individual trust may be high while organisational integration remains weak because data are fragmented, approval rights are unclear or departments optimise conflicting KPIs. Conversely, a strong mandate for AI adoption can coexist with low user legitimacy and symbolic compliance. Bankins et al. (2024) argue more broadly that organisational AI research needs multilevel and field-based analysis; this is particularly relevant to supply chains where decisions span individual planners, internal functions and external partners.

An interpretivist reading makes visible meanings that performance studies bracket out. An override may express distrust, but it may also embody tacit knowledge about a promotion, supplier failure or management commitment absent from training data. A critical-realist reading asks which mechanism operates in a particular context. A post-positivist study can then test whether observable patterns are consistent with that mechanism. These perspectives can therefore be complementary across a research programme even though their assumptions about explanation differ.

Axiology becomes unavoidable when decision rights are shared. Efficiency and accuracy matter, but so do service continuity, employee autonomy, accountability and legitimacy. Preserving human control can be rational when it improves auditability or responsibility; repeated discretionary overrides can also be costly if no learning or rationale is captured. Ethics should therefore concern the accountable allocation of authority rather than defending humans or algorithms in the abstract.

Procedural ethics—consent, confidentiality and data protection—remain important, but AI-enabled supply-chain research raises substantive issues: commercially sensitive data, model bias, opaque accountability, potential deskilling and the quality of human oversight. Oversight is meaningful only when users possess the competence, time and authority to challenge recommendations. NIST's AI Risk Management Framework emphasises governance and transparency across the AI lifecycle (Tabassi, 2023), while UNESCO (2021) places human oversight and accountability alongside technical performance. These normative frameworks clarify why responsible AI cannot be inferred from accuracy alone.

Axiology also has a reflexive research implication. Researchers choose the outcome variables that later appear to demonstrate 'success'. If a study measures only cost reduction, it may make employee autonomy or service inequality invisible; if it measures only trust, it may overlook poor operational performance. Ethical reflexivity therefore requires transparency about whose

interests are represented by the dependent variables and which trade-offs are outside the study's frame.

The managerial implication is calibration rather than simple adoption. Organisations need evidence about when overrides add value, feedback loops that let planners learn from adjustments, explanations matched to task and expertise, and governance specifying decision rights. Training should develop metaknowledge rather than merely confidence. Data-quality governance and coordination are equally important because trust cannot compensate for unreliable inputs or a replenishment process unable to act on predictions. The organisation is thus part of the causal system through which algorithmic capability becomes, or fails to become, value.

3.4 Synthesis, contradictions and research gaps

The reviewed evidence rejects a simple progression from more sophisticated AI to more accurate forecasts and then to better business performance. Technical research establishes measurable predictive gains, inventory research shows that their economic value depends on policy and cost conditions, human–AI research shows that output is accepted or modified according to uncertainty, trust and expertise, and organisational research adds culture, coordination, slack and managerial beliefs as mechanisms (Goltsos et al., 2022; Theodorou, Spiliotis and Assimakopoulos, 2025; Revilla et al., 2023; Wang et al., 2024; Cadden et al., 2022).

Six gaps follow. The gaps are related rather than independent: a modelling-dominated evidence base makes it easier to treat accuracy as value, while limited organisational evidence leaves human and contextual mechanisms underspecified. First, an accuracy–utility gap persists where forecast error is treated as a proxy for replenishment value. Second, a technical–organisational evidence gap remains because demand-planning research is still modelling-heavy (Walter, Ahsan and Rahman, 2025). Third, a human–AI mechanism gap spans conflicting aversion, appreciation, transparency and delegation findings. Fourth, organisational mechanisms such as data quality, authority, culture, slack and coordination remain unevenly integrated. Fifth, a philosophical/axiological gap persists because studies rarely interrogate whose definition of 'successful AI' is embedded in their outcome measures. Sixth, ecological and longitudinal evidence is limited, leaving insufficient understanding of how trust, capability and governance co-evolve (Bankins et al., 2024).

Table 4 converts these gaps into implications for the DBA. Its purpose is not to manufacture novelty from missing topics, but to show where philosophical assumptions have produced partial

explanations. The most significant opportunity is to connect forms of evidence that are usually studied separately while retaining clear boundaries between what each can support.

No single paradigm fully captures these gaps. Post-positivist evidence is required to avoid replacing measurement with opinion; interpretive inquiry is required to understand meaning and legitimacy; critical realism supports mechanism and context; and pragmatism can organise these forms of evidence around a consequential DBA problem. The defensible position is not an unqualified hybrid, but a problem-centred orientation in which each knowledge claim is explicit about its assumptions, quality criteria and limits.

