



Original Article

User and Store Behavior Analytics for Improving E-Commerce Business Decisions: A Design-Science Behavior-to-Decision Framework

Nitin Patki

Engineering Project Manager at Apple Inc. USA.

Received On: 22/08/2026

Revised On: 18/09/2026

Accepted On: 26/09/2026

Published On: 02/10/2026

Abstract - E-commerce platforms capture fine-grained evidence of how customers discover, search, compare, purchase, return, and seek service for products, while also generating store-level evidence on assortment, price, promotion, inventory, fulfillment, margin, and service conditions. Many organizations, however, lack an integrated method for translating these observations into economically justified and governable decisions. This study develops the Behavior-to-Decision Analytics Framework (BDAF) as a conceptual design-science artifact for user and store behavior analytics in e-commerce. The method synthesizes prior work in clickstream prediction, customer-journey analysis, customer-base modeling, recommender systems, causal inference, online experimentation, and responsible analytics to define a canonical data and journey representation, analytical progression, decision logic, governance controls, and implementation architecture. The resulting BDAF integrates consent-aware event contracts, identity and session management, multi-resolution journeys, store-context features, descriptive and diagnostic analytics, predictive and survival models, recommendation and uplift methods, experimentation, and constrained decision optimization. Its design outputs include a unified data model, method-selection logic, KPI and validation frameworks, a phased enterprise adoption roadmap, and eight testable research propositions. A worked scenario illustrates application across acquisition, search, checkout, retention, merchandising, and fulfillment. The paper reports conceptual design results rather than empirical estimates of business lift; empirical validation is specified as future research.

Keywords - Behavior Analytics, Clickstream Analytics, Customer Journey, E-Commerce, Store Analytics, Purchase Propensity, Customer Lifetime Value, Causal Inference, Recommender Systems, Experimentation, Decision Support.

1. Introduction

Digital commerce produces a detailed behavioral record of how users discover, compare, evaluate, and purchase products. Recommended e-commerce event schemas now cover product-list impressions and selections, product-detail views, cart actions, checkout steps, purchases, promotions, and refunds, enabling analysis at both session and item

levels. The analytical opportunity is therefore broader than reporting traffic and conversion: behavioral data can support intent estimation, experience diagnosis, recommendation, retention, customer-value management, and operational planning.

The principal difficulty is not the absence of data but the absence of an integrated decision architecture. Event data are commonly fragmented across web analytics, commerce platforms, customer data platforms, order systems, campaign tools, and service channels. Metrics may be computed with inconsistent identities, product definitions, attribution rules, and time windows. Predictive models are then trained against these unstable representations and deployed without explicit treatment controls, leading to interventions that may correlate with conversion but do not necessarily cause incremental value.

Recent clickstream studies demonstrate that purchase prediction improves when aggregated session attributes are combined with granular recent actions. Other work shows that anonymous and identified sessions require different feature sets because customer history is unavailable for a substantial share of traffic. Theory-informed sequence models further indicate that latent shopping phases, such as goal-directed search and exploratory browsing, can improve exit-risk prediction and interpretability. These findings motivate a multi-resolution framework rather than a single universal feature representation.

This paper develops the Behavior-to-Decision Analytics Framework (BDAF), a conceptual framework intended to improve business decision support using user- and store-behavior data. It addresses four research questions: RQ1, what canonical representation can integrate customer events, transaction history, and changing store conditions while preserving consent, identity boundaries, and point-in-time correctness? RQ2, which descriptive, predictive, causal, and optimization methods are appropriate for different e-commerce journey stages and business decisions? RQ3, how should predictive outputs be translated into feasible actions and validated for incremental commercial and customer value? RQ4, which governance and operating-model

capabilities moderate the realized value of e-commerce behavior analytics?

The contribution is fourfold. First, the paper defines a canonical behavioral representation and data-quality contract. Second, it introduces an analytical progression from measurement to causal action. Third, it links model outputs to an incremental decision-value objective. Fourth, it provides KPI, model-selection, validation, governance, and implementation frameworks suitable for enterprise e-commerce environments.

In this paper, the term user behavior denotes observable actions and inferred states associated with a shopper or business buyer, whereas store behavior denotes the commercial and operational conditions that shape what the platform exposes and can feasibly offer. Store behavior includes changes in assortment, price, promotion, inventory position, delivery promise, recommendation exposure, content placement, service quality, and return policy. Treating these two domains jointly is important because the same user action may have different meanings under different store conditions. A low product-detail conversion rate, for example, may indicate weak intent, poor content, price disadvantage, insufficient inventory, an unattractive delivery promise, or biased exposure. The proposed framework therefore models behavior as an interaction among user state, store state, treatment, and outcome rather than as a user-only classification problem.

1.1. Related Work

1.1.1. Clickstream and Purchase-Intent Modeling

Clickstream research models user interactions as either engineered session summaries or ordered action sequences. Comparative evidence indicates that hybrid representations can capture both persistent session characteristics and short-horizon behavioral momentum. Studies on anonymous versus identified shoppers show that device, channel, timing, and page depth are essential for anonymous sessions, whereas historical purchases and customer-level features improve modeling for recognized users. Session- and journey-level clustering has also been used to distinguish shopper types and support targeted communication strategies.

Sequence models extend this work by treating browsing behavior as a temporally ordered process. ClickstreamDMM combines attention with latent shopping states and demonstrates that behavioral phase representation can improve exit-risk prediction while retaining managerial interpretability. Conversion prediction remains difficult because purchase events are sparse and browsing traces are noisy, which makes precision-recall tradeoffs, probability calibration, and intervention timing more important than accuracy alone.

