ISRG JOURNAL OF HUMANITIES AND CULTURAL STUDIES (ISRGJHCS)

ISRG PUBLISHERS
Abbreviated Key Title: ISRG J Humanit Cult Stud
ISSN: 3048-5436 (Online)
Journal homepage: https://isrgpublishers.com/gjhcs/
Volume – III Issue- V (September-October) 2026
Frequency: Bimonthly



Beyond Forecast Accuracy: A Responsible AI Framework for Provenance-Aware Fine Jewelry Retail Planning

Katayoun Kiarsi[1*], James Hutson[2]

[1, 2] Lindenwood University, Saint Charles, Missouri, United States

| **Received:** 14.10.2026 | **Accepted:** 14.10.2026 | **Published:** 17.10.2026

*__Corresponding author:__ Katayoun Kiarsi

Abstract

Fine jewelry retailers must reconcile uncertain demand, high-value inventory, complex material histories, and expectations of responsible sourcing. This article develops a responsible artificial intelligence framework that connects demand forecasting, provenance evidence, and governance within retail planning. A conceptual framework-development approach informed by design-science research combines a targeted literature update through October 9, 2026, with descriptive analysis of published market statistics and an external retail forecasting benchmark. The analysis distinguishes original design propositions from empirical findings reported by other researchers; no retailer-level experiment or environmental impact assessment was conducted. Three sourced bar charts illustrate changing gold jewelry consumption and the methodological relevance of inventory-based model evaluation. The resulting architecture integrates data inputs, predictive analytics, provenance and traceability, ethical governance, and human-reviewed decisions with feedback. Five design principles and five testable propositions specify how forecasts can inform assortment, replenishment, allocation, and sourcing while preserving evidence quality and accountability. The framework shifts evaluation from forecast accuracy alone toward decision quality under explicit operational and provenance constraints. It also separates inventory efficiency, sourcing documentation, and environmental outcomes, preventing improvements in one category from being misrepresented as proof of another. Managerial implications include staged implementation, evidence-qualified sourcing rules, meaningful review authority, and documented overrides. Future validation should compare complete decision policies, preserve temporal information boundaries, measure calibrated reliance, and assess environmental consequences independently. The article contributes a sector-specific architecture and evaluation agenda for accountable retail intelligence without claiming that the integrated framework has already demonstrated operational or sustainability benefits across fine jewelry businesses and markets.

Keywords: responsible artificial intelligence; fine jewelry retail; demand forecasting; supply chain provenance; sustainable inventory planning

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

1. Introduction

Fine jewelry retail presents a distinctive problem for artificial intelligence because the value of an inventory decision exceeds the value of an accurate forecast. A ring, necklace, or gemstone carries commercial attributes, material histories, and cultural meanings that cannot be reduced to expected sales alone. Retailers must decide which products to acquire, where to allocate them, and when to replenish them while maintaining credible sourcing commitments. Han (2025) characterizes jewelry inventory management through high product values, emotionally influenced purchasing, and seasonal fluctuations. Perret et al. (2025) similarly identify sustainability and diversification as strategic concerns for high jewelry businesses. Taken together, these characteristics suggest that planning systems should connect demand estimation with the conditions under which products can responsibly enter an assortment. A forecast may identify a profitable opportunity without establishing whether the proposed supplier, material, or purchasing quantity is acceptable. This article therefore treats fine jewelry planning as a problem of accountable decision design rather than prediction in isolation.

The sustainability implications of planning deserve attention alongside the better established emphasis on extraction and responsible sourcing. Inventory accumulation can commit working capital and materials to products for which demand is uncertain; repeated replenishment or reallocation can also generate additional handling and transport. These mechanisms are plausible pathways, not evidence that every reduction in inventory produces an environmental benefit. Jewelry may retain material value, remain desirable over long periods, or circulate through repair and resale. Conversely, high residual value does not erase the environmental and social consequences of producing and moving an additional item. The relevant question is whether a particular planning decision changes material use, sourcing exposure, or product circulation relative to a defensible alternative. CIBJO (2025) situates environmental, social, and governance responsibilities within a sector-specific framework that also recognizes the needs of smaller businesses. The present article extends that orientation into the routines through which retailers translate demand expectations into purchasing commitments.

Recent market evidence makes the distinction between financial and physical demand particularly important. The World Gold Council's full-year report shows that gold jewelry consumption fell from 1,866.9 tonnes in 2024 to 1,542.3 tonnes in 2025, while demand value increased by 18% (World Gold Council, 2026a). This divergence cautions against treating revenue growth as equivalent to growth in material throughput. It also raises a practical modeling question: should a retailer forecast transactions, item quantities, metal weight, expenditure, or margin? Each target answers a different operational question, and none substitutes automatically for another. These aggregate statistics concern gold jewelry rather than the entire fine jewelry market. They cannot establish the demand profile of a particular retailer, gemstone category, or customer segment. Nevertheless, they provide a concrete reason to preserve separate volume, value, and provenance measures within the proposed architecture.

Advances in forecasting expand the available technical options without resolving these managerial distinctions. Research has moved beyond comparisons of conventional time-series methods and machine learning toward models that incorporate covariates, share information across related series, and evaluate downstream consequences. Ansari et al. (2025) introduce Chronos-2 as a pretrained model for univariate, multivariate, and covariate-informed forecasting; this article treats that contribution as a preprint rather than established jewelry evidence. Elmachtoub and Grigas (2022) provide a more foundational argument for evaluating predictions through the optimization decisions they induce. These developments make an accuracy-only account of responsible retail planning increasingly inadequate. However, a technically sophisticated forecasting system still needs an explicit representation of provenance, uncertainty, decision rights, and operational constraints. The integration problem is therefore organizational and informational as well as computational. A sector-specific framework should identify how those elements interact without presuming that a particular algorithm supplies the solution.

This article develops a responsible AI framework for provenance-aware fine jewelry retail planning through conceptual synthesis and descriptive use of published evidence. Its first research question asks how demand forecasts should be connected to assortment, sourcing, replenishment, and allocation decisions. Its second asks how provenance information can become an operational constraint or decision attribute rather than a reporting attachment. Its third asks how governance and human review can make resulting recommendations contestable and measurable. The proposed contribution is an integrated decision architecture with five functions, five design principles, and five propositions for subsequent testing. The article does not introduce a validated software implementation or report an original retailer experiment. Instead, it distinguishes the empirical findings of cited studies, the descriptive statistics displayed in its figures, and the design claims developed through synthesis. This distinction provides the evidentiary basis for the literature review, methods, results, discussion, and conclusion that follow.

2. Literature Review

2.1 Forecasting as a Retail Decision Capability

Retail analytics provides the informational foundation for connecting forecasts with business decisions. Chen et al. (2012) position predictive modeling within the broader development of business intelligence and analytics, while Grewal et al. (2017) emphasize technology, data, and customer engagement in retail transformation. Their contributions establish why transaction records should be interpreted alongside product, customer, channel, and contextual information. They do not establish that additional data are always useful or that automated recommendations should replace professional judgment. For the present framework, an input is valuable when it improves a specified decision or supports accountability for that decision. Historical sales may inform replenishment, whereas an upcoming event may alter the timing of demand. A sourcing certificate performs a different function because it can determine whether replenishment is permissible under a retailer's commitments. Treating these inputs as interchangeable predictors would obscure the distinction between estimating demand and governing the actions taken in response.

Model choice should follow the demand structure and decision horizon rather than assumptions about technological sophistication. Martins and Galegale (2022) compare traditional retail forecasting approaches with machine learning and emphasize the relevance of endogenous and exogenous variables. Nasseri et al. (2023) report

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

that tree-based ensembles outperformed a long short-term memory model across their evaluated retail categories, while also showing variation across categories. These findings support benchmarking, not a universal claim that one model family dominates. A jewelry retailer may face relatively stable demand for a standard chain but highly uneven demand for a distinctive gemstone setting. Store-level and product-level series may therefore require different representations or pooling strategies. Lead times also determine whether daily prediction improvements matter to purchasing decisions made months before delivery. Within the proposed framework, the practical unit of comparison is a complete forecasting-and-decision policy operating under the same information and inventory constraints.

