International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 3, May-June 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

LLM-Powered Dynamic Pricing Engine with Causal Inference for Retail and E-Commerce

Sundeep Kumar P[1], Harish Patil[2], Pratibha S [3], Sai Shivani K[4], Anita Patil[5]
[1,2,3,4] UG Student, Dept. of Computer Science and Artificial Intelligence, Ballari Institute of Technology and Management, Ballari, Karnataka, India
[5] Professor, Dept. of Computer Science and Artificial Intelligence, Ballari Institute of Technology and Management, Ballari, Karnataka, India

--************************----------------------------------

Abstract:

Stop relying on predictive machine learning models that find correlations in historical data. Use causal models, like econometric models. Short-term demand forecasting systems fail to distinguish between actual cause and no more than coincidence, which can lead to inadequate pricing measures and a widely known downward price spiral. LLM-DPECI is proposed in this paper, which is an integrated framework which uses causal machine learning, Bayesian Optimization, and retrieval-augmented generation to deliver accurate price decisions which are causally ground and also provide human-readable reports. This study proposes a method that employs the DoWhy and CausalML framework to estimate heterogeneous price elasticities using double ML while explicitly controlling for the confounders, such as competitor pricing, seasonal events, and weather effects. A module for Bayesian Optimization establishes a pseudo-price space using a Gaussian process surrogate to optimize expected revenue under margin constraints. A large language model (GPT/Claude) based RAG pipeline generates easy to understand natural language driven pricing report from complex model outputs for technical actions by non-technical manager and makes them act based on causal insights. Experiments on the Instacart and Walmart M5 public datasets indicate that LLM-DPECI significantly outperforms correlation-based baselines in terms of off-policy demand estimation accuracy and revenue uplift. Users also rated the reports produced by LLM as highly interpretable.This work highlights causal inference as the critical next step beyond predictive ML in commercial pricing systems.

Keywords — **dynamic pricing, causal inference, large language models, RAG, Bayesian Optimization, double machine learning, DoWhy, CausalML, demand forecasting, e-commerce, price elasticity, confounders, RAG.**

--************************----------------------------------

I. INTRODUCTION

This Pricing is one of the most consequential and continuously evolving decisions in retail and e-commerce. Online retailers now have unparalleled access to transaction histories, competitor pricing feeds, weather signals, local event calendars, and detailed customer behavior logs due to the widespread availability of real-time data collection infrastructure. Despite this data abundance, the dominant approach to automated pricing remains predictive in nature—systems learn statistical associations between observable features and historical demand and then extrapolate to set future prices [1].

The fundamental limitation of such predictive systems is their inability to differentiate causation from correlation. A product may appear to sell more on rainy days not because rain causes increased buying intent but because competitors close their physical outlets, reducing alternatives. If a pricing engine fails to identify this underlying mechanism, it may incorrectly raise prices on rainy days expecting sustained higher demand only to face inventory surpluses when competitors adapt. This confusion between correlation and causation is the root-cause of erratic pricing behavior, revenue loss, and the much-documented race-to-the-bottom dynamic in competitive e-commerce environments [2].

Causal inference offers a principled solution. DoWhy [3] and CausalML [4], for example, are based on Pearl's do-calculus and potential outcome theory, offering tools for encoding expert knowledge into causal graphs as well as estimating the causal impact of the price on demand after controlling for other variables. With such a causal approach to pricing, it becomes possible not just to answer "How much demand will be observed at price $12?" but also to reason about counterfactual scenarios, such as "What was the demand expected to be had the price been set at $12, holding everything else constant?" Counterfactual reasoning is crucial for any reliable pricing optimization task.

Alongside that, the recent advent of large language models (LLMs) like GPT-4 and Claude has created a whole new way of delivering complex analysis results. When combined with retrieval-augmented generation workflows [5], an LLM will be able to combine structured output of a predictive model with context information (e.g., competitors' behavior, local events, and weather patterns) into a human-readable narrative about pricing strategy. This bridges the gap between quantitative pricing models and business managers who must act on their recommendations.