Gap	Evidence / contradiction	Philosophical significance	DBA implication
Accuracy–utility gap	Accuracy and inventory rankings diverge under different costs/policies	Predictive truth is not identical to decision value	Measure forecast and operational outcomes together
Technical–organisational evidence gap	Demand-planning literature remains modelling-heavy	External/ecological validity is underdeveloped	Increase longitudinal and field evidence
Human–AI mechanism gap	Aversion, appreciation, transparency and delegation produce conflicting patterns	Meaning and causal mechanism cannot be inferred from overrides alone	Combine behavioural logs with explanatory inquiry
Organisational mechanism/context gap	Culture, coordination, slack, authority and data quality condition integration	Causal powers are context-dependent	Use mechanism-sensitive, multi-level designs
Philosophical/axiological gap	Studies operationalise “success” differently	Dependent-variable choice encodes values	Make stakeholder values and success criteria explicit
Longitudinal/governance gap	Many studies are short-term, cross-sectional or simulated	Trust, capability and accountability may co-evolve	Track implementation, learning and responsible governance over time

Table 4: Research gaps, philosophical significance and DBA implications

4. Philosophical and Methodological Analysis of Wang et al. (2024)

4.1 Study rationale, problem and contribution

Wang et al. (2024) are particularly relevant because they examine AI at the point where analytics becomes a material replenishment decision. Using safety stock during supply-chain disruption, the study asks whether algorithm aversion occurs, how it affects performance and why users deviate from analytical decision support. It combines operational data from a European online grocery retailer with simulation and interview-informed theorising. Pre-disruption analysis covers 17 fresh-product SKUs and post-disruption analysis 28 grocery SKUs. The authors report that during disruption algorithmic decisions outperform human judgement and propose reduced perceived algorithm fitness and insufficient slack resources for users and developers as mechanisms. The design therefore sits directly within the report's central question: why technically stronger support may fail to become realised organisational value.

The data are particularly useful because the setting is not a hypothetical vignette. Purchasing managers worked with daily forecasts and could adjust safety stock or ordering decisions; lead times varied across products and the disruption altered the environment in which historical routines had previously operated. This makes the paper a strong case for examining how claims about algorithm superiority are constructed when the decision context itself is unstable.

4.2 Ontological assumptions

The paper does not explicitly declare an ontology, so any classification is interpretive. Operational records, safety-stock decisions and performance outcomes are treated as phenomena that can be measured independently of participants' descriptions, exhibiting realist/post-positivist characteristics. Yet the explanation also depends on perceived fitness and organisational slack, meaning that identical technical advice can acquire different practical significance under different conditions. A critical-realist reading is therefore plausible, but it would be inaccurate to attribute critical realism to the authors. The design is better understood as ontologically layered: material outcomes are treated as real, while use is mediated through social and organisational mechanisms.

That layered reading is preferable to calling the study simply 'positivist'. The interviews are not used merely to measure an attitude; they help theorise conditions that may have causal force. At the same time, the paper does not develop a formal stratified ontology or retroduction process. Critical-realist affinities should therefore be treated as an analytical interpretation of its mechanism language, not as evidence of an author-declared paradigm.

This boundary supports strong comparative claims within the case but gives less attention to professional identity, power and wider legitimacy. The ontology privileges a decision-performance reality while only partly opening the socially constructed reality through which recommendations are interpreted.

4.3 Epistemology

The study's epistemology is also unstated, but the design exhibits post-positivist characteristics. Operational data and simulation provide the principal evidence for whether algorithmic rules would have outperformed observed human decisions. This is a major strength because it avoids asking managers to self-report whether their judgement was effective and permits counterfactual comparison. Interviews and information-systems theory then explain why deviations occurred.

The combination is complementary, but it creates an implicit hierarchy: quantitative evidence establishes the performance problem; qualitative evidence explains behaviour. The paper

provides more detail about operational data and simulation than about qualitative analysis, making dependability and confirmability of the mechanism claims harder to audit. This matters because knowing that an algorithm performs better is epistemologically different from knowing why managers reject or adapt it. A more symmetrical design would test whether users held relevant contextual information or accountabilities absent from the simulation rather than assuming deviation represents a cognitive failure.

This is also a question of inference. The simulation can establish that specified algorithmic rules would have performed differently under the modelled conditions, but it does not directly observe the unchosen future. Counterfactual knowledge is conditional on policy assumptions and the quality of recorded data. The epistemic strength of the result therefore lies in transparent comparative modelling rather than absolute proof of what 'would' have happened in every operational detail.