1.1.2. Customer Lifecycle and Value Analytics

Session prediction explains immediate intent but not the economic value of the customer relationship. Probabilistic customer-base models such as BG/NBD estimate repeat-

purchase behavior and the probability that a customer remains active in noncontractual settings. Customer lifetime value research further shows that value estimation should jointly consider purchase timing, purchase amount, and defection risk rather than treating longevity as equivalent to profitability. This implies that behavior analytics should connect short-term conversion signals with repeat behavior, margin, service cost, return risk, and future value.

1.1.3. Recommendation, Bias, and Omni channel Behavior

Recommender systems typically learn from implicit feedback such as impressions, clicks, dwell time, and purchases. These signals are affected by exposure, display position, and interface design. Propensity-based learning demonstrates that correcting implicit-feedback bias can improve recommendation models across different item types and layouts. Moreover, omnichannel settings require the integration of online sequences and store transactions through compatible pipelines and feature representations. Consequently, behavioral analytics must record not only user actions but also what the system exposed, ranked, suppressed, or made unavailable.

1.1.4. Causal Decisioning and Governance

Prediction estimates what is likely to happen; decisioning requires an estimate of what will change because of an intervention. Uplift modeling explicitly estimates heterogeneous treatment effects for actions such as coupons, messages, or recommendations and can model effects across impression, click, and conversion stages. Off-policy evaluation provides methods for estimating the value of a new policy from logged interactions, although large action spaces create support and variance challenges. These approaches are necessary to prevent high-propensity users from receiving unnecessary incentives and low-propensity users from receiving ineffective treatments.

Behavior analytics also creates privacy and accountability obligations. GDPR provisions address profiling, direct-marketing objections, information about automated logic, and safeguards for solely automated decisions with significant effects. NIST guidance recommends treating security and privacy as a fabric that spans the full big-data architecture and highlights risks from combining diverse data sources, streaming analytics, and algorithmic discrimination. Differential privacy offers one potential privacy-enhancing technology, but its guarantees require careful interpretation and implementation.

1.1.5. Customer Journey and Process Analytics

Customer-journey research emphasizes that experience develops across interconnected touchpoints rather than within a single session or channel. Journey analytics consequently requires representations that preserve temporal order, transitions, interruptions, channel switches, and post-purchase interactions. Reviews of the literature identify recurring themes around experience, service recovery, customer response, channel integration, and technological disruption. Big-data-based customer-experience management extends this perspective by linking customer signals to

organizational learning and action. These streams support the use of event logs, path analysis, and stage-specific outcomes, but they rarely provide a complete mapping from instrumentation to constrained business decisions.

Process mining offers a complementary lens because it treats observed events as traces of an underlying process rather than forcing them into a predetermined linear funnel. Discovery algorithms can expose dominant and exceptional paths; conformance methods can compare actual journeys with an intended process; enhancement methods can attach time, cost, and outcome measures to the process model. In e-commerce, this is particularly useful for repeated search, cart removal and re-addition, failed payments, split shipments, returns, and service recovery. Event abstraction is still required because raw interface events may be too granular for interpretable process models. BDAF therefore retains both the atomic event sequence and an analytically meaningful journey-state sequence.

1.2. Research Gap and Positioning

The research gap is broader than predictive performance. Existing work is often partitioned into customer-experience analysis, clickstream classification, recommendation, customer value, or experimentation. Store-level conditions are commonly treated as static covariates, even though price, exposure, availability, fulfillment, and service policies are themselves changing decision variables. The combined framework addresses this gap by representing the store as a dynamic decision environment and by linking user outcomes to the actions and constraints that produced them. The literature provides strong methods for individual tasks, but enterprise implementations often lack a unified structure linking event quality, journey representation, model selection, treatment assignment, business economics, and governance. BDAF addresses this gap by defining behavior analytics as a closed-loop decision system rather than a collection of reports and models.

2. Materials and Methods

2.1. Research Design and Evidence Base

This study uses a conceptual design-science approach. The framework is synthesized from peer-reviewed research on clickstream prediction, customer-base analysis, recommendation bias, causal modeling, process and journey analysis, and off-policy evaluation, together with official event-measurement, privacy, and risk-management guidance. The objective is theory building and implementation design; no proprietary dataset is analyzed and no empirical performance claim is made for BDAF itself. The literature mapping is purposive rather than a systematic review or meta-analysis: it is used to identify recurring analytical problems, constructs, and candidate methods relevant to artifact design rather than to estimate pooled effects.

The design process followed four steps: (1) identify recurring analytical entities and failure modes; (2) define a canonical data and journey representation; (3) connect analytical methods to business decisions and evaluation criteria; and (4) formulate propositions that can be tested

using future field experiments or multi-site observational studies. The resulting framework is intentionally platform-neutral and can be implemented with batch, streaming, or hybrid architectures.

2.2. Scope and Unit of Analysis

The unit of analysis depends on the decision. Event-level records support immediate intent and error detection; sessions support conversion and abandonment analysis; user or account histories support retention and lifetime value; item and category panels support merchandising; and store-day or market-week panels support operational and commercial planning. The framework permits these units to coexist but requires an explicit observation window, prediction time, outcome window, treatment definition, and decision owner for every use case. This specification prevents ambiguous labels such as “churn,” “high intent,” or “successful search” from being used without a business time horizon and operational consequence.