This paper makes the following contributions**:**

- We propose **LLM-DPECI**, a modular end-to-end pricing framework that integrates causal inference, Bayesian Optimization, and retrieval-augmented generation (RAG) for LLM-based reporting.
- We demonstrate the application of **Double Machine Learning (DML)** using EconML [6] and DoWhy on public e-commerce datasets to estimate heterogeneous price elasticities.
- We show that Bayesian Optimization, using Gaussian process surrogates over causally estimated demand functions, outperforms both rule-based and reinforcement learning baselines in simulated pricing environments.
- We design and evaluate a RAG pipeline that generates human-readable reports, and we assess interpretability through a structured user study.
- We publicly document the proposed architecture, experimental setup, and evaluation protocol to ensure reproducibility.

The remainder of this paper is organized as follows. Section II reviews related literature. Section III describes the proposed framework in detail. The experimental results are presented in Section IV. Section V discusses applications, challenges, and limitations. Section VI concludes the paper.

II. LITERATURE SURVEY

A. Sharma and Kiciman (2020) DoWhy: An End-to-End Library for Causal Inference

Sharma and Kiciman [3] introduced DoWhy as an open-source Python library that provides a unified, four-step interface for causal inference: model, identify, estimate, and refute. The library synthesizes graphical causal models (structural causal models) with the potential outcomes framework, enabling practitioners to explicitly encode domain assumptions in a causal graph and then automatically derive valid identification strategies. A key differentiator is DoWhy's refutation API, which subjects causal estimates to a battery of adversarial tests—random common cause addition, placebo treatment, and data subset validation—to assess the robustness of identified effects.

DoWhy allows analysts to model the price-demand relationship as a causal graph in pricing contexts, identifying confounders like seasonality, competitor actions, and external shocks, and then estimating the direct causal effect of price on demand using methods from ordinary least squares to doubly-robust estimators. The 2024 extension, DoWhy-GCM [7], further automates causal graph discovery and supports root-cause attribution, allowing the system to precisely explain which drivers were responsible for an observed demand drop. This root-cause capability is central to the explainability goals of the proposed LLM-DPECI framework.

B. Chen et al. (2020) CausalML: Python Package for Causal Machine Learning

Chen et al. [4] from Uber introduced CausalML as a Python library focused on uplift modeling and conditional average treatment effect (CATE) estimation. The CausalML framework offers numerous meta-learners, including the S-, T-, X-, and R-learner models, as well as causal trees and causal random forests approaches. In the price domain, for instance, the treatment may be considered a change in price (such as reducing the price from \$15 to \$12), and the CATE is described by the demand variation among different customers or goods across different customer segments or product categories. This customer-level heterogeneity is essential for value-based pricing strategies that avoid uniform discounting.

The CausalML toolkit also includes evaluation metrics, such as the Area Under the Uplift Curve (AUUC) and Qini coefficient, which allow practitioners to compare the effectiveness of different causal estimators in targeting the most price-sensitive customer segments. This facilitates smarter promotional strategies that maximize revenue uplift while minimizing unnecessary margin sacrifice on customers who would have purchased regardless of a discount.

C. Chernozhukov et al. (2018) – Double/Debiased Machine Learning

Chernozhukov et al. [8] established the theoretical foundation of double machine learning (DML), a semiparametric framework that uses cross-fitting and machine learning residualization to produce unbiased estimates of causal treatment effects in the presence of high-dimensional confounders. DML first estimates and removes the influence of confounders on both the treatment variable (price) and the outcome variable (demand) using separate machine learning models and then estimates the causal effect from the residualized relationship. The key property of DML is that it achieves n^(1/2) convergence rates for causal parameters even when nuisance functions are estimated at slower nonparametric rates.

In retail pricing, DML is particularly powerful because it allows any ML model—random forests, XGBoost, or neural networks—to control for the complex, high-dimensional confounders present in real-world data while still yielding statistically valid confidence intervals for the price elasticity estimate. The DoubleML Python package [9] and EconML [6] implement this framework and are central components of the proposed LLM-DPECI architecture.

D. Stephan et al. (2023) Causal Forecasting for Pricing

Stephan et al. [10] proposed an approach that integrates double machine learning with transformer-based time-series forecasting models to produce causally valid demand forecasts in a pricing context. Their DML Forecaster first deconfounds historical price-demand data using DML residualization, then trains a temporal transformer on the cleaned residuals to forecast demand under counterfactual price levels multiple weeks into the future. Experiments on real-world online retail data demonstrated that the DML Forecaster substantially outperforms conventional forecasting approaches in off-policy evaluation settings—precisely the situation encountered when a retailer wishes to test a new pricing policy without running a live A/B experiment.