4.4 Axiology and researcher positionality

The study privileges decision performance and successful use of analytically superior safety-stock rules. This is appropriate to operations management, but it is also axiological: safety stock embodies trade-offs among availability, inventory cost and resilience. Labelling deviation 'algorithm aversion' can imply that adherence is preferable when the algorithm wins under the selected performance criteria.

That framing risks deficit-framing users if alternative rationalities are underexplored. Managers may respond to supplier information, reputational risk or commitments that are not represented in historical data. Wang et al. (2024) partly avoid this simplification through perceived fitness and slack, but the study could distinguish more explicitly between irrational aversion and contextually defensible non-adherence.

Researcher positionality and reflexivity are not explicitly developed. This does not invalidate the quantitative comparison, but it matters because researchers select performance benchmarks, classify deviations and choose explanatory constructs. Reflexive disclosure would make these interpretive choices visible.

Teleology is equally relevant. If the ultimate purpose is to maximise algorithm adoption, user deviation appears mainly as a barrier. If the purpose is resilient and accountable inventory decision-making, a justified override can be valuable even when adherence is usually superior. A more explicit statement of the desired end state would help separate adoption as a means from organisational performance and responsible control as ends.

4.5 Ethics

Research ethics and AI-practice ethics should be separated. The paper reports that company data were normalised for confidentiality and cannot be publicly shared because of commercial sensitivity (Wang et al., 2024). In the accessible article, a formal ethics-approval statement, consent procedure or detailed interview confidentiality protocol is not prominently reported. It is therefore inappropriate either to assume their absence or claim that they were documented.

Substantively, safety-stock algorithms redistribute decision authority. Responsibility for poor outcomes can become ambiguous among model, developer, planner and organisation. Meaningful human oversight requires competence, time and authority, while unlogged overrides can weaken accountability. Data quality may create uneven service consequences and automation may deskill planners. Responsible use therefore requires traceability, decision rights and governance alongside performance optimisation (Tabassi, 2023; UNESCO, 2021).

4.6 Methodological coherence and design quality

The combination of a real organisational case, simulation and interviews is well matched to the research problem. The disruption setting adds ecological value because it tests decision support when historical regularities are threatened and intuition becomes especially salient. It is stronger than a laboratory judgement task for understanding actual replenishment behaviour.

Constraints remain. The study is based on one online grocery setting and different product sets across pre- and post-disruption periods. Sales are used as a proxy for demand, creating censored-demand risk when stock-outs hide unmet demand; the authors acknowledge this. Simulation also depends on assumptions about replenishment logic, lead times and cost relationships. The study infers aspects of the incumbent historical policy from interview description and data fit. These choices are reasonable but mean the counterfactual is a disciplined representation, not direct observation of an alternative reality.

Method integration is therefore the key coherence test. The simulation identifies a performance discrepancy and the interview-informed theory explains why that discrepancy may arise. The design is stronger when the strands genuinely constrain each other—for example, when qualitative explanation accounts for patterns visible in the operational evidence—than if interviews merely provide illustrative quotations after a quantitative conclusion. The paper moves in the stronger direction, although fuller qualitative reporting would make the integration easier to evaluate.

4.7 Reliability, validity and trustworthiness

Operational records, explicit simulation logic and repeatable policy comparison strengthen quantitative reliability. Exact replication is limited by the non-public dataset, a defensible confidentiality trade-off that reduces independent verification. Construct validity is more contested. 'Algorithm aversion' is evidenced through rejection or adjustment combined with poorer performance, yet deviation may also represent rational adaptation, competing objectives or model misfit. Interviews help discriminate among explanations but do not eliminate construct ambiguity.

Internal validity is stronger than in a purely observational case because simulation creates counterfactual comparisons, but the design is not randomised. Disruption coincides with changes in demand, supply and managerial attention, so causal claims about why users behaved as they did should remain bounded. External validity is limited by the single retailer and exceptional COVID-era context, whereas ecological validity is a clear strength because the behaviour occurred within a real replenishment system. The study therefore supports analytical rather than statistical generalisation.

Bias also requires a nuanced assessment. Selection into one retailer limits representativeness, while disruption-specific managerial decisions may be affected by unobserved pressures. Conversely, using actual behavioural records reduces common-method bias compared with asking the same respondents to report both AI use and performance. The design consequently trades population representativeness for behavioural authenticity, a defensible choice for theory-building when its boundaries are acknowledged.

Qualitative trustworthiness is strengthened by linking interview explanation to observed behaviour, but comparatively limited reporting of interview sampling and coding makes credibility and dependability difficult to assess to the same standard as the simulation. Clearer sampling, negative-case analysis and reflexivity would strengthen mechanism claims.