Sparse demand creates additional measurement problems that should not be hidden by aggregate accuracy scores. A product can record zero sales because no customer requested it, because it was unavailable, or because an acceptable substitute absorbed the purchase. These possibilities have different implications for forecasting and stock control. Hyndman and Athanasopoulos (2021) explain that percentage errors become undefined or unstable when actual observations approach zero and that evaluation requires data not used in model fitting. The framework consequently retains absolute and scaled measures alongside carefully qualified percentage measures. It also requires explicit treatment of missing observations, stockout periods, discontinued items, and cold-start products. A model that predicts repeated zeros can appear successful while failing to support a retailer's intended service level. Forecast evaluation must therefore identify the exposure, aggregation level, and business consequence represented by each reported metric.

2.2 Decision-Focused and Probabilistic Forecasting

Decision-focused forecasting distinguishes a useful estimate from an estimate that performs well under an arbitrary statistical loss. Elmachtoub and Grigas (2022) formalize this distinction by connecting predictive learning with the objective and constraints of a downstream optimization problem. Their work demonstrates why prediction error and decision error need not identify the same preferred model. The implication for jewelry is conceptual rather than a claim that their experimental results transfer directly to retail sourcing. Underestimating demand for a readily replenished product may have a different consequence from underestimating demand for a unique piece with a long lead time. Similarly, overestimating demand can have different consequences when an item is returnable, repairable, or difficult to remarket. The proposed architecture therefore evaluates forecasts through the actions they induce under explicitly stated costs and constraints. This approach preserves accuracy as an important diagnostic while refusing to treat it as the final operational objective.

Recent work further strengthens the case for linking forecasting benchmarks with inventory evaluation. Alsaif (2026) compares retail forecasting models and evaluates their outputs through holding and shortage costs in a published simulation. The study illustrates that model assessment can move beyond prediction scores toward consequences for stock decisions. Its results remain conditional on the selected datasets, inventory policy, and cost assumptions. They do not validate jewelry-specific service targets or responsible sourcing outcomes. The present article uses selected reported costs only as an external methodological illustration. It does not reproduce the experiment, combine those estimates with jewelry market statistics, or claim that equivalent savings would occur in practice. This boundary is essential because a general retail benchmark can motivate an evaluation design without establishing the effectiveness of a new sector-specific framework.

Probabilistic forecasting adds information about uncertainty that a single expected value cannot express. For a retailer deciding how much to order, the distribution of possible demand can matter more than a small improvement in average error. Zambon et al. (2026) propose coherent probabilistic forecasting across large retail hierarchies and report evaluations on M5 and Favorita data. Their preprint addresses the problem of producing forecasts that remain consistent across aggregation levels while representing uncertainty. This direction is relevant to jewelry planning because store, category, and product decisions should not rely on contradictory totals. Nevertheless, historical disaggregation patterns may become unreliable when product mixes or customer preferences change. A provenance-aware system must therefore distinguish uncertainty in demand from uncertainty in source documentation. The two can interact in a decision, but they represent different knowledge gaps and require different corrective actions.

Foundation models broaden the comparison set but introduce their own evaluation responsibilities. Chronos-2 supports inference across multiple forecasting configurations and uses relationships among series and covariates to inform predictions (Ansari et al., 2025). For smaller retailers, such models may warrant testing because they offer alternatives to building every model from the beginning. That possibility is not evidence that external pretrained systems are always cheaper, more accurate, or easier to govern. Vendor access conditions, data handling, reproducibility, and inference requirements remain deployment questions. Moreover, a model's broad benchmark performance does not establish that it understands jewelry materials, brand codes, or sourcing obligations. In this framework, foundation models enter the same controlled comparison as conventional baselines and tree-based approaches. No model receives exemption from local evaluation merely because its training corpus or architecture is larger.

2.3 Jewelry-Specific Analytics and Sustainability

Jewelry-specific studies provide relevant examples while also exposing the limits of the available evidence. Han (2025) develops a decision-support approach linking customer relationship management data with jewelry inventory management. Munde and Kaur (2024) examine machine learning for predicting sustainable jewelry purchases. These contributions address different outcomes: customer-centered inventory decisions in one case and purchase behavior in the other. Neither should be treated as direct validation of a complete architecture combining forecasting, provenance constraints, and responsible AI governance. Their value for the present study lies in identifying plausible data relationships and sector-specific implementation questions. Customer information may inform assortment, but it also requires rules governing access and acceptable use. Likewise, predicting interest in sustainable products does not establish that the products purchased have lower environmental impacts.

The meaning of sustainability in fine jewelry requires separation of material attributes, supply-chain practices, and product outcomes. Perret et al. (2025) use expert interviews to examine sustainability and diversification in high jewelry, providing a strategic perspective rather than a forecasting evaluation. Their findings support considering sourcing and product positioning together, but not assuming that all consumers make equivalent trade-offs. An

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

item may contain recycled metal while relying on a gemstone whose origin is poorly documented. Another may have strong source documentation without a product-level estimate of energy or water use. For the proposed framework, these are separate attributes rather than competing versions of a single sustainability score. Preserving that separation makes trade-offs visible and reduces the temptation to present incomplete evidence as comprehensive environmental performance. It also allows planners to distinguish what the organization knows, what it has verified, and what remains uncertain.

Gemological identification illustrates why provenance must not be confused with environmental superiority. Eaton-Magaña et al. (2024) describe the identification and characterization of laboratory-grown diamonds through gemological evidence and advanced testing. Their research supports accurate classification of material origin, not a universal environmental ranking of laboratory-grown and mined stones. Such a ranking would require comparable system boundaries, production information, and impact measures. A retailer should consequently maintain separate fields for origin, treatment, certification, and environmental documentation. Those fields can support informed comparison without asserting that a single label resolves every sustainability concern. The framework also preserves the distinction between verified records and supplier declarations. This approach makes responsible communication dependent on the evidence available for a particular product rather than on a favorable category narrative.

Emerging computational work on environmental assessment should be interpreted with similar restraint. Jiang et al. (2025) investigate jewelry eco-design using an artificial neural network and a small literature-derived design dataset. Their study links material and design attributes with environmental indicators, offering an illustration of computational assessment beyond demand prediction. However, its scale and heterogeneous evidence base limit its usefulness as a universal source of product impact factors. The present framework therefore does not transfer its numerical estimates into retailer dashboards or treat them as measured impacts for fine jewelry assortments. Instead, the study motivates a separate environmental evidence layer that records assumptions and uncertainty. Product-level comparisons would require compatible functional units and traceable calculations. Environmental modeling can complement retail planning only when the meaning and limitations of its outputs remain visible.

2.4 Traceability, Verification, and Operational Provenance

Traceability systems create the possibility of connecting products with their material and transactional histories. Patel et al. (2025) develop a jewelry supply-chain prototype using Ethereum and certificate records supported by the InterPlanetary File System. Fortuna and Gaspar (2022) examine automatic identification and data capture in luxury metal-parts manufacturing. Together, these studies show that different infrastructures can support the recording and retrieval of product information. Their technical feasibility does not mean that all stored claims are independently verified or that planners will use the records. A durable record of an inaccurate declaration remains an inaccurate declaration. The proposed framework therefore distinguishes record integrity from the truth and completeness of the underlying claim. It also treats blockchain as one possible implementation rather than a necessary condition of responsible jewelry planning.

Operational provenance requires information that can alter a decision at the moment an action is considered. A certificate stored in an archive may support later reporting without influencing which supplier receives an order. By contrast, a structured record with material scope, issuer, validity period, and product linkage can identify whether a proposed purchase meets a stated sourcing requirement. The Responsible Jewellery Council's chain-of-custody approach supplies a sector-specific reference for traceable precious-metal handling (Responsible Jewellery Council [RJC], 2024). The framework does not treat organizational membership, certification, and product-level documentation as interchangeable. It proposes that each claim be represented at the level to which its supporting evidence actually applies. Unknown or expired information should trigger review rather than silent conversion into a favorable score. This representation turns provenance from a communication asset into a governed input to assortment and replenishment.