This work is highly relevant to LLM-DPECI as it provides both conceptual and empirical validation for combining causal ML with sequence modeling in retail pricing and serves as one of the key baselines against which the proposed system is benchmarked.

E. Zhang et al. (2025) Delta Method for Debiased Price Promotion

Zhang et al. [11] introduced a Delta Method approach for panel-data causal inference in promotional pricing, presented at the KDD 2025 Applied Data Science track. Their method demeans product-level features to mimic fixed-effects regression, isolating within-product price variation and removing time-invariant confounders. Applied to a furniture e-commerce retailer, the approach improved out-of-sample demand prediction accuracy and provided clearer causal estimates of the discount-demand relationship. A live A/B test validated the causal estimates, showing a 3% revenue uplift and a 2% profit gain compared to a predictive ML baseline that ignored the causal structure.

This finding directly supports the central thesis of the proposed system: causally informed pricing decisions yield measurable improvements over correlational approaches, even on the same dataset, because the causal model correctly attributes demand changes to their true drivers rather than confounded associations.

F. Anand et al. (2025) Bayesian Optimization for Dynamic Pricing

Anand et al. [12] formally modeled dynamic pricing as a black-box revenue Optimization problem and applied Gaussian process-based Bayesian Optimization to sequentially propose prices, balancing exploration of uncertain price regions with exploitation of currently profitable regions. Their theoretical analysis establishes sublinear Bayesian regret bounds, and empirical experiments on simulated finite-inventory pricing environments demonstrated that the BO approach outperforms standard reinforcement learning baselines—including Q-learning and policy gradient methods—while requiring significantly fewer domain-specific assumptions about the demand distribution.

The superiority of BO over RL in this setting is attributed to BO's sample efficiency: the Gaussian process surrogate explicitly encodes uncertainty and directs price trials toward informative regions, whereas RL requires far more environment interactions to converge. In the LLM-DPECI framework, the causal demand model from the DML module serves as the objective function for Bayesian Optimization, making the price search both sample-efficient and causally grounded.

G. Lewis et al. (2020) EconML: A machine learning library for estimating heterogeneous treatment effects

Microsoft's EconML library [6], developed under the ALICE (Automated Learning and Intelligence for Causation and Economics project, provides a comprehensive suite of heterogeneous treatment effect estimators. These include DML-based estimators (linear DML, sparse linear DML, and forest DML), doubly-robust learners, causal forests (orthogonal random forests), and instrumental variable estimators. EconML is designed specifically for economic and business applications where heterogeneous causal effects—such as the variation in price sensitivity across customer demographics and product categories—are of central importance.

A. In LLM-DPECI, EconML's DML estimator is used as the primary engine for price elasticity estimation. The per-product or per-segment elasticities estimated by EconML feed directly into the Bayesian Optimization module, which uses them to compute expected revenue as a function of price and then proposes the next price to evaluate. The tree interpreter functionality in EconML further provides feature-level attributions ingested into the RAG pipeline to enrich LLM-generated reports with data-driven explanations.

III. RESEARCH GAP

Despite significant advances in both causal machine learning and large language models, the existing literature addresses these technologies in isolation with respect to dynamic pricing. Traditional pricing models rely on correlational demand forecasting without accounting for causal confounders, leading to unreliable off-policy predictions and susceptibility to the pricing death spiral under competitive conditions [1], [2].

While causal inference tools such as DoWhy, CausalML, and EconML provide rigorous frameworks for estimating price elasticities [3], [4], [6], they do not include a downstream price Optimization module and do not address the challenge of communicating causal insights to non-technical decision-makers. Conversely, LLM-based report generation systems [5] lack integration with structured causal outputs and cannot ground their narratives in verified causal findings, risking hallucinated explanations.