4.8 Theoretical framing and alignment

Wang et al. (2024) use information-systems reasoning to explain how disruption reduces perceived algorithm fitness and how insufficient slack limits adaptation. This relocates aversion from an isolated cognitive bias into an organisational setting and aligns with evidence on culture, facilitating conditions and coordination (Cadden et al., 2022; Cao et al., 2021; Li et al., 2024).

Socio-Technical Systems thinking could extend this explanation by treating algorithm, users, task, routines and governance as jointly interdependent. Critical realism could organise the

account as context–mechanism–outcome reasoning: disruption and limited slack create conditions in which perceived misfit activates non-adherence, affecting inventory outcomes. These perspectives clarify how objective performance evidence and socially situated explanation can coexist without being treated as the same kind of knowledge.

The theoretical alignment is therefore promising but incomplete. The paper's strongest contribution is precisely its move from an individual-bias label towards organisational conditions. A fuller socio-technical model would ask how task redesign, interfaces, incentives, cross-functional routines and governance change the mechanisms over time, rather than treating slack and fitness as relatively static explanatory factors.

4.9 Alternative philosophical approaches

An interpretivist redesign would ask how planners construct trust, expertise and accountability around AI during disruption. Longitudinal interviews and observation could reveal tacit knowledge and legitimacy, gaining depth while sacrificing direct estimation of comparative performance.

A critical-realist design would retain operational outcomes but compare cases to identify mechanisms and boundary conditions such as data quality, disruption severity, authority, slack and expertise. Retroduction would test competing explanations of non-adherence rather than treating aversion as the endpoint.

A pragmatist design would explicitly integrate longitudinal forecast/inventory metrics, system logs of overrides and qualitative accounts of exceptions. Integration would occur around the consequential question—when and how the intervention improves replenishment—rather than by simply combining methods. This is particularly suitable for an applied DBA when justified through disciplined inquiry rather than methodological convenience (Morgan, 2014).

A stricter post-positivist replication could add multi-site field experimentation and pre-specified outcomes, strengthening causal inference about interface or override design while still requiring complementary evidence to explain meaning and legitimacy.

Table 6 makes the consequences of these alternatives explicit. The point is not that one philosophy would produce a universally better study. Each changes the research question, quality criteria and legitimate conclusion. Interpretivism would deepen meaning, critical realism mechanism, post-positivism causal comparison and pragmatism problem-oriented integration. The future DBA can therefore justify methodological plurality only after specifying which knowledge deficit each strand addresses.

4.10 Overall evaluation

Wang et al. (2024) make a strong contribution by linking real deviation from algorithmic safety-stock recommendations to measurable performance and an organisational explanation. Its ecological credibility exceeds many laboratory studies, while its behavioural insight exceeds a technical simulation alone. Limitations include philosophical implicitness, single-case generalisability, simulation assumptions, restricted data replication, thinner qualitative reporting and limited visible procedural-ethics detail. These limitations make the paper especially useful for RPP: they show that performance evidence and contextual explanation are both necessary but neither is sufficient. The study therefore supports the wider framing of AI-enabled replenishment as a socio-technical value-realisation problem and provides a bridge to a future DBA design integrating rigorous outcomes with mechanism-sensitive explanation.

Dimension	Evidence / status	Critical interpretation	DBA implication
Ontology	No explicit ontology; operational outcomes treated as real/measurable, while perceived fit/slack are contextual	Layered design exhibits post-positivist characteristics with mechanism sensitivity	Future DBA should state ontological assumptions explicitly
Epistemology	Operational data and simulation establish performance; interviews/theory explain behaviour	Complementary evidence, but quantitative evidence is more transparently reported	Integrate forms of evidence without creating an implicit hierarchy
Axiology	Performance and analytically superior decisions are privileged	"Aversion" may deficit-frame users if alternative responsibilities/values are hidden	Define value across service, cost, resilience, autonomy and accountability
Positionality	Researcher reflexivity not explicitly developed	Classification and benchmark choices still embody researcher judgement	Include reflexive account of constructs and interpretation
Reliability	Operational records and explicit simulation improve repeatability; data remain confidential	Strong within-study consistency; exact external replication constrained	Use auditable code/logic and transparent data governance
Validity	Counterfactual simulation strengthens comparison; single retailer and disruption constrain generalisation	High ecological relevance but limited statistical external validity	Use multi-site/longitudinal field evidence where feasible
Qualitative trustworthiness	Interview-informed mechanisms provide context; reporting is thinner than quantitative methods	Credibility is plausible but hard to audit fully	Report sampling, coding, negative cases and reflexivity
Ethics	Commercial confidentiality reported; formal ethics/consent details not prominent in accessible article	Procedural reporting is limited; substantive decision-accountability issues remain	Separate research ethics from responsible AI practice