Adoption conditions matter because a technically sound system can remain disconnected from everyday planning. Kamble et al. (2019) examine blockchain adoption in supply chains and identify the relevance of perceived usefulness and ease of use. Their findings support attention to how an information system fits organizational practice, rather than assuming that technical capability guarantees use. In jewelry retail, the framework proposes interfaces that present only the sourcing evidence needed for the decision under review. Buyers should not need to interpret an entire ledger before comparing two replenishment options. At the same time, a simplified dashboard should retain links to the underlying evidence and its limitations. The design problem is therefore to reduce unnecessary complexity without erasing meaningful uncertainty. Effective provenance integration depends on usable decision routines as much as on the infrastructure that stores the records.

2.5 Responsible AI and the Integration Gap

Responsible AI research shifts attention from general commitments to the controls needed throughout an operating system. Morley et al. (2020) identify the challenge of translating ethical principles into practical methods across machine-learning development. The National Institute of Standards and Technology's AI Risk Management Framework organizes risk activities around govern, map, measure, and manage (Tabassi, 2023). Shaik et al. (2026) apply responsible AI concerns to demand forecasting, including transparency, privacy, fairness, and accountability. These sources provide a vocabulary for designing controls but do not establish a ready-made jewelry implementation. The present framework uses them to specify ownership, review thresholds, documentation, and feedback. It treats ethical governance as a continuing decision function rather than a final approval step. A forecast can be statistically acceptable while the system that uses it remains poorly governed.

Human oversight should be designed as a meaningful capacity to question and change a recommendation. The updated OECD principles emphasize transparency, accountability, and human-centered values as features of trustworthy AI (OECD, 2024). In a retail setting, those commitments imply more than placing an approval button beside an automated order. The reviewer needs relevant evidence, sufficient time, defined authority, and an escalation path for unresolved conflicts. Otherwise, nominal oversight may conceal an operational expectation that employees accept the model. The proposed framework also separates predictive explanations from causal claims. Identifying a variable

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

associated with a forecast does not prove that changing that variable would produce the expected business outcome. A responsible explanation should clarify what the model used, what remains uncertain, and why the recommended action satisfies the applicable decision rules.

The research gap is consequently narrower than a general absence of forecasting, traceability, or governance scholarship. Decision-focused research already connects prediction with downstream optimization, and recent retail work evaluates inventory consequences. Jewelry studies already examine customer analytics, sustainable purchase behavior, and traceability infrastructure. What remains insufficiently established in the reviewed evidence is how these capabilities should operate together in a repeatable jewelry planning process. Specifically, the literature does not validate the combined effect of demand uncertainty, evidence-qualified provenance constraints, and meaningful human review on operational and sustainability outcomes. The present contribution addresses that integration problem through a conceptual artifact, not a claim of empirical resolution. Its novelty rests in the relationships among components and the conditions under which those relationships should be tested. This positioning avoids presenting a collection of existing technologies as though their coexistence alone constituted demonstrated innovation.

3. Methods

3.1 Research Design and Evidentiary Boundaries

The study uses conceptual framework development informed by design-science research and supplemented by descriptive secondary evidence. Hevner et al. (2004) describe design science as the development and evaluation of artifacts intended to address organizational problems. Here, the artifact is a decision architecture specifying information requirements, operational functions, governance controls, and testable relationships. The work does not complete a full implementation-and-evaluation cycle. It instead establishes a design rationale and an evaluation agenda that later studies can operationalize. The analytical starting point was the integration of forecasting, provenance, and governance rather than the evaluation of a preferred software platform or vendor. The analysis develops that structure through current literature and explicitly sourced visualizations. No retailer was recruited, no participant data were collected, and no forecasting model was trained for this article.

The evidence base combines foundational scholarship with a targeted literature update completed on October 9, 2026. Searches used combinations of jewelry or jewellery, demand forecasting, inventory planning, provenance, traceability, responsible AI, decision-focused learning, foundation models, and probabilistic hierarchical forecasting. Publisher websites, author-hosted research repositories, and institutional sources supplied the materials reviewed. Priority was given to original studies, official guidance, and directly reported data relevant to the framework's components. Searches also followed citations and verified bibliographic details where accessible. Publication and online-release dates were checked to exclude material not yet available at the review cutoff. This procedure is a purposive narrative update, not a systematic review or a claim of exhaustive database coverage. Accordingly, the article does not report invented screening counts, a formal meta-analysis, or a comprehensive ranking of all available studies.

Evidence was interpreted according to its source, research design, and relevance to the proposed application. Peer-reviewed studies informed the discussion of established methods and reported implementations, while preprints were explicitly identified as emerging research. Institutional guidance informed governance and sourcing requirements but was not treated as evidence of intervention effectiveness. Where full text was unavailable, interpretation was confined to the accessible abstract, publisher description, or reported findings. Numerical results were retained only when their definitions and original context could be identified. Findings from general retail, product design, and jewelry analytics were not pooled into a synthetic effect estimate. Instead, each source contributed to a specific design requirement or an explicit boundary condition. This approach supports conceptual integration while preserving the limits of heterogeneous evidence.

3.2 Descriptive Data and Visualization Procedures

Three bar charts provide contextual and methodological evidence without implying an original experiment. Figure 1 uses annual global gold jewelry consumption reported in the World Gold Council's full-year 2025 publication. Figure 2 uses the same calendar quarter across two years for India, mainland China, and the United States from the Q2 2026 publication. These markets provide geographically distinct illustrations rather than a representative sample of fine jewelry retailers. Figure 3 redraws selected total inventory costs from Alsaif's (2026) published retail simulation. The figures use the values reported by their sources, with zero-based axes and explicit measurement units. No missing values were imputed, no confidence intervals were invented, and no source results were presented as observations generated by this study. Figure captions identify the relevant reporting periods, data provenance, and limits of interpretation.

The descriptive calculations preserve comparability within each source rather than creating a combined historical series across publication vintages. Changes in annual consumption were calculated as the later value minus the earlier value, divided by the earlier value. Quarterly comparisons use Q2 against Q2 rather than mixing quarterly and annual totals. Published estimates can be revised, so figures from different release dates are not assumed to reconcile exactly. Gold weight is not treated as the number of jewelry items sold, and source-simulation dollars are not treated as retailer revenue or observed savings. The benchmark chart illustrates an evaluation approach rather than estimating the value of the proposed framework. These choices prevent incompatible units from creating misleading visual comparisons. They also preserve a clear distinction between descriptive arithmetic, reported external findings, and conceptual interpretation.

3.3 Framework Construction and Evaluation Logic

Framework construction followed a structured analytical sequence rather than a formal coding study. First, the planning problem was specified in terms of actions, information gaps, and consequences. Second, relevant evidence was mapped to data inputs, prediction, provenance, governance, and human-reviewed outputs. Third, the relationships among those functions were examined through hypothetical planning conflicts, including incomplete documentation and disagreement between a forecast and a buyer. Fourth, design principles were formulated to identify requirements that should remain stable across alternative technologies. Fifth, propositions were paired with measurable outcomes and comparison conditions. These steps produced an analytic design

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

argument; they did not produce independent coder agreement or empirical validation. Table 1 summarizes the architecture, and Table 2 translates its propositions into an evaluation agenda.