2.3. Conceptual Design Procedure

The framework follows design-science principles in which an artifact is constructed to address a relevant organizational problem and is evaluated against explicit objectives. The synthesis proceeded through five activities: problem diagnosis, literature mapping, artifact design, internal consistency assessment, and definition of an empirical validation agenda. Problem diagnosis identified recurring failure modes: inconsistent event semantics, identity overreach, leakage in model training, optimization of proxy metrics, exposure bias, inability to estimate incrementality, and weak ownership of activated decisions. Literature mapping connected these failure modes to methods in information systems, marketing science, machine learning, recommender systems, process mining, and causal inference.

2.4. Artifact Evaluation Criteria

The resulting artifact was assessed using four criteria. Completeness requires coverage from data generation through outcome measurement. Traceability requires every decision to be linked to data, model, policy, and experiment versions. Economic coherence requires model outputs to be translated into contribution margin, cost, or service value rather than accuracy alone. Governance compatibility requires purpose limitation, minimization, explainability proportional to risk, and the ability to suppress or review an action. These criteria are design requirements rather than empirical findings.

Accordingly, the Results section reports the BDAF artifact, its analytical selection logic, validation and governance structures, implementation roadmap, and testable propositions. The worked application scenario is used as an illustrative consistency check rather than as an empirical case study; quantitative effect estimates require future field or observational validation.

3. Results

The design-science process produced BDAF as the primary research artifact. The results below specify its data and journey representation, store-state model, decision logic, implementation architecture, analytical method-selection rules, validation requirements, governance controls, enterprise adoption roadmap, and empirical propositions. These outputs are conceptual design results rather than estimates of realized business lift.

3.1. Behavior-to-Decision Analytics Framework (BDAF)



Fig 1: Behavior-to-Decision Analytics Framework (BDAF)

3.1.1. Behavioral Event Layer

The atomic unit is a behavioral event represented as $e_i = (u, s, t, a, o, c, x)$, where u is a user or pseudonymous identifier, s is a session, t is event time, a is the action, o is the object or item, c is the channel or context, and x is a vector of event parameters. The event vocabulary should cover acquisition, navigation, search, product exposure, selection, product-detail engagement, cart modification, checkout, payment, purchase, cancellation, return, service contact, and campaign exposure. Recommended analytics events provide a useful baseline, but enterprise implementations also require domain-specific events such as inventory unavailability, configuration errors, delivery-promise changes, promotion rejection, and payment decline.

3.1.2. Identity, Consent, and Data-Quality Layer

Behavioral analysis must distinguish anonymous browser or device identifiers, sessions, authenticated accounts, households, and organizational buyers. Identity stitching should be purpose-limited and consent-aware. Every event contract should specify ownership, semantic definition, trigger condition, required parameters, timestamp source, identity scope, retention class, consent requirement, and quality checks. Core controls include duplicate detection, event-order validation, purchase reconciliation,

currency normalization, bot filtering, test-traffic exclusion, and item-master conformance.

A data-quality score $q(e)$ can be attached to each event or session, allowing downstream models to exclude or down-weight unreliable observations. This prevents instrumentation defects from appearing as behavioral changes. Data-quality monitoring should be treated as a production control with service-level objectives for event completeness, latency, uniqueness, schema validity, and transaction reconciliation.

3.1.3. Multi-Resolution Journey Representation

For user u , the journey is the time-ordered sequence $J_u = \langle e_1, e_2, e_n \rangle$. BDAF maintains parallel representations because different decisions operate at different time scales. The event level supports real-time triggers; the session level represents immediate intent; the customer level captures recency, frequency, value, and longitudinal preferences; the cohort level supports comparative diagnosis; and the state-transition level models movement through discovery, browse, consideration, checkout, purchase, repeat, and inactivity states.

$$P(S_{t+1} = j \mid S_t = i, X_t = x) = p_{ij}(x) \quad (1)$$

The conditional transition probability in (1) enables funnel analysis without assuming that all journeys are linear. Repeated search, cart removal, product comparison, channel switching, and delayed purchase can be represented explicitly. Anonymous and identified journeys should be modeled separately when historical features are unavailable or when consent limits linkage.

3.1.4. Feature Layer

The feature layer is organized into five families. Session-intent features include elapsed time, event velocity, page depth, search refinements, product diversity, cart intensity, and recent action sequence. Customer features include recency, frequency, monetary value, tenure, purchase cadence, returns, service contacts, affinity, and estimated activity status. Product features include category, price, margin, availability, promotion, lifecycle, and substitutability. Context features include device, source, geography, time, fulfillment promise, and campaign exposure. Friction features include latency, zero-result search, validation errors, unavailable items, payment failures, and repeated backtracking.

Features must be computed with point-in-time correctness to prevent leakage. For any prediction time t , only information available at or before t may be used. The same feature definitions should serve training, batch scoring, and online inference to reduce training-serving skew.

3.1.5. Analytical Progression

BDAF organizes analytics into five levels. Descriptive analytics measures what occurred; diagnostic analytics identifies where and under what conditions behavior changed; predictive analytics estimates future events or value; causal analytics estimates incremental effects of

actions; and prescriptive analytics selects actions under economic and operational constraints. Higher levels do not replace lower levels. A causal model trained on unreliable funnel events remains unreliable, and a prescriptive policy without guardrail metrics can optimize revenue while degrading margin, customer experience, or fairness.