Furthermore, Bayesian Optimization approaches pricing [12] have been evaluated primarily in simulated environments with simplified demand models that ignore causal structure. No existing system simultaneously: (a) estimates causally valid, heterogeneous price elasticities from observational retail data; (b) uses those estimates as input to a principled price Optimization routine; and (c) generates natural language reports that explain the causal drivers of pricing decisions to business stakeholders. This integrated gap is precisely what LLM-DPECI is designed to address.

IV. PROPOSED FRAMEWORK

The LLM-Powered Dynamic Pricing Engine with Causal Inference (LLM-DPECI) is a flexible and end-to-end framework that takes raw retail data and outputs optimal pricing advice along with reports that are comprehensible to humans. The proposed architecture is composed of 4 closely integrated blocks: Data Ingestion & Feature Engineering Block, Causal Inference Engine, Bayesian Optimization Engine, and LLM-powered RAG-based Report Generator.

A. Data Ingestion and Feature Engineering

Data collection involves collecting multivariate time series data from various sources such as historical transaction records (daily sales and price of products), competitor pricing information (gathered using structured web crawling or third-party APIs), weather information (temperatures, rainfall, UV index) associated with specific store locations, local event calendar (such as festivals, sport events, and holidays), and product information.All streams are aligned on a common daily time index per store-product combination.

Feature engineering constructs relative price indices (product price relative to category median), competitor price ratios, weather severity scores, event proximity indicators (days to nearest major event), and lagged demand features. A dedicated confounders DataFrame is constructed containing all variables that plausibly affect both price-setting decisions and demand outcomes, forming the input to the causal inference module.

B. Causal Inference Engine

The causal inference engine is the methodological core of LLM-DPECI. It uses the DoWhy library [3] to specify a causal graph in which price is the treatment, demand is the outcome, and all confounders (seasonality, competitor prices, weather, and events) are explicitly represented as nodes. DoWhy's identify_effect function determines which estimand is non-parametrically identified from the graph, and estimate_effect computes the average treatment effect and, via integration with EconML [6], the conditional average treatment effect (CATE) using a DML estimator.

The DML estimator first regresses price (P) on confounders (Z) using a machine learning model $\hat{f}_P$ to obtain price residuals $\tilde{P} = P - \hat{f}_P(Z)$, and independently regresses demand (D) on confounders to obtain demand residuals $\tilde{D} = D - \hat{f}_D(Z)$. The price elasticity θ is then estimated from the partially linear regression $\tilde{D} = \theta\tilde{P} + \varepsilon$. By construction, θ captures only the direct causal effect of price on demand, purged of confounder influence. CausalML [4] is additionally used to estimate uplift curves across customer segments, identifying which segments exhibit the highest price sensitivity—essential for targeted promotional pricing.

All causal estimates are subjected to DoWhy's refutation tests to validate robustness: the random common cause test adds a random variable as a confounder and verifies that the estimate remains stable; the placebo treatment test replaces price with a random variable and verifies that the estimated effect collapses to zero. These tests provide confidence that estimated elasticities are not artifacts of modeling assumptions.

C. Bayesian Optimization Module

Given the causally estimated demand model—which maps price to expected demand conditional on current contextual features—the system faces a price Optimization problem: find the price p* that maximizes expected profit $\pi(p) = (p - c) \times \hat{D}(p)$, where c is unit cost and $\hat{D}(p)$ is the DML-predicted demand at price p. The action space is continuous or discretized into fine price increments.

Rather than exhaustively evaluating all price candidates, the module employs Bayesian Optimization using a Gaussian process (GP) surrogate model over the profit function [12]. At each iteration, the GP is conditioned on previously evaluated (price, profit) pairs, and a Gaussian process posterior is computed. An acquisition function—specifically expected improvement (EI)—is maximized to select the next price to evaluate. This process continues for a predefined budget of evaluations, after which the price with the highest posterior mean profit is selected.

The BO module additionally enforces margin constraints (minimum acceptable gross margin) and competitor proximity constraints (not pricing below the lowest known competitor price by more than a configurable percentage), preventing the algorithm from inadvertently triggering destructive price wars. This constraint-aware BO formulation directly addresses the race-to-the-bottom problem identified in the literature [2].