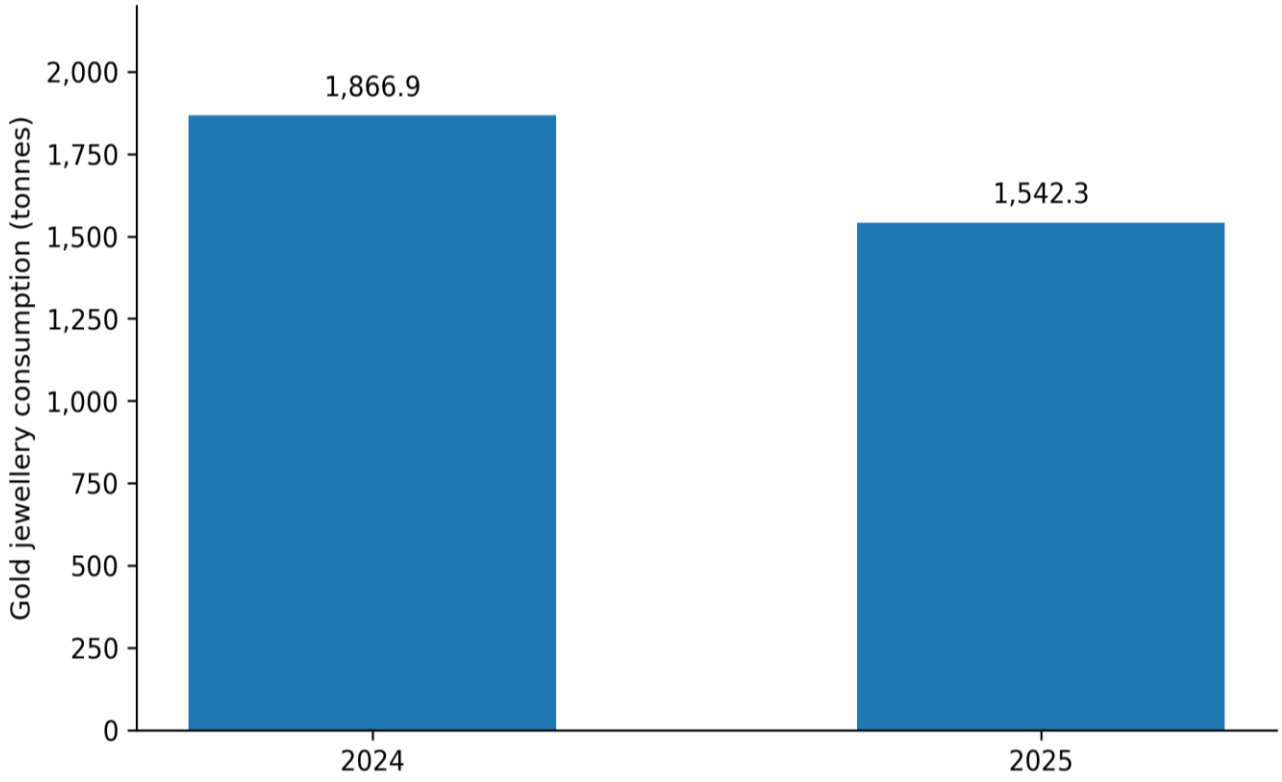
The proposed unit of future evaluation is the decision episode, linked to a product, location, planning horizon, and responsible reviewer. A decision episode records the information available before the recommendation, the action proposed, any human modification, and the eventual outcome. Forecast-only, provenance-aware, and fully governed policies can then be compared under equivalent operational conditions. Temporal holdouts should preserve the sequence of information availability, and prospective studies should define service and sourcing thresholds before observing results. Actual sales cannot automatically be treated as unconstrained demand, especially when stockouts censor purchases. Environmental outcomes require additional product and process information beyond inventory records. These requirements remain a prospective protocol rather than a description of experiments completed here. Before deployment, researchers should register the primary outcome, minimum meaningful improvement, permitted exclusions, and stopping conditions. They should also document how delays, unavailable products, substitutions, and missing provenance affect the comparison, so that favorable results cannot be manufactured through retrospective changes to the evaluation rules or population. The resulting framework is therefore analytically specified and empirically testable, but its effectiveness remains an open research question.

4. Results

4.1 Descriptive Evidence and Its Interpretation

The descriptive evidence demonstrates why a planning system should preserve multiple measures of demand. Figure 1 displays the annual gold jewelry volumes introduced above; the reported values imply a decline of 324.6 tonnes, or approximately 17.4%, between 2024 and 2025. The calculation uses the report's displayed tonnages rather than substituting its rounded percentage summary. It does not measure a change in unit sales or environmental impact. The analytical result is therefore limited but useful: a revenue-oriented system can face a materially different trajectory from a system tracking physical gold consumption. That divergence warrants separate targets and dashboards rather than a single undifferentiated demand indicator. In the proposed architecture, those targets inform different planning decisions and should be reconciled explicitly. The chart supplies contextual evidence for that design choice, not proof that AI would improve the resulting decisions.

Figure 1 Global Gold Jewelry Consumption, 2024 and 2025



Quarterly comparisons also indicate why planners should avoid imposing a single aggregate trend on every market. Figure 2 displays the source's Q2 values for three selected countries, preserving the same measurement unit and comparison period (World Gold Council, 2026b). The bars show declines in all three examples, but their levels and absolute changes differ. These differences do not reveal individual retailers' customer mixes, product strategies, or supply constraints. They therefore cannot justify uniform percentage adjustments to store orders. Instead, the framework treats aggregate evidence as contextual information that must be evaluated alongside local records. A planner should be able to distinguish a broad market signal from a product-specific inference. This separation prevents a visually persuasive macroeconomic trend from acquiring more decision authority than its evidentiary scope warrants.