Table 5: Philosophical and methodological audit of Wang et al. (2024)

Approach	Reframed question	Likely design	Knowledge gained / trade-off
Interpretivist	How do planners construct trust, expertise and accountability around AI during disruption?	Longitudinal interviews, observation, planning narratives	Rich understanding of meaning/legitimacy; weaker performance generalisation
Critical realist	Which mechanisms generate adherence or non-adherence under which contexts?	Comparative cases + operational evidence + retrodution	Explains context–mechanism–outcome configurations
Pragmatist	What combination of evidence best determines whether and how AI improves replenishment?	Integrated operational metrics, behavioural logs and qualitative inquiry	Problem-centred, managerially actionable knowledge when integration is explicit
Stricter post-positivist	What is the causal effect of decision-support design or override rights?	Multi-site field experiment with pre-specified outcomes	Stronger causal inference, but meaning/context still partly bracketed

Table 6: Alternative philosophical redesigns of the focal problem

5. Discussion and Contribution

The review indicates that the central managerial problem is not whether AI can produce accurate demand forecasts, but whether predictive capability survives the sequence of decisions through which an organisation converts a forecast into inventory action. Theodorou, Spiliotis and Assimakopoulos (2025) show that accuracy and inventory performance can diverge under different cost and policy conditions, while Goltsoos et al. (2022) identify the structural separation between forecasting and inventory research. Human–AI studies add another source of contingency: intervention can improve or damage performance depending on task uncertainty, horizon, expertise and the allocation of control (Revilla et al., 2023; Fügener et al., 2022). Wang et al. (2024) makes this translation problem concrete during disruption, where analytically stronger safety-stock decisions did not automatically become realised organisational decisions. Importantly, interpreting the behaviour only as irrational aversion would be too narrow; perceived fitness, limited slack, accountability and context can make non-adherence an organisational phenomenon rather than a simple cognitive defect.

Philosophically, the synthesis demonstrates that different traditions answer different parts of the problem. Post-positivist evidence is indispensable for testing forecast accuracy, policy outcomes and performance relationships; removing it would make claims of AI value impossible to evaluate. Its limitation arises when measurable outcomes are treated as the whole phenomenon. Interpretivism reveals how trust, legitimacy, expertise and accountability shape the meaning of a recommendation. Critical realism adds the question of mechanism: why does the same technical capability produce different outcomes across contexts? Pragmatism offers a defensible applied orientation when the inquiry is organised around the consequences of a managerial problem and integrates evidence deliberately. Following Morgan (2014), this does not mean that any mixed-method design is automatically pragmatic. Coherence depends on being explicit about what each form of evidence can establish and how their claims are brought together.

The theoretical contribution is therefore a socio-technical reframing of AI forecasting. Figure 2 positions predictive performance as only one stage in a chain extending from AI capability through decision support and human–AI integration to replenishment, operational outcomes and organisational value. Data quality, culture, skills, governance, system integration, authority and uncertainty condition the chain. Socio-Technical Systems thinking captures the joint optimisation problem, while dynamic-capability and supply-chain-integration perspectives help explain how organisations reconfigure resources and coordinate information. This architecture develops the earlier accuracy–performance logic into a mechanism-sensitive account of value realisation rather than assuming that technology produces outcomes directly.

Methodologically, the future DBA should therefore avoid both forecast benchmarking in isolation and perception-only inquiry. A defensible design would combine longitudinal operational evidence—forecast errors, stock-outs, service, inventory and cost—with behavioural evidence about overrides and system use, and qualitative evidence about trust, perceived fit, accountability and organisational routines. Figure 3 represents this evidence architecture. Such a design would permit triangulation without pretending that all data answer the same question: quantitative evidence evaluates performance, behavioural data reveal enactment, and qualitative inquiry explains meaning and mechanism. Explicit integration, careful construct definition, attention to causality and reflexivity would be more important than simply increasing the number of methods.

For managers, the implication is calibrated collaboration rather than either 'trust the algorithm' or 'keep a human in the loop' as universal rules. Forecast accuracy should be monitored alongside operational KPIs; override rationales should be logged and fed back into learning; explanations should be matched to the task and user expertise; and decision rights should specify when discretion is appropriate. Organisations also require data-quality governance, cross-functional coordination and enough slack to adapt systems when disruption changes model fit. These practices address both performance and responsible AI because accountability, transparency, employee autonomy and customer availability are values alongside cost and efficiency.