D. RAG-Enabled LLM Report Generator

The final component translates quantitative causal and Optimization outputs into natural language pricing strategy reports accessible to business managers. The RAG pipeline [5] is constructed using LangChain with a vector store (FAISS or Pinecone) indexing a curated knowledge base consisting of: (a) DoWhy causal graph outputs and refutation test results, (b) EconML. feature importance attributions, (c) per-product elasticity estimates and their confidence intervals, (d) competitor pricing context retrieved from the data layer, and (e) external contextual snippets, such as upcoming events and weather forecasts.

When The query that will be generated when a pricing strategy is chosen for any particular product is "Give a detailed explanation for the best price strategy of $X.XX for [product] based on the current market situation," and the system will perform a semantic search in the index of knowledge using this structured query. The chunks returned by the search engine will be used to create a prompt to feed into the LLM (GPT-4 or Claude).

V. EXPERIMENTAL SETUP AND RESULTS

A. Datasets

Experiments were conducted on 2 public retail datasets. The Instacart Market Basket dataset [13] contains approximately 3 million orders from over 200,000 users, with detailed product-level purchase histories suitable for demand modeling at the category level. The Walmart M5 Forecasting dataset, referenced in [14], comprises hierarchical daily unit sales for 3,049 products across 10 stores in California, Texas, and Wisconsin, along with a price history file and a calendar file that details SNAP food stamp events, holidays, and sports events. The M5 dataset is particularly valuable as it contains the contextual variables—event type, event name, and price history—that are required to construct realistic confounders for the causal inference engine.

Data preprocessing involves log-transforming demand (to model multiplicative price effects consistent with the economic definition of elasticity), forward-filling missing prices, and constructing relative price features. The training set comprises the first 80% of the time-series, with the remaining 20% set aside for off-policy evaluation, in which model predictions are compared to actual demand under a pricing policy that was not utilised to train the model.

B. Baselines

LLM-DPECI is evaluated against 3 baselines. The naïve baseline sets prices equal to the historical category-average price. The predictive ML baseline trains an XGBoost regressor on historical price-demand data and uses it with a grid search over prices to select the revenue-maximizing price, without causal debiasing. The RL baseline implements a tabular Q-learning agent that interacts with a demand simulator built from historical data and learns a pricing policy through repeated episodes. All baselines use the same data splits and feature sets.

C. Evaluation metrics

Causal validity is assessed via the mean absolute percentage error in off-policy demand prediction—the error when predicting demand under prices not observed in training. This metric directly captures whether the model correctly extrapolates causal effects on new price levels. Revenue uplift is computed as the percentage improvement in simulated revenue over the evaluation period relative to each baseline. For the LLM report generation module, the interpretability of the module is assessed through a structured user study, in which 15 business analysts evaluated reports on a 5-point Likert scale across 3 key dimensions: factual accuracy, clarity, and actionability.

D. Results

On the Walmart M5 dataset, the LLM-DPECI causal inference module achieves an off-policy demand MAPE of 8.3%, compared to 14.7% for the predictive XGBoost baseline—a 43.5% reduction in prediction error. This improvement is consistent with the findings of Stephan et al. [10], who showed that DML-based debiasing systematically improves off-policy demand forecast accuracy for elastic products. The Bayesian Optimization module achieves an average revenue uplift of 4.2% over the predictive ML baseline and 6.8% over the naïve baseline, whereas the RL baseline trails behind the LLM-DPECI by 1.1% due to its poor sample efficiency within the finite evaluation budget.

Regarding the LLM report generation, user study participants gave LLM-DPECI reports a mean score of 4.3/5 for factual accuracy (reflecting the grounding of reports in verified causal outputs), 4.1/5 for clarity, and 4.0/5 for actionability. Free-text comments highlighted the value of explicit causal explanations—for example, reports attributing a 12% demand spike to a local sports event rather than to price reduction—as enabling managers to make more confident pricing decisions. These results confirm that the RAG architecture substantially reduces LLM hallucination risk compared to zero-shot LLM pricing narratives.

VI. APPLICATIONS

The LLM-DPECI framework has broad applicability across retail and e-commerce verticals. The grocery retail system can dynamically adjust prices for perishable items based on factors such as proximity to expiry, weather conditions, and competing promotions, thereby preventing waste through underpricing and margin loss through unnecessary discounts.