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

Figure 2 Gold Jewelry Consumption in Three Selected Markets, Q2 2025 and Q2 2026

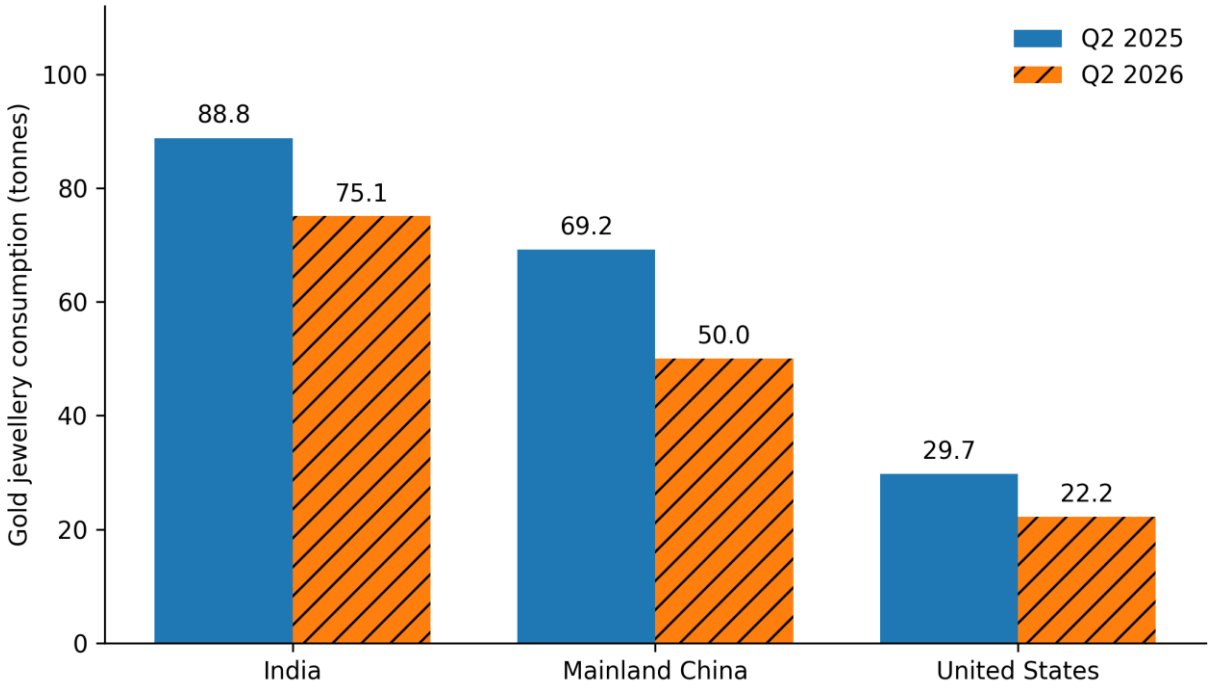
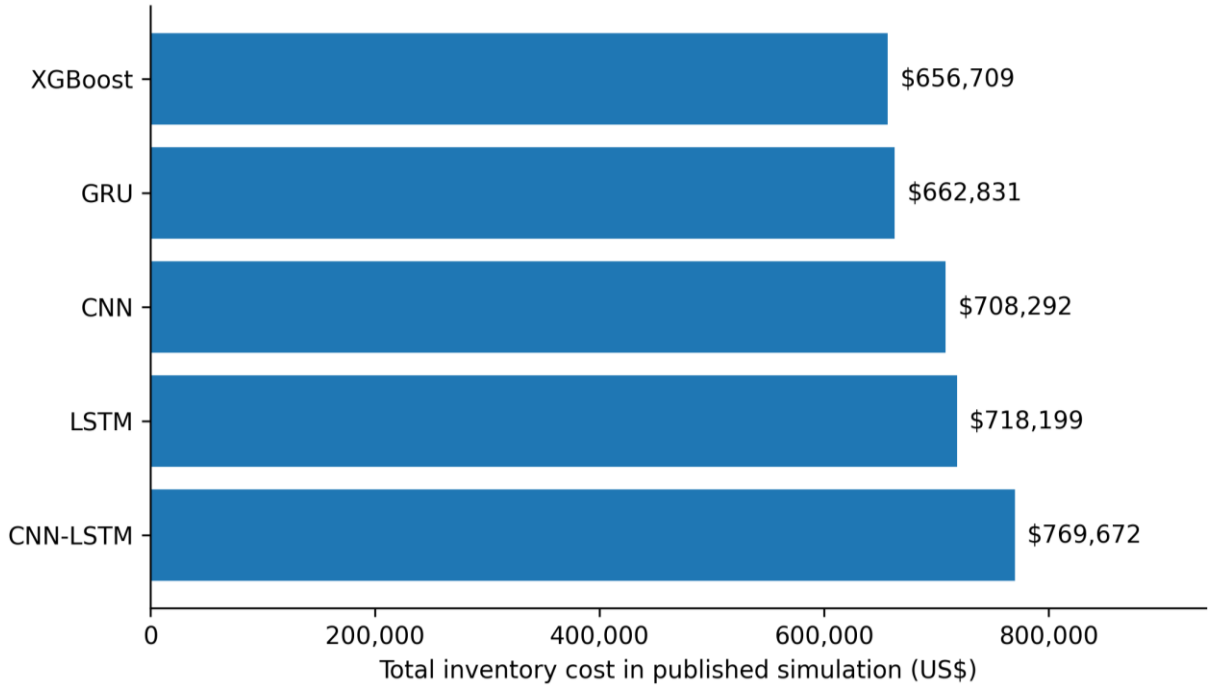


Figure 3 illustrates an additional distinction between forecasting performance and operational consequences. The plotted values reproduce reported total costs for five models in Alsaif's (2026) inventory simulation; they are not new experimental results. XGBoost has the lowest displayed total, but that ordering should not be interpreted as a recommendation for jewelry deployment. The values depend on the source's particular policy, horizon, and cost assumptions. Neither provenance restrictions nor jewelry-specific service obligations are represented by the comparison. Its contribution here is to make the decision-evaluation layer visible and concrete. A corresponding jewelry study would need independently justified parameters and locally relevant baselines. The chart therefore supports the proposed evaluation logic while leaving the framework's predicted benefits unconfirmed.

Figure 3 Inventory Costs Reported in an External Retail Forecasting Simulation



4.2 Integrated Decision Architecture

The principal conceptual result is a five-function architecture in which governance operates across the planning cycle rather than only after prediction. Table 1 summarizes the data-input layer, predictive analytics module, traceability and provenance module, ethical governance module, and decision outputs with feedback loops. These functions remain analytically distinct even when a single software platform implements several of them. Data preparation determines what information can legitimately enter the process. Prediction characterizes possible demand rather than authorizing purchases by itself. Provenance qualifies the sourcing

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

claims relevant to candidate actions. Governance assigns responsibility and determines which actions require review, while the output function records decisions and subsequent outcomes. The architecture's proposed value arises from these connections, not from requiring a particular database, algorithm, or commercial product.

Table 1 Proposed Architecture for Provenance-Aware Retail Planning

Function	Required information or controls	Decision role and failure check
Data inputs	Time-stamped sales, availability, product attributes, permissible customer information, sourcing records, and missing-data flags.	Establish the information available before a decision. Check identifier mismatches, unavailable items, and unsupported assumptions.
Predictive analytics	Benchmarks, appropriate horizons, forecast uncertainty, temporal holdouts, and product or location hierarchies.	Estimate demand and compare complete planning policies. Check leakage, poorly calibrated uncertainty, and aggregation errors.
Provenance and traceability	Product-linked origin claims, issuer, scope, validity, verification status, and unresolved evidence gaps.	Constrain or qualify sourcing options. Check expired evidence, mistaken scope, and conflation of documentation with environmental performance.
Ethical governance	Named owners, access rules, escalation thresholds, explanation requirements, and meaningful review authority.	Determine who can approve or contest an action. Check rubber-stamping, excessive alerts, and unreviewed high-consequence exceptions.
Decisions and feedback	Recommended and approved actions, overrides, execution records, outcomes, and follow-up responsibilities.	Link forecasts to actions and learning. Distinguish model error from policy, execution, supplier, or data failure.

The data-input function establishes a common decision record without assuming that every available datum should be collected. Required retail fields include product identifiers, location, observation date, availability, sales, returns, and relevant replenishment information. Additional attributes may include material composition, price band, collection, planned events, and channels, subject to documented usefulness. Customer information should enter only where its purpose and permissible use have been specified. Provenance records require an explicit link to the product or material lot rather than an informal association with a supplier's name. Each field should retain a timestamp and an indication of its source. Missing, disputed, expired, and verified values should remain distinguishable. This structure prevents a superficially complete dataset from concealing gaps that could change the acceptability of a recommendation.

The predictive function estimates demand at the horizon and aggregation level needed for the decision. Its comparison set should include a defensible baseline, a suitable statistical method, and any machine-learning or pretrained alternative justified by the available data. Forecast distributions or prediction intervals should accompany point estimates when they can be estimated and evaluated appropriately. Future-known variables must be separated from information unavailable when the order would actually have been placed. A model should not receive credit for retrospectively using a promotion outcome or certification update that the planner could not have known. Evaluation should also distinguish new products from established series and record the consequences of aggregation. An apparent category-level improvement may coexist with poor allocation among individual items. The proposed module therefore reports both predictive diagnostics and the decisions generated under the specified planning policy.

The provenance function determines how sourcing evidence changes the available action set. Some attributes operate as hard constraints, such as an organization's requirement for valid documentation before a specified purchase. Others operate as preferences or review triggers when the organization has explicitly authorized a trade-off. A missing record is not equivalent to a negative environmental finding, but neither should it be silently treated as proof of acceptability. The system should identify what evidence is absent, whether an alternative exists, and who can resolve the exception. It should also distinguish the quality of documentation from the underlying material or supplier characteristic. Two products may therefore have similar forecast demand while differing in what the retailer can substantiate about their origins. By making that difference operational, the module turns provenance into an input that can change assortment and replenishment decisions.

The governance function defines the conditions under which a recommendation can proceed. It assigns owners for data quality, model performance, sourcing evidence, customer-data use, and final approval. It also establishes thresholds for escalation, including uncertainty, unusual order size, disputed provenance, or material disagreement with a baseline. Reviewers receive the evidence relevant to those thresholds rather than an unexplained aggregate score. They may approve, modify, defer, or reject a recommendation, with a recorded rationale proportional to the decision's consequence. A high-risk exception should not disappear within a routine batch approval. Conversely, low-consequence decisions should not require documentation so burdensome that staff bypass the process. The design objective is a review structure that makes responsibility practicable and meaningful rather than merely nominal.