The report's contribution is conceptual and methodological rather than a claim of new empirical discovery. It reframes AI-enabled forecasting as a philosophically conditioned socio-technical value-realisation problem, specifies why the accuracy–utility relationship is contingent, integrates human and organisational mechanisms into the explanatory chain, and shows how ontology, epistemology and axiology shape definitions of valid evidence and successful AI. This provides a disciplined basis for a future DBA inquiry that asks not only whether AI works, but

how organisations can know when it works, why it works in one context rather than another, and whether the resulting decision system is managerially useful and responsibly governed.

This contribution also refines the researcher's own positioning. A practice-oriented study can begin pragmatically while remaining alert to critical-realist questions about mechanism and interpretive questions about meaning; what must be avoided is presenting this as a loose philosophical blend. The coherence test is whether each method answers a distinct aspect of the research problem and whether the final inference integrates those answers transparently.

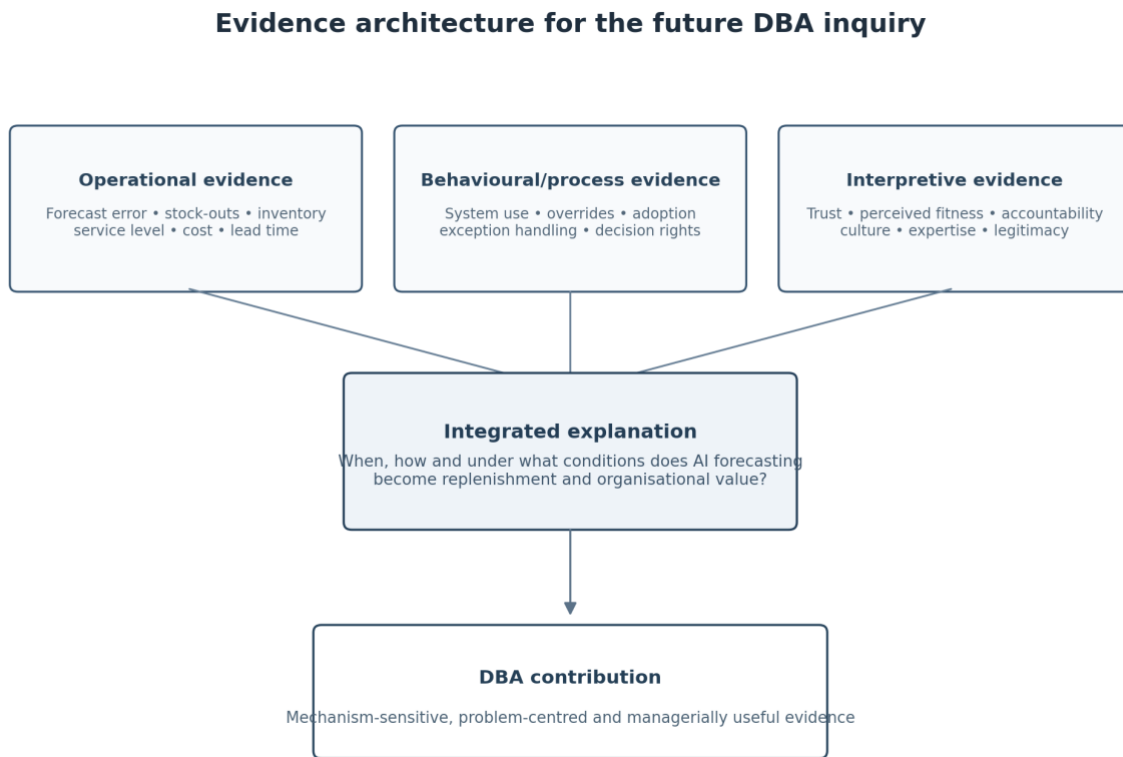


Figure 3: Evidence architecture for the future DBA inquiry. Source: Author-developed synthesis.

6. Conclusion

This report has argued that AI-enabled forecasting in retail supply chains cannot be understood adequately as a prediction-accuracy problem alone. Forecast accuracy remains an essential technical criterion, but its translation into replenishment performance and organisational value is contingent on inventory policy, cost asymmetries, demand and lead-time conditions, data quality, system integration, organisational capability, decision authority and the way human decision-makers engage with algorithmic advice. The literature therefore supports a distinction

between predictive performance and decision value rather than treating one as an automatic proxy for the other.

The philosophical analysis explains why this distinction matters. Positivist and post-positivist approaches provide indispensable evidence about measurable performance, yet they can under-explain the meanings, organisational structures and mechanisms through which technology is enacted. Interpretivism contributes insight into trust, expertise, legitimacy and accountability; critical realism directs attention to generative mechanisms and contextual conditions; and pragmatism can provide a defensible applied DBA orientation when it integrates different forms of warranted evidence around a consequential problem rather than merely equating pragmatism with mixed methods.