3.1.6. Decision and Feedback Layer

A business action a is selected from a feasible set $A(x)$ that reflects inventory, channel eligibility, contact policy, budget, and customer permissions. Let $\tau_a(x)$ denote the estimated incremental effect of action a , $m(x)$ the expected

contribution margin, $k_a(x)$ the treatment cost, and $r_a(x)$ a risk or policy penalty. The decision rule is:

$$a^*(x) = \arg \max_{a \in A(x)} [\tau_a(x)m(x) - k_a(x) - \lambda r_a(x)] \quad (2)$$

Equation (2) prevents the system from equating predicted conversion with decision value. For example, a shopper with high natural purchase probability may be suppressed from a discount, while a moderate-propensity shopper with positive incremental response and sufficient margin may receive it. Outcomes, exposures, costs, overrides, and policy decisions are returned to the data layer to support experimentation, drift detection, and accountability.

Table 1: Stage-Based Behavioral KPI Framework

Stage	Behavioral Question	Primary KPI / Formula
Discovery	Are qualified users arriving?	Revenue/session; new-user quality
Browse	Can users find relevant products?	List-to-detail; search success; zero-result rate
Consideration	Is intent strengthening?	PDP-to-cart; cart intensity; comparison depth
Checkout	Where is purchase blocked?	Checkout completion; payment failure
Post-purchase	Is value sustained?	Repeat rate; return-adjusted margin; CLV

3.1.7. Store-State and Commercial-Context Layer

User behavior is generated within a changing store state. Let z_t denote the store context at time t , including price, promotion, available-to-promise inventory, delivery estimate, content version, ranking position, channel, payment options, service capacity, and expected contribution margin. A behavior record is therefore more completely represented as $e_i = (u, s, t, a, o, c, x, z, v)$, where v identifies the experiment or decision-policy version that determined exposure. Without z and v , the analyst cannot distinguish customer preference from what the store made visible, available, or attractive.

Store-state features serve three purposes. First, they improve explanation: a conversion decline can be decomposed into traffic quality, relevance, availability, price, fulfillment, and technical factors. Second, they constrain action: a recommendation should not promote an unavailable item, and an incentive should not be issued when margin is insufficient. Third, they support policy learning: the system can estimate how outcomes vary with exposure, placement, promotion, and service conditions. The framework treats these variables as governed commercial facts with effective timestamps, not as manually copied dashboard attributes.

A concise commercial decomposition is useful for diagnosis. For a period g , expected contribution can be expressed as traffic multiplied by conversion probability, units per order, and unit margin, less fulfillment, return, service, and treatment costs. Each term can be segmented by source, device, market, category, inventory state, and policy version. The decomposition does not prove causality, but it localizes where a business change should be investigated and identifies which intervention can be tested.

$$CM_g = N_g p_g U_g m_g - F_g - R_g - S_g - T_g \quad (3)$$

Where N is qualified traffic, p is conversion probability, U is units per order, m is unit contribution, and F , R , S , and

T represent fulfillment, return, service, and treatment costs. Equation (3) provides a business metric hierarchy that connects behavioral stages to contribution rather than revenue alone.

3.1.8. Reference Implementation Architecture

The reference architecture in Fig. 2 separates collection, governed data products, feature and model services, experimentation, decisioning, and activation. Event schemas are registered and versioned; identity and consent services determine permitted linkage; the event store retains immutable behavioral evidence; and curated data products expose session, customer, item, and store-state views. A point-in-time feature store supplies identical definitions to training and online scoring. Model outputs are passed to a decision service that applies eligibility, inventory, and frequency, budget, and policy rules before activation. Outcomes flow back through the same event contract, creating an auditable learning loop.

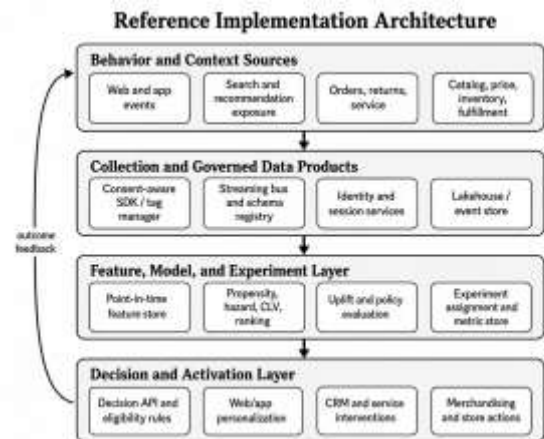


Fig 2: Reference Implementation Architecture for Governed Behavior Analytics and Decision Activation

Table 2: Behavioral Event and Context Taxonomy

Family	Representative Signals	Primary Analytical Use
Discovery	Source, Campaign, Landing, Device	Acquisition Quality And Attribution
Browse/search	Query, Filters, List Exposure, PDP View	Relevance, Findability, Intent
Consideration	Compare, Review, Stock Check, Save	Preference And Friction Diagnosis
Cart/checkout	Add/Remove, Coupons, Payment, Error	Abandonment And Intervention Timing
Post-purchase	Shipment, Return, Support, Review	Retention, Cost-To-Serve, Satisfaction
Store context	Price, Inventory, Placement, SLA, Margin	Commercial Feasibility And Policy Constraints

3.2. Analytical Methods and Model Selection

3.2.1. Descriptive and Diagnostic Methods

The foundation consists of stage-level funnels, path analysis, cohort retention, RFM segmentation, product and channel decomposition, search diagnostics, and error-event joins. Funnel metrics should be evaluated by device, source, and geography, new versus returning status, product category, price band, fulfillment promise, and experiment assignment. Path analysis should preserve loops and repeated actions rather than forcing all users into a single ordered funnel.