The decision-output function converts analysis into an accountable operational record. A replenishment recommendation should identify the product, proposed quantity, horizon, expected uncertainty, relevant sourcing conditions, and available

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

alternatives. It should distinguish the model's forecast from the organization's policy for translating that forecast into an order. The final record should then capture the actual action, reviewer, rationale for any change, and evidence available at approval. Later observations can reveal whether outcomes differ because of forecast error, execution failure, supplier disruption, or human modification. Those distinctions matter because retraining a model cannot correct every operational problem. Feedback should therefore improve data, decision rules, and organizational practice as well as predictive parameters. This closes the planning cycle without assuming that every observed outcome is an unbiased measure of the original recommendation's quality.

4.3 Design Principles and Testable Propositions

Five design principles translate the architecture into requirements that should survive changes in technology. First, forecasting must remain decision-linked, with an identifiable connection between an estimate and an operational action. Second, provenance must be operationalized through evidence-qualified attributes, constraints, or review conditions. Third, responsible AI governance must be embedded across data preparation, prediction, approval, and monitoring. Fourth, sustainability claims must distinguish operational proxies, sourcing documentation, and independently measured environmental outcomes. Fifth, human expertise must remain actionable through informed review, justified overrides, and opportunities to correct the system. These principles extend the original integration argument by specifying what an implementation must make observable. They do not prescribe that every retailer use identical thresholds or assign equal importance to every objective. Instead, they require organizations to disclose the choices through which a locally appropriate planning policy becomes responsible and reviewable.

The propositions specify relationships for future testing rather than reporting relationships established by this study. Proposition 1 links multi-signal forecasting to improved decision usefulness only when relevant information and suitable baselines are available. Proposition 2 treats inventory and purchasing actions as the pathway through which forecast improvements could reduce operational inefficiency. Proposition 3 predicts that evidence-qualified provenance changes sourcing alignment when its attributes actually constrain or inform decisions. Proposition 4 links contestable recommendations and meaningful oversight with calibrated reliance rather than unquestioning adoption. Proposition 5 connects documented feedback with improvements in data quality, planning policy, and sustainability alignment over time. Table 2 pairs each proposition with outcomes and conditions that could disconfirm it. Together, the propositions create an evaluation agenda while explicitly permitting null findings, trade-offs, and implementation failure.

Table 2 Propositions and Prospective Evaluation Logic

Proposition	Comparison or mechanism	Candidate outcomes	Boundary or disconfirmation
P1. Multi-signal forecasting improves decision usefulness.	Compare added signals with a common baseline and the same planning policy.	Absolute or scaled error; calibrated intervals; service; decision costs.	Additional signals provide no material benefit, introduce leakage, or worsen allocation.
P2. Actions mediate operational benefits of better forecasts.	Record whether improved forecasts actually change orders, allocations, or purchasing.	Aged inventory; turnover; shortages; replenishment frequency; markdown exposure.	Accuracy improves but actions or operational outcomes do not improve.
P3. Operational provenance improves sourcing alignment.	Compare evidence-qualified sourcing rules with documentation-only reporting.	Purchases meeting declared sourcing rules; unresolved exceptions; evidence completeness.	Documentation improves without changing decisions, or unacceptable service costs result.
P4. Contestability supports calibrated reliance.	Compare meaningful review and usable evidence with minimally explained recommendations.	Justified approvals and overrides; review time; unresolved risks; inappropriate reliance.	Approval rises without better decisions, or reviewers lack effective authority.
P5. Documented feedback improves system alignment.	Compare repeated planning cycles with explicit records, owners, and corrective actions.	Corrected data gaps; drift responses; policy revisions; sourcing and service outcomes.	Feedback is collected but ignored, or interventions do not improve relevant outcomes.

A hypothetical replenishment conflict illustrates the architecture without introducing synthetic performance results. Suppose a buyer considers two otherwise comparable products with similar anticipated demand but different documentation status and delivery dates. An accuracy-only process might rank them by expected sales without exposing the documentation gap. The proposed framework instead presents the sourcing condition, the uncertainty surrounding each option, and any permissible alternatives. The buyer could wait for verification, select an acceptable substitute, or escalate a documented exception under organizational policy. No outcome is assumed to be universally optimal because service obligations and sourcing commitments differ. The example demonstrates how the framework structures a decision rather than proving that its decision will be better. Its value becomes an empirical question once the retailer's objectives, constraints, and eventual outcomes are observed.

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

5. Discussion

5.1 Theoretical Contribution and Scope

The framework's first contribution is to define responsible retail intelligence at the level of a decision system rather than an isolated model. Existing forecasting research provides important methods for estimating demand, while traceability and governance research provide different capabilities. The present synthesis specifies how these capabilities must interact before a recommendation becomes an acceptable action. Forecast accuracy remains relevant, but it occupies a narrower role within a larger account of decision quality. A product recommendation can be predictable and profitable while failing the retailer's sourcing requirements. Conversely, a fully documented product can be commercially inappropriate or poorly allocated. The integration therefore cannot be reduced to adding an ethical paragraph to a forecasting model or attaching certificates to a dashboard. It requires a shared representation of the action, its supporting evidence, and the authority under which it is approved.

A second contribution is the distinction between three forms of evidence that sustainability discussions can otherwise conflate. Operational evidence concerns outcomes such as inventory aging, service levels, replenishment frequency, and markdown exposure. Provenance evidence concerns the origin, custody, verification, and documentation of products and materials. Environmental evidence concerns consequences such as emissions, resource use, or other impacts assessed within defined boundaries. Improvement in one category does not logically establish improvement in the others. For example, a retailer could increase inventory turnover by promoting rapid replacement without reducing material throughput. Alternatively, stronger provenance documentation could improve accountability while leaving measured environmental impacts unchanged. By separating the categories, the framework makes both positive findings and uncomfortable trade-offs more interpretable.

A third contribution concerns how responsibility is distributed across people, models, and organizational policies. Forecasting software estimates patterns, but managers determine which objectives matter, what evidence is sufficient, and which exceptions are permissible. Human review does not automatically resolve those choices because reviewers can face workload, incentives, and information limitations. The framework therefore treats oversight as a designed organizational capability with measurable conditions. This perspective differs from equating responsibility with the mere presence of a human approver. It also differs from assuming that the most interpretable model necessarily yields the most responsible decision. A transparent forecast can still be used under an inappropriate policy or supported by unreliable provenance records. Responsibility arises from the relationship among information quality, decision rules, accountable actors, and continuing evaluation.

The scope of the contribution should remain explicit because conceptual integration is not equivalent to demonstrated effectiveness. The reviewed research establishes the relevance or feasibility of individual components under different conditions. It does not supply a pooled effect size for the combined architecture. Nor does the framework prove that retailers will accept its governance requirements or that consumers will reward the resulting practices. Table 3 summarizes the evidentiary roles and boundaries of selected sources to make that distinction visible. The conceptual contribution is strongest where it identifies mechanisms that can fail and conditions that can be tested. An implementation that adds documentation but does not change decisions would not support the central integration claim. Similarly, improved accuracy without better operational performance would challenge the assumption that the chosen model contributes meaningful planning value.