The analysis of Wang et al. (2024) reinforces this conclusion. Their study shows that analytically stronger safety-stock recommendations can fail to become realised decision value during disruption when users resist or adapt algorithmic guidance. However, behaviour labelled algorithm aversion should not automatically be interpreted as irrational failure: perceived model fitness, slack resources, contextual knowledge and accountability can shape non-adherence. The focal study consequently illustrates both the value of performance measurement and the need for mechanism-sensitive organisational explanation.

Conceptually, the report reframes AI forecasting as a socio-technical value-realisation chain: technical capability informs prediction, prediction enters decision support, human–AI integration shapes replenishment action, and organisational conditions determine resulting operational value. Axiology and ethics run through the entire chain because 'better' can mean accuracy, cost, service, resilience, autonomy, accountability or responsible oversight. Future DBA research should therefore align ontology, epistemology, axiology and methodology explicitly while evaluating technical outcomes, organisational mechanisms and stakeholder consequences together. The central question is not merely whether AI can forecast more accurately, but under what human, organisational and philosophical conditions predictive capability becomes responsible and sustainable decision value.

References

- Alon-Barkat, S. and Busuioc, M. (2023) 'Human–AI interactions in public sector decision making: “Automation bias” and “selective adherence” to algorithmic advice', *Journal of Public Administration Research and Theory*, 33(1), pp. 153–169. doi:10.1093/jopart/muac007.
- Alvesson, M. and Sandberg, J. (2011) 'Generating research questions through problematization', *Academy of Management Review*, 36(2), pp. 247–271. doi:10.5465/amr.2009.0188.
- Bankins, S., Ocampo, A.C., Marrone, M., Restubog, S.L.D. and Woo, S.E. (2024) 'A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice', *Journal of Organizational Behavior*, 45(2), pp. 159–182. doi:10.1002/job.2735.
- Cadden, T., Dennehy, D., Mäntymäki, M. and Treacy, R. (2022) 'Understanding the influential and mediating role of cultural enablers of AI integration to supply chain', *International Journal of Production Research*, 60(14), pp. 4592–4620. doi:10.1080/00207543.2021.1946614.
- Cao, G., Duan, Y., Edwards, J.S. and Dwivedi, Y.K. (2021) 'Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making', *Technovation*, 106, 102312. doi:10.1016/j.technovation.2021.102312.
- Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015) 'Algorithm aversion: People erroneously avoid algorithms after seeing them err', *Journal of Experimental Psychology: General*, 144(1), pp. 114–126. doi:10.1037/xge0000033.
- Dietvorst, B.J., Simmons, J.P. and Massey, C. (2018) 'Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them', *Management Science*, 64(3), pp. 1155–1170. doi:10.1287/mnsc.2016.2643.
- Fügener, A., Grahl, J., Gupta, A. and Ketter, W. (2022) 'Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation', *Information Systems Research*, 33(2), pp. 678–696. doi:10.1287/isre.2021.1079.
- Frederiksen, D.J. and Kringelum, L.B. (2021) 'Five potentials of critical realism in management and organization studies', *Journal of Critical Realism*, 20(1), pp. 18–38. doi:10.1080/14767430.2020.1846153.
- Glikson, E. and Woolley, A.W. (2020) 'Human trust in artificial intelligence: Review of empirical research', *Academy of Management Annals*, 14(2), pp. 627–660. doi:10.5465/annals.2018.0057.