Diagnostic analysis should distinguish demand weakness from experience friction. A lower add-to-cart rate may reflect poor relevance, price, inventory, shipping promise, content quality, or technical failure. Joining behavioral events with product, inventory, order, and performance data allows these mechanisms to be separated before predictive modeling begins.

3.2.2. Purchase Propensity and Exit Risk

Purchase propensity is a binary or time-to-event problem. Logistic regression provides a transparent baseline; gradient-boosted trees are strong for heterogeneous tabular features; sequence models are appropriate when action order and timing contain additional signal. The prediction horizon must be defined explicitly, such as purchase within the current session, 24 hours, or seven days. Because conversion is often imbalanced, area under the precision-recall curve, calibration, precision at an intervention budget, and expected profit are more decision-relevant than raw accuracy.

Exit and abandonment are naturally modeled with survival or hazard methods because users who have not purchased are right-censored rather than necessarily lost. Hazard estimates can trigger assistance or reminders when the probability of disengagement rises, while contact policies prevent excessive messaging.

3.2.3. Repeat Purchase, Churn, and CLV

In no contractual commerce, churn is latent. Recency-based rules are easy to operationalize but should be

calibrated by category-specific purchase cycles. BG/NBD-type models estimate future transactions and activity probability, while supervised survival and machine-learning models can incorporate customer, product, and context covariates. CLV should be margin-based and return-adjusted whenever possible. A generic discounted formulation is:

$$CLV_u = \sum_{h=1}^H E[M_{u,h} - C_{u,h}] / (1 + d)^h \quad (4)$$

Where M is expected contribution margin, C is expected service or treatment cost, d is the discount rate, and H is the planning horizon. CLV is most useful for ranking and resource allocation when uncertainty intervals and calibration by value decile are reported.

3.2.4. Recommendation and Next-Best Action

Recommendation models may combine rules, collaborative filtering, content features, sequential representations, and business constraints. Offline ranking metrics are insufficient because logged clicks are exposure-biased and do not fully measure satisfaction or incremental value. The training data should include impression and rank information, and propensity correction should be considered where exposure bias is material. Candidate generation and ranking should enforce inventory, compatibility, margin, policy, diversity, and customer-intent constraints.

3.2.5. Causal Targeting and Policy Optimization

A response model predicts outcomes under observed treatment; an uplift model estimates the difference between potential outcomes. For binary treatment T , conditional uplift is:

$$\tau(x) = E[Y | T = 1, X = x] - E[Y | T = 0, X = x] \quad (5)$$

Randomized experiments provide the strongest identification when feasible. Uplift modeling can then prioritize persuadable users and suppress sure-things, lost causes, or users who may react negatively. Multi-stage e-commerce journeys may require separate treatment effects for impression, click, cart, conversion, and longer-term outcomes. Logged-policy evaluation can support screening before online deployment, but action-space coverage and estimator variance must be assessed.

Table 3: Predictive Analytics Model-Selection Framework

Business problem	Baseline	Stronger method	Optimize
Purchase propensity	Logistic regression	GBDT; hybrid sequence	PR-AUC, calibration, profit
Exit / churn	Tree classifier	Survival / deep survival	Precision at budget; hazard calibration
CLV	RFM / GBM	BG/NBD + value; uncertainty-aware ML	MAE; decline lift; margin
Recommendation	Rules + CF	Sequential / graph ranker	CTR; conversion; diversity; revenue
Offer targeting	Response model	Uplift / causal forest / policy	Incremental profit; uplift

3.2.6. Journey Mining, Segmentation, and Behavioral States

Funnel analysis is most effective when used as a controlled aggregation of journeys rather than as an assumption that every user follows the same path. Transition matrices, Markov chains, sequence clustering, and process-mining models reveal loops, skips, and rework. Useful states may be defined from business semantics—discovery, exploration, comparison, commitment, payment, fulfillment, recovery—or learned as latent states. The state definition should remain stable enough for longitudinal reporting while retaining the event-level trace required for model development.

Behavioral segmentation can combine value, need state, journey pattern, and response to interventions. RFM remains a practical baseline, but its compression can omit timing structure; clumpiness and category cadence improve the representation of repeat behavior. Unsupervised clusters should be evaluated for stability, actionability, and economic separation rather than silhouette score alone. A segment is operationally useful only when a decision owner can specify a differentiated policy and when that policy can be measured.

For real-time decisions, latent-state or sequence models can summarize the evolving session. However, model complexity should be proportional to evidence. Gradient-boosted trees using recent-action counts and order-aware features may outperform sequence models when data are sparse or instrumentation changes frequently. Deep sequence models are more justified when long histories, repeated behavior types, and consistent item representations are available. The framework therefore recommends a champion–challenger approach that compares models at the same latency, calibration, and intervention budget.

3.2.7. Store, Merchandising, and Operations Analytics

Store behavior analytics examines how the digital store allocates exposure and service capacity. Search and recommendation logs should retain the candidate set, ranking scores, displayed positions, availability, and user response. This enables diagnosis of zero-result queries, low-recall retrieval, popularity concentration, position bias, and lost

demand due to unavailable products. Recommender-system research shows that implicit feedback is affected by exposure and other biases; naive learning may therefore reproduce the prior policy rather than customer preference.