Table 3 Evidence Roles and Limits in the Framework

Evidence or source	Role in the synthesis	Inference not warranted
World Gold Council (2026a, 2026b)	Published context for physical gold jewelry demand and differences among selected markets.	Retailer-level causal effects, unit sales, or changes in environmental burden.
Alsaif (2026)	External example of evaluating forecasts through inventory simulation costs.	Expected jewelry savings or validation of provenance-aware planning.
Han (2025); Munde and Kaur (2024)	Jewelry-specific analytics for inventory and sustainable purchase behavior.	An established effect of the complete integrated framework.
Patel et al. (2025); Fortuna and Gaspar (2022)	Examples of traceability infrastructure and information capture.	Automatic truth of recorded claims or routine use in buying decisions.
Morley et al. (2020); Shaik et al. (2026); Tabassi (2023)	Principles and methods informing governance across the AI life cycle.	Empirically demonstrated effectiveness of the proposed local controls.
Ansari et al. (2025); Zambon et al. (2026)	Emerging preprint research informing forecasting alternatives and uncertainty.	Peer-reviewed validation in fine jewelry or guaranteed transferability.
Jiang et al. (2025)	Illustrative computational treatment of jewelry design and environmental indicators.	Universal product impact factors or verified impacts for retailer assortments.

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

5.2 Managerial Implementation and Organizational Readiness

Implementation should begin with a narrow, consequential decision rather than a purchase of the most sophisticated technology. A retailer might select replenishment for a repeatable product category, allocation across a small group of stores, or sourcing review for a defined assortment. The initial task is to identify the current decision owner, available information, recurrent failures, and existing baseline. This mapping should include how employees currently use client knowledge, supplier relationships, and product narratives. Otherwise, a technical project may unknowingly remove information that supports effective judgment. The team should also identify which outcomes would justify the cost and disruption of implementation. A smaller pilot with credible records and explicit responsibilities can be more informative than a broad deployment with ambiguous success criteria. This recommendation follows from the framework's emphasis on observable decision episodes rather than technological coverage.

A staged implementation can operationalize the proposed framework's five-part roadmap while making its dependencies clearer. The first stage establishes data readiness through consistent identifiers, time records, availability information, and sourcing fields. The second documents baseline forecasts and current inventory policies before introducing a new model. The third tests AI-assisted alternatives under the same horizons, inputs, and operational assumptions. The fourth connects provenance evidence to actual decision rules, including treatment of missing or disputed records. The fifth evaluates governance and feedback before expanding the system's scope. These stages describe capability development rather than a justification for postponing safeguards until deployment. Privacy, access, and accountability requirements begin in the first stage and become more operationally specific as the pilot develops.

A practical dashboard should expose the differences among forecast, recommendation, and approved action. These are separate objects even when a user interface presents them in a single row. The forecast describes anticipated demand; the recommendation applies a policy to that forecast; the approved action reflects the organization's final decision. Combining them would make it difficult to determine whether a poor result originated in estimation, policy, or execution. The dashboard should therefore show relevant uncertainty, provenance conditions, and exceptions alongside the action record. It should also provide access to supporting evidence without forcing every reviewer to inspect the entire underlying system. The goal is a concise decision interface with a traceable evidentiary foundation. A high volume of notifications is not evidence of good governance when employees cannot identify the few issues that require meaningful attention.

Implementation also requires a distinction between useful personalization and unjustified preference in service allocation. Customer information may help a retailer anticipate an occasion or identify an appropriate product, but optimization objectives can privilege already well-served groups. The framework does not assume that equal inventory distribution across all stores or customers is the appropriate fairness criterion. Differences in demand, logistics, and product strategy may justify unequal allocations. Nevertheless, those differences should be examined rather than automatically accepted because a model reproduces them. Planners should specify which service disparities warrant review and which variables are legitimate for the intended decision. Sensitive or unnecessary customer information should not become a default input simply because it improves a statistical score. This makes fairness evaluation context-dependent without making it optional or unaccountable.

Supplier participation introduces another implementation tension because documentation capacity can differ across businesses. A strict evidence threshold may reduce exposure to unsupported claims while making participation more difficult for smaller or less digitized suppliers. The framework therefore separates the identification of missing evidence from an unsupported judgment about a supplier's overall responsibility. Retailers can establish remediation pathways, phased documentation requirements, and additional verification where appropriate. Such pathways should not become a mechanism for repeatedly approving unresolved high-consequence gaps. They should instead clarify what evidence is needed, who will provide it, and when the decision must be reconsidered. A responsible implementation consequently combines credible requirements with a transparent process for improving evidence quality. The objective is not to replace sourcing judgment with a single administrative score.

5.3 Empirical Validation and Research Priorities

A first validation study should establish whether the proposed information and decision records can be maintained reliably. Before testing performance benefits, researchers should determine whether products, sourcing records, recommendations, and actions can be linked without material ambiguity. They should also assess whether reviewers understand the explanations and possess actual authority to change a recommendation. A retrospective data audit can identify missing availability histories, inconsistent product identifiers, and undocumented policy changes. A prospective feasibility pilot can then reveal whether recordkeeping requirements are practical during routine work. These studies would evaluate the architecture's implementability rather than its causal effectiveness. Their findings should inform the next study instead of being presented as evidence of improved sustainability. This sequence reduces the risk that an elaborate comparison rests on data relationships that do not function in practice.

A second study should compare forecasting-and-decision policies under a common evaluation protocol. Candidate policies could include the existing process, an improved forecast without provenance integration, and a provenance-aware process with documented human review. The comparison should preserve relevant budget, lead-time, assortment, and service constraints. Temporal validation must ensure that predictions use only information available before the corresponding decision. Where a prospective randomized rollout is feasible, assignment should minimize spillovers among stores or shared inventories. Where randomization is not feasible, matched comparisons or phased implementation require careful assessment of alternative explanations. Results should report practical effects and uncertainty, not only whether a statistical threshold was crossed. A failure to improve performance should be examined as evidence about the proposed mechanism rather than dismissed as resistance to innovation.

A third research stream should evaluate reliance and override behavior as outcomes in their own right. Higher adoption is not necessarily better when users accept unreliable recommendations,

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

and frequent overrides are not necessarily worse when staff detect genuine problems. Researchers should distinguish justified corrections, unsupported departures, and cases in which the available evidence is insufficient to classify an override. Review time and documentation burden should be measured alongside decision quality. Interviews and observation can help explain how professional knowledge enters a decision, but they should not substitute for operational evidence of the consequences. Studies should also examine whether employees can raise sourcing concerns without being penalized for slowing a transaction. These questions connect human-centered AI with the institutional conditions under which oversight becomes effective. They also allow the framework's fourth proposition to be tested through calibrated reliance rather than simple approval rates.

Environmental evaluation requires an additional study design rather than an automatic interpretation of inventory improvement. Researchers must specify the relevant functional unit, organizational boundary, material flows, and comparison scenario. Reduced holdings may reflect postponed purchases, faster sales, transfers between stores, or genuine reductions in unnecessary production. Those mechanisms can have different environmental consequences even when an inventory metric moves in the same direction. Repair, resale, and take-back activities similarly require evidence about what they displace rather than only counts of participation. A credible assessment should therefore connect planning records with product and process information where those data exist. Where they do not, the study should report operational and provenance outcomes without converting them into unmeasured environmental benefits. This restraint preserves the usefulness of partial evidence while preventing sustainability claims from outrunning the analysis.

5.4 Limitations and Boundary Conditions

Several limitations constrain the interpretation of this article and its proposed research agenda. The narrative search was targeted rather than exhaustive, and accessible full texts were not available for every cited source. Emerging preprints may change through revision and peer review. The market figures cover gold jewelry, not all materials or the complete economics of fine jewelry retail. The external benchmark uses general retail data and simulation assumptions that cannot be imported unchanged into a jewelry setting. Moreover, the framework was not evaluated with practitioners, organizational records, or environmental measurements. Its five propositions should consequently be understood as reasoned expectations rather than established relationships. These limitations define what the article contributes: a transparent architecture and evaluation logic, not evidence that the integrated system already works.

The framework may also perform differently across business models that the broad category of fine jewelry can conceal. A large chain, an independent jeweler, a bespoke atelier, and a resale specialist do not share identical products, data volumes, or service obligations. Some businesses may benefit more from improving documentation and baseline planning than from introducing complex prediction. Others may require strong integration across stores, suppliers, and product hierarchies before a local pilot becomes informative. The architecture is intended to accommodate those differences, but adaptability itself requires testing. Future work should identify which functions are indispensable, which can remain lightweight, and which depend on organizational scale. Researchers should also document implementation costs and unintended exclusions rather than reporting benefits alone. Sector specificity becomes analytically valuable only when it clarifies these conditions rather than treating jewelry retail as a homogeneous setting.