The causal inference module in airline and hospitality pricing can separate the impacts of route competition, seasonal travel patterns, and macroeconomic indicators on booking demand, allowing for more stable and value-aligned fare setting than current yield management systems. Surge pricing in ride-hailing apps will enable the detection of real supply and demand mismatches, such as an event at a stadium coming to an end, rather than false connections, making the pricing system more balanced and reliable for consumers.

The LLM-based report generator provides considerable value for small and medium-sized businesses that lack the ability to conduct their own data science in their companies by giving them valuable insights into how to price products in simple terms.

VII. CONCLUSIONS

This paper presents LLM-DPECI, an integrated dynamic pricing framework that combines causal machine learning, Bayesian Optimization, and retrieval-augmented LLM report generation. By replacing correlational demand models with causally debiased estimates from double ML via DoWhy and EconML, the system achieves substantially lower off-policy prediction error and higher revenue uplift than predictive ML and reinforcement learning baselines. The Bayesian Optimization module efficiently searches the price space while enforcing margin and competitor constraints, preventing the race-to-the-bottom pricing behavior endemic to naïve automated pricing systems.

This RAG-based LLM report generation system helps bridge the crucial gap between the quantitative causal output and its practical application to decision-making processes through the creation of interpretative reports on pricing that were rated highly by business analysts in a user study. All these pieces show that causal inference is not just an improvement to theoretical approaches; it is a meaningful step forward in terms of improving production systems, made possible thanks to advanced open-source tools available in such packages as DoWhy, CausalML, EconML, and modern LLMs.

The future direction for the project will be expanding the capabilities of the system to include such techniques as dealing with streaming datasets, web scraping, instrumental variables, and multi-product pricing scenarios.

VIII. REFERENCES

[1] P. K. Kopalle, K. Pauwels, L. Y. Akella, and M. Gangwar, "Dynamic pricing: Definition, implications for managers, and future research directions," J. Retail., vol. 99, pp. 580–593, 2023.

[2] M. Neubert, "A systematic literature review of dynamic pricing strategies," Int. Bus. Res., vol. 15, no. 1, pp. 1–17, 2022.

[3] A. Sharma and E. Kiciman, "DoWhy: An end-to-end library for causal inference," arXiv preprint arXiv:2011.04216, 2020.

[4] H. Chen, T. Harinen, J.-Y. Lee, M. Yung, and Z. Zhao, "CausalML: Python package for causal machine learning," arXiv preprint arXiv:2002.11631, 2020.

[5] P. Lewis, E. Perez, A. Piktus, et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. NeurIPS, 2020, pp. 9459–9474.

[6] R. Keith Pace and P. E. Lewis, "EconML: A machine learning library for estimating heterogeneous treatment effects," Microsoft Research, 2020. [Online]. Available: https://www.microsoft.com/en-us/research/project/econml/.

[7] P. Blöbaum, P. Götz, K. Budhathoki, A. A. Mastakouri, and D. Janzing, "DoWhy-GCM: An extension of DoWhy for causal inference in graphical causal models," J. Mach. Learn. Res., vol. 25, no. 147, pp. 1–7, 2024.

[8] V. Chernozhukov, D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, "Double/debiased machine learning for treatment and structural parameters," Econom. J., vol. 21, no. 1, pp. C1–C68, 2018.

[9] P. Bach, V. Chernozhukand attributionsrz, and M. Spindler, "DoubleML – An object-oriented implementation of double machine learning in Python," J. Mach. Learn. Res., vol. 23, no. 53, pp. 1–6, 2022.

[10] J. Stephan, M. Kunz, the S. Birr, et al., "Causal forecasting for pricing," arXiv preprint arXiv:2312.15282, 2023.

[11] Y. Zhang, et al., "A debiased machine learning framework for optimizing price promotion within e-commerce," in Proc. KDD Applied Data Science Track, 2025.

[12] R. Anand, A. Tran, and L. Zhang, "Bayesian Optimization for dynamic pricing and learning," arXiv preprint arXiv:2510.12447, 2025.

[13] Instacart, "Instacart Market Basket Analysis," Kaggle, 2017. [Online]. Available: https://www.kaggle.com/c/instacart-market-basket-analysis.

[14] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M5 accuracy competition: Results, findings, and conclusions," Int. J. Forecast., vol. 38, no. 4, pp. 1346–1364, 2022.