- Goltsos, T.E., Syntetos, A.A., Glock, C.H. and Ioannou, G. (2022) 'Inventory–forecasting: Mind the gap', *European Journal of Operational Research*, 299(2), pp. 397–419. doi:10.1016/j.ejor.2021.07.040.
- Kumar, N. and Kumar, R.R. (2025) 'Human–AI collaboration in operations and supply chain management: A systematic literature review', *Management Review Quarterly*. doi:10.1007/s11301-025-00575-9.
- Kumar, N., Kumar, R.R. and Raj, A. (2026) 'Managerial beliefs and human-artificial intelligence collaboration in supply chain: A configurational perspective', *International Journal of Production Economics*, 297, 110005. doi:10.1016/j.ijpe.2026.110005.
- Li, L., Liu, Y., Jin, Y., Cheng, T.C.E. and Zhang, Q. (2024) 'Generative AI-enabled supply chain management: The critical role of coordination and dynamism', *International Journal of Production Economics*, 277, 109388. doi:10.1016/j.ijpe.2024.109388.
- Logg, J.M., Minson, J.A. and Moore, D.A. (2019) 'Algorithm appreciation: People prefer algorithmic to human judgment', *Organizational Behavior and Human Decision Processes*, 151, pp. 90–103. doi:10.1016/j.obhdp.2018.12.005.
- Makarius, E.E., Mukherjee, D., Fox, J.D. and Fox, A.K. (2020) 'Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization', *Journal of Business Research*, 120, pp. 262–273. doi:10.1016/j.jbusres.2020.07.045.
- Mingers, J. (2000) 'What is it to be critical? Teaching a critical approach to management undergraduates', *Management Learning*, 31(2), pp. 219–237. doi:10.1177/1350507600312005.
- Morgan, D.L. (2014) 'Pragmatism as a paradigm for social research', *Qualitative Inquiry*, 20(8), pp. 1045–1053. doi:10.1177/1077800413513733.
- Page, M.J. et al. (2021) 'The PRISMA 2020 statement: An updated guideline for reporting systematic reviews', *BMJ*, 372, n71. doi:10.1136/bmj.n71.
- Revilla, E., Saenz, M.J., Seifert, M. and Ma, Y. (2023) 'Human–artificial intelligence collaboration in prediction: A field experiment in the retail industry', *Journal of Management Information Systems*, 40(4), pp. 1071–1098. doi:10.1080/07421222.2023.2267317.
- Saunders, M., Lewis, P. and Thornhill, A. (2019) *Research Methods for Business Students*. 8th edn. Harlow: Pearson.

Tabassi, E. (2023) *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology. doi:10.6028/NIST.AI.100-1.

Theodorou, E., Spiliotis, E. and Assimakopoulos, V. (2025) 'Forecast accuracy and inventory performance: Insights on their relationship from the M5 competition data', *European Journal of Operational Research*, 322(2), pp. 414–426. doi:10.1016/j.ejor.2024.12.033.

Toorajipour, R., Sohrabpour, V., Nazarpour, A., Oghazi, P. and Fischl, M. (2021) 'Artificial intelligence in supply chain management: A systematic literature review', *Journal of Business Research*, 122, pp. 502–517. doi:10.1016/j.jbusres.2020.09.009.

UNESCO (2021) *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO.

Vössing, M., Kühl, N., Lind, M. and Satzger, G. (2022) 'Designing transparency for effective human-AI collaboration', *Information Systems Frontiers*, 24, pp. 877–895. doi:10.1007/s10796-022-10284-3.

Walter, A., Ahsan, K. and Rahman, S. (2025) 'Application of artificial intelligence in demand planning for supply chains: A systematic literature review', *The International Journal of Logistics Management*, 36(3), pp. 672–719. doi:10.1108/IJLM-02-2024-0120.

Wang, X., Sanchez Rodrigues, V., Demir, E. and Sarkis, J. (2024) 'Algorithm aversion during disruptions: The case of safety stock', *International Journal of Production Economics*, 278, 109442. doi:10.1016/j.ijpe.2024.109442.

Appendices

Appendix A. Search concept families

Search family	Concept string
Family A – AI forecasting/performance	("artificial intelligence" OR "machine learning" OR "predictive analytics") AND ("demand forecast*" OR "demand planning" OR "forecast accuracy") AND (retail OR "supply chain") AND (replenish* OR inventory OR "service level" OR stockout*)
Family B – human-AI integration	("human-AI" OR "human-in-the-loop" OR "algorithm aversion" OR "algorithm appreciation" OR "judgmental adjustment*" OR explainab* OR trust) AND (forecast* OR replenish* OR inventory OR "supply chain decision**")
Family C – organisational context	("artificial intelligence" OR AI) AND ("supply chain" OR retail) AND (governance OR culture OR readiness OR capability OR integration OR adoption OR "decision authority" OR accountability OR ethics)
Family D – research philosophy	(positivis* OR post-positivis* OR interpretiv* OR pragmat* OR "critical realism" OR ontolog* OR epistemolog* OR axiolog* OR reflexiv*) AND ("artificial intelligence" OR algorithm*) AND (management OR organi?ation* OR "supply chain")

Appendix B. Extended evidence-extraction fields

- Full Harvard citation
- Sector / country / organisational context
- Research purpose and questions
- Theoretical framework
- Research design and sample/data
- AI/forecasting construct
- Forecast-performance measure
- Replenishment / operational outcome
- Human–AI factor
- Main finding
- Stated ontology/epistemology (if any)
- Carefully inferred philosophical characteristics
- Axiological assumptions / valued outcomes
- Research ethics / AI-practice ethics
- Reliability / dependability
- Construct, internal, external and ecological validity
- Limitations and alternative explanations
- Contradictions with other evidence
- Research gap
- DBA relevance