6. Conclusion

Responsible AI in fine jewelry retail should be evaluated through the decisions it enables and the evidence those decisions can sustain. This article develops a framework that connects demand forecasting, provenance, and governance within a single planning architecture. Its five functions preserve the distinctions among data preparation, prediction, source verification, accountable review, and feedback. Its five design principles identify requirements for decision-linked forecasting, operational provenance, embedded governance, cautious sustainability claims, and meaningful human expertise. The accompanying propositions specify how those requirements can be examined empirically. Published market data and an external forecasting benchmark illustrate the relevance of the problem without constituting validation of the framework. The central contribution is therefore a structured and testable account of integration rather than an assertion of demonstrated technical or environmental performance. Fine jewelry provides a consequential setting in which that distinction matters.

The article's practical implication is that retailers should not treat a more accurate model, a traceability ledger, or a governance policy as a complete solution by itself. Forecasts need decision rules, provenance records need a defined role in those rules, and oversight needs evidence and authority. Operational improvement should be measured separately from documentation quality and environmental outcomes. These distinctions make it possible to identify progress without overstating what the available data demonstrate. They also reveal failure modes that an accuracy-only evaluation could miss, including unsupported sourcing claims and ineffective human review. A responsible implementation can begin with a modest decision scope while preserving a clear pathway to broader evaluation. Its success depends on the quality of the relationships among components rather than the number of technologies deployed. The appropriate next step is empirical investigation of those relationships under real organizational constraints.

Future research should examine whether the proposed integration improves planning usefulness, sourcing alignment, and calibrated reliance while maintaining acceptable service and implementation costs. Studies should compare complete decision policies, preserve the timing of information, and distinguish observed effects from assumptions. They should also report null findings and trade-offs that challenge the framework's expectations. Environmental claims require their own measurements and cannot be inferred merely from lower inventory or stronger certification coverage. Human expertise should remain visible in the evidence rather than disappearing into an unexplained override field. The resulting research agenda is deliberately demanding because responsibility cannot be established by technical plausibility alone. An accountable planning system must show what it knows, how it acts, who can challenge it, and what follows from its decisions. That is the standard through which AI-supported fine jewelry planning can advance beyond forecast accuracy.

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

References

1. Alsaif, S. A. (2026). A systematic benchmarking pipeline for demand forecasting and inventory-aware supply chain decision support. *Frontiers in Artificial Intelligence*, *9*, Article 1936360. https://doi.org/10.3389/frai.2026.1936360
2. Ansari, A. F., Shchur, O., Küken, J., Auer, A., Han, B., Mercado, P., Rangapuram, S. S., Shen, H., Stella, L., Zhang, X., Goswami, M., Kapoor, S., Maddix, D. C., Guerron, P., Hu, T., Yin, J., Erickson, N., Desai, P. M., Wang, H., ... Bohlke-Schneider, M. (2025). *Chronos-2: From univariate to universal forecasting* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2510.15821
3. Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165–1188. https://doi.org/10.2307/41703503
4. CIBJO. (2025, January 18). *An ESG model for the jewellery industry*. https://cibjo.org/an-esg-model-for-the-jewellery-industry/
5. Eaton-Magaña, S., Hardman, M. F., & Odake, S. (2024). Laboratory-grown diamonds: An update on identification and products evaluated at GIA. *Gems & Gemology*, *60*(2), 146–167. https://doi.org/10.5741/GEMS.60.2.146
6. Elmachtoub, A. N., & Grigas, P. (2022). Smart "predict, then optimize." *Management Science*, *68*(1), 9–26. https://doi.org/10.1287/mnsc.2020.3922
7. Fortuna, G., & Gaspar, P. D. (2022). Implementation of industrial traceability systems: A case study of a luxury metal pieces manufacturing company. *Processes*, *10*(11), Article 2444. https://doi.org/10.3390/pr10112444
8. Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The future of retailing. *Journal of Retailing*, *93*(1), 1–6. https://doi.org/10.1016/j.jretai.2016.12.008
9. Han, T.-T. (2025). Intelligent decision-making system for jewelry retail prediction and inventory management integrating CRM data. *Edelweiss Applied Science and Technology*, *9*(5), 1576–1592. https://doi.org/10.55214/25768484.v9i5.7235
10. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625
11. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. Retrieved October 9, 2026, from https://otexts.com/fpp3/
12. Jiang, A., Huan, M., Choi, D.-O., & Kang, Y. (2025). Optimizing eco-friendly jewelry design through an integrated eco-innovation approach using artificial neural networks. *Scientific Reports*, *15*, Article 6441. https://doi.org/10.1038/s41598-024-84477-y
13. Kamble, S. S., Gunasekaran, A., & Arha, H. (2019). Understanding the blockchain technology adoption in supply chains: Indian context. *International Journal of Production Research*, *57*(7), 2009–2033. https://doi.org/10.1080/00207543.2018.1518610
14. Martins, E., & Galegale, N. V. (2022). Retail sales forecasting information systems: Comparison between traditional methods and machine learning algorithms. In M. B. Nunes, P. Isaías, & P. Powell (Eds.), *Proceedings of the IADIS International Conference Information Systems 2022* (pp. 30–38). IADIS.
15. Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, *26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5
16. Munde, A., & Kaur, J. (2024). Predictive modelling of customer sustainable jewelry purchases using machine learning algorithms. *Procedia Computer Science*, *235*, 683–700. https://doi.org/10.1016/j.procs.2024.04.066
17. Nasseri, M., Falatouri, T., Brandtner, P., & Darbanian, F. (2023). Applying machine learning in retail demand prediction: A comparison of tree-based ensembles and long short-term memory-based deep learning. *Applied Sciences*, *13*(19), Article 11112. https://doi.org/10.3390/app131911112
18. OECD. (2024). *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
19. Patel, A., Sai, S., Daiya, A., Akolekar, H., & Chamola, V. (2025). Blockchain enabled traceability in the jewel supply chain. *Scientific Reports*, *15*, Article 3837. https://doi.org/10.1038/s41598-025-88245-4
20. Perret, J. K., Beyerhaus, C., & Ding, S. (2025). Navigating the high jewelry industry's future: Sustainability and brand diversification. *Businesses*, *5*(4), Article 49. https://doi.org/10.3390/businesses5040049
21. Responsible Jewellery Council. (2024). *Chain of custody standard* (2024 ed.). https://www.responsiblejewellery.com/standards/chain-of-custody/
22. Shaik, R. S., Mohammad, A. F., Gunti, N., & Kalva, S. (2026). Responsible AI in demand forecast of supply chain management. In H. R. Arabnia, L. Deligiannidis, S. Amirian, F. Ghareh Mohammadi, & F. Shenavarmasouleh (Eds.), *AI revolution: Research, ethics and society: AIR-RES 2025* (Communications in Computer and Information Science, Vol. 2723, pp. 372–382). Springer. https://doi.org/10.1007/978-3-032-13056-3_27
23. Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1
24. World Gold Council. (2026a, January 29). *Gold demand trends: Q4 and full year 2025: Jewellery*. https://www.gold.org/goldhub/research/gold-demand-trends/gold-demand-trends-full-year-2025/jewellery
25. World Gold Council. (2026b, July 30). *Gold demand trends: Q2 2026: Jewellery*. https://www.gold.org/goldhub/research/gold-demand-trends/gold-demand-trends-q2-2026/jewellery

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334

26. Zambon, L., Azzimonti, D., & Corani, G. (2026). *End-to-end probabilistic hierarchical forecasting of large hierarchies via probabilistic top-down* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2606.26774

Copyright © ISRG Publishers. All rights Reserved.
DOI: 10.5281/zenodo.23432334