Merchandising decisions operate across multiple horizons. Within a session, the platform can reorder results, select content, or change assistance. Daily decisions can adjust campaign allocation, featured products, inventory buffers, and fulfillment routing. Longer-horizon decisions address assortment, supplier commitments, content investment, and return-policy design. The analytical unit should match the control horizon: a millisecond ranking decision cannot depend on slowly refreshed inventory, while a weekly assortment model should not optimize a transient click signal.

Inventory and fulfillment variables are not merely operational constraints; they affect observable behavior. Delivery promise, pickup availability, split-shipment risk, and stock confidence influence click-through, carting, and purchase. Conversely, behavioral demand signals can improve inventory prioritization. A joint decision policy can use expected incremental margin multiplied by fulfillment probability and subtract expected return and service cost. When uncertainty is high, robust or conservative policies are preferable to point estimates that over-allocate scarce inventory.

Price and promotion analytics require particular causal discipline. Customers exposed to a discount are selected by past rules, campaigns, or sales events, so observed response cannot be interpreted as the treatment effect. Randomized tests, geographic or temporal holdouts, or credible quasi-experiments are needed. Uplift models can prioritize customers whose behavior is likely to change because of an offer, while decision rules protect margin and prevent repeated incentive conditioning. Dynamic pricing introduces additional fairness, transparency, and legal concerns; the framework therefore treats it as a high-governance application rather than a default maturity target.

Table 4: User- And Store-Level Decision Applications

Decision domain	Behavioral evidence	Decision output	Primary guardrail
Acquisition	Source, Landing, Early Depth	Budget And Landing Allocation	Incremental Margin
Search/merchandising	Query, Exposure, Click, Stock	Ranking And Assortment Action	Coverage And Availability
Checkout	Cart, Error, Payment, ETA	Assistance Or Suppression	Discount Leakage
Retention	Recency, Returns, Service, Category Cycle	Reactivation Or Contact Policy	Fatigue And Complaints
Store operations	Traffic, Conversion, Margin, SLA	Inventory And Fulfillment Priority	Service Level And Cost

3.2.8. Decision Optimization under Constraints

A model score becomes a decision only after it is combined with action eligibility and expected economics. Let $\hat{y}_a(x,z)$ be the predicted outcome under action a , $\tau_a(x,z)$ the estimated incremental effect, m_a the contribution per

successful outcome, c_a the direct cost, and p_a a risk or customer-experience penalty. The selected action maximizes expected net value subject to inventory, budget, contact, consent, latency, and business-policy constraints.

$$a^*(x,z) = \arg \max_{a \in A(x,z)} [\tau_a(x,z)m_a(x,z) - c_a(x,z) - \lambda \rho_a(x,z)] \quad (6)$$

The feasible set $A(x,z)$ is as important as the objective. It can exclude products without reliable availability, offers outside margin thresholds, contacts that violate frequency limits, and actions requiring unavailable service capacity. Where causal estimates are unavailable, the system may use a conservative rule-based baseline or limit the model to prioritization within an already approved treatment. Logged propensity and policy versions are required for later off-policy evaluation.

3.3. Validation and Measurement Framework

3.3.1. Offline Validation

Validation splits must respect time and identity. Random event-level splits can leak later behavior from the same user into training. Recommended designs include temporal holdouts, user-level separation, campaign or geography holdouts, and backtesting across multiple seasonal windows. Model comparison should include a simple baseline and assess discrimination, calibration, stability, latency, and interpretability.

For ranking and recommendation, evaluation should include recall or normalized discounted cumulative gain, catalog coverage, diversity, novelty, and exposure

concentration. For CLV, absolute error should be complemented by rank correlation and value-decile calibration. For treatment models, uplift curves, Qini coefficients, policy value, and cost-sensitive incremental profit are required.

3.3.2. Online and Causal Validation

Online experiments should predefine a primary business metric, guardrails, sample-ratio checks, minimum detectable effect, and decision horizon. Immediate conversion may conflict with contribution margin, returns, customer support burden, or long-term retention. Persistent holdouts are valuable for measuring cumulative policy effects and detecting treatment fatigue.

3.3.3. Monitoring and Drift

Production monitoring should separate data drift, concept drift, policy drift, and economic drift. Data drift changes feature distributions; concept drift changes the relationship between behavior and outcome; policy drift occurs when upstream ranking or campaign rules alter exposure; and economic drift changes the value of actions through margin, inventory, or fulfillment conditions. Monitoring should include event quality, feature freshness, score distributions, calibration, treatment allocation, override rates, realized uplift, fairness slices, and business constraints.

Table 5: Model and Decision Evaluation Dimensions

Dimension	Representative measures	Failure detected
Predictive	PR-AUC, ROC-AUC, log loss	Weak discrimination
Calibration	Brier score, reliability curve	Misstated probabilities
Economic	Incremental margin, expected profit	High score but low value
Causal	ATE/CATE, uplift, policy value	Correlation mistaken for impact
Operational	Latency, coverage, override rate	Unusable decisions
Governance	Fairness slices, consent, auditability	Policy or trust risk

3.3.4. Business-Value and Decision Validation

Offline model quality is a necessary but insufficient condition for deployment. A decision scorecard should connect four levels: data reliability, model performance, policy compliance, and business outcome. Data reliability includes schema completeness, event timeliness, identity coverage, and store-state freshness. Model performance includes discrimination, calibration, ranking quality, uncertainty, and subgroup behavior. Policy compliance measures eligibility failures, suppression, manual overrides, and explanation availability. Business outcome measures incremental contribution, conversion, retention, service cost, inventory utilization, and customer-experience guardrails.

Online controlled experiments remain the preferred method for measuring product and policy changes when randomization is practical. Experiment design should define the unit of randomization, exposure event, primary metric, guardrails, analysis window, and stopping policy before launch. Sample-ratio mismatch, novelty effects, network interference, repeated testing, and metric peeking can invalidate conclusions. Pre-experiment covariates and

variance-reduction methods such as CUPED can increase sensitivity without changing the treatment assignment.

For decisions that cannot be randomized, the analysis should explicitly state the identification assumptions. Difference-in-differences, interrupted time series, synthetic controls, matching, or instrumental-variable methods may be appropriate, but each depends on conditions that must be assessed. The framework rejects the practice of labeling a pre/post dashboard change as “lift” without a defensible counterfactual.

3.3.5. Portfolio Measurement and Decision Accounting

Behavior analytics programs often overstate value by summing model-specific gains that overlap. A shopper may be influenced by search ranking, recommendation, email, and checkout assistance during the same journey. Decision accounting should therefore assign each activated policy a treatment identifier and use mutually exclusive experiments, factorial designs, or portfolio-level holdouts where feasible. Reported value should distinguish gross observed revenue, modeled attribution, experimentally estimated incremental value, and realized contribution after cost. This hierarchy

supports financially credible prioritization and prevents multiple systems from claiming the same outcome.

3.4. Governance, Privacy, and Responsible Use

Governance is an architectural layer rather than a final compliance review. Purpose limitation should be mapped from each event and feature to approved analytical uses. Data minimization should remove parameters that are not required for the business objective. Access should be role-based, and raw identifiers should be separated from analytical identifiers. Retention periods should differ for raw events, aggregates, training data, model scores, and decision logs.

For automated personalization or targeting, the organization should retain the input version, model version, decision rule, eligible action set, reason codes, treatment, and outcome. Human review is required for high-impact decisions, and users should have appropriate transparency and objection mechanisms where applicable. NIST guidance supports embedding security and privacy controls across data providers, processing, applications, and orchestration rather than relying on perimeter controls alone.

Fairness assessment should be aligned with the decision context. Behavioral variables can proxy for sensitive characteristics even when those fields are removed. Evaluation should therefore compare allocation, benefit, error, and adverse-outcome metrics across legally and ethically relevant groups. Privacy-enhancing methods such as aggregation, clean rooms, federated analysis, and differential privacy may reduce exposure, but each changes utility, complexity, or interpretability and must be evaluated against the specific risk.

3.4.1. Data Protection and Identity Boundaries

Consent and identity must be resolved before features are computed, not after a model is trained. Anonymous session analysis can often support experience improvement without persistent cross-device linkage. Authenticated customer histories may be appropriate for account service, repeat-purchase analysis, and preferences where the purpose and retention period are defined. Household or organizational-account aggregation requires additional controls because the inferred unit may include multiple people. The identity graph should retain source, confidence, purpose, and effective dates, and downstream data products should expose only the level required for the approved use.

3.4.2. Model Risk, Fairness, and Human Oversight

Risk classification should consider both the model and the action. A purchase-propensity model used only for aggregate planning has lower individual impact than the same model used to vary price or suppress service. High-impact policies require stronger documentation, subgroup testing, uncertainty handling, reason codes, approval, and appeal or human review. The NIST AI Risk Management Framework provides a useful govern–map–measure–manage structure for this operating model, while the NIST Privacy Framework supports privacy-risk identification and control.

3.5. Enterprise Implementation Roadmap

Implementation should proceed in maturity stages because sophisticated models cannot compensate for unstable events or unclear decision ownership. Each phase should deliver operational value while creating reusable assets for the next phase.

Table 6: Phased Implementation Roadmap

Phase	Horizon	Primary deliverables	Success criterion
Foundation	0–3 mo.	Event taxonomy, identity, QA, core funnel	Trusted stage measurement
Optimization	3–6 mo.	Search/PDP/cart/checkout diagnostics; tests	Measured funnel and margin lift
Predictive	6–9 mo.	Intent, hazard, repeat, CLV scoring	Better prioritization and calibration
Causal	9–12 mo.	Uplift targeting; persistent holdouts	Incremental profit with guardrails
Enterprise	12+ mo.	Omnichannel features; real-time decisioning	Consistent cross-channel activation

The operating model should assign ownership for event semantics, data products, models, experiments, business policies, and privacy controls. A cross-functional decision council can approve high-impact use cases, define guardrails, and resolve conflicts between local metrics and enterprise objectives. Model deployment should be integrated with product-release processes, observability, incident response, and rollback mechanisms.

3.5.1. Target Operating Model and Lifecycle

An enterprise implementation requires product ownership for decisions, not only technical ownership for models. The decision owner defines the action, eligibility, objective, guardrails, and acceptable failure modes. Data-product owners govern event and feature definitions. Model owners are accountable for training, validation, monitoring, and documentation. Experiment owners protect

randomization and metric integrity. Privacy, legal, security, merchandising, finance, and customer-service stakeholders participate according to the action risk.

The lifecycle begins with a decision canvas that states the business problem, unit of action, eligible population, intervention, comparator, outcome window, economic equation, and constraints. Only then should the team select features and models. The release process should include shadow scoring, replay tests against historical logs, policy simulation, limited exposure, experiment launch, guardrail review, and rollback criteria. Models and business rules must be versioned together because a change in either can change realized outcomes.

Production monitoring should include event-contract tests, feature freshness, model latency, calibration drift,

treatment allocation, inventory conflicts, customer-contact frequency, and incremental outcome estimates. Monitoring is tied to an action playbook: pause activation on consent or schema failure; fall back to rules on model failure; retrain on sustained concept drift; and revisit the objective when economic relationships change. This makes the analytical system operationally resilient rather than merely statistically accurate.

3.6. Research Propositions

The conceptual framework produces the following propositions for empirical testing.

- P1: Models using combined session summaries, recent event sequences, and customer history will produce higher decision value than models using any single representation, subject to latency and data-availability constraints.
- P2: Stage-specific behavioral models will provide better calibration and managerial interpretability than a single end-to-end conversion model when journey states have distinct mechanisms.
- P3: Requiring causal validation before incentive deployment will improve incremental margin by reducing unnecessary treatment of users who would purchase without intervention.
- P4: Event-quality and identity-governance maturity will moderate the relationship between analytical sophistication and realized business impact.
- P5: Continuous monitoring of exposure, drift, and treatment effects will reduce performance degradation caused by feedback loops and nonstationary merchandising or marketing policies.

These propositions can be tested using randomized field experiments, stepped-wedge deployments across markets, synthetic-control designs, or longitudinal comparisons of business units adopting different levels of BDAF maturity. Outcome variables should include incremental contribution margin, customer retention, experience quality, operational cost, and governance incidents.

- P6: Incorporating time-varying store context and exposure into behavioral models will improve calibration and reduce false managerial attribution compared with user-only models.
- P7: Decision policies optimized for incremental contribution under inventory and contact constraints will outperform policies optimized for conversion probability alone.
- P8: Organizations that maintain a versioned decision registry linking treatment, model, rule, and outcome will achieve more reliable experimentation and faster incident resolution than organizations that monitor models independently.

4. Discussion

4.1. Interpretation and Contributions

BDAF reframes behavior analytics from observation to accountable intervention. Its central implication is that data, modeling, and activation cannot be designed independently.

A model score has no business meaning until the action, counterfactual, cost, constraints, and feedback horizon are defined. Similarly, a dashboard metric has limited value when event quality, identity scope, and decision ownership are ambiguous.

The framework also clarifies when complex models are justified. Tree-based models may outperform deep sequence models when features are tabular, labels are stable, and latency is constrained. Sequence or latent-state models become more useful when order, timing, and browsing phase carry incremental information. Probabilistic customer-base models remain valuable when repeat-purchase behavior and latent inactivity are central [8], while causal and policy-learning methods are required when the objective is to choose treatments rather than predict outcomes.

For managers, the framework provides a portfolio view of use cases. Conversion optimization, search diagnostics, churn prevention, CLV allocation, recommendation, promotion targeting, and service interventions share the same event and identity foundation. Reusing governed behavioral data products lowers implementation cost and allows metrics to remain comparable across functions. For researchers, BDAF identifies a need for studies that measure end-to-end decision value rather than isolated offline model gains.

4.2. Worked Application Scenario

Consider a hypothetical multi-category retailer that observes declining mobile contribution despite stable traffic. The descriptive layer shows that product-list click-through is unchanged, but search-to-product-detail transition and checkout completion have fallen. Store-state joins reveal that a subset of high-demand items has low available-to-promise inventory, delivery estimates have lengthened in two markets, and a recently deployed ranking policy continues to promote those items. Error logs also show an increase in payment retries for one wallet type. A user-only propensity model would identify lower conversion, but it would not identify the controllable causes.

Under BDAF, the retailer first validates the event and store-state contracts, then constructs stage-specific cohorts by device, market, query class, inventory state, and ranking-policy version. Process analysis shows additional loops between search results and product details for unavailable items. A diagnostic decomposition attributes the contribution decline to three candidate mechanisms: exposure of low-availability items, fulfillment promise deterioration, and payment friction. The decision council assigns separate interventions because each mechanism has a different owner and risk profile.

For search, the candidate-ranking policy is constrained by availability confidence and expected margin. A randomized experiment compares the incumbent ranker with an availability-aware ranker while monitoring relevance, catalog coverage, cancellation, and margin. For checkout, a separate experiment changes wallet error recovery and message sequencing. For customers who abandoned high-

value carts, an uplift model prioritizes assistance, but discounts are excluded when the estimated purchase probability is already high or margin is insufficient. Each treatment is logged with its policy version so that outcomes are not double counted.

The scenario illustrates three distinctions. First, prediction identifies where outcomes are likely, whereas diagnosis identifies plausible mechanisms. Second, store conditions determine both the meaning of behavior and the actions that are feasible. Third, business value is established through incremental contribution and guardrails rather than through an increase in clicks or an offline F1 score. The example is intentionally illustrative; empirical effect sizes must be obtained from a real deployment.

4.3. Managerial Implications and Use-Case Prioritization

Managers should prioritize behavior-analytics use cases by combining decision frequency, addressable economic value, data readiness, action feasibility, and governance risk. High-frequency decisions such as search ranking or checkout error recovery can create substantial cumulative value, but they require reliable low-latency features and rapid rollback. Lower-frequency decisions such as assortment revision or retention-budget allocation can tolerate batch processing but require longer outcome windows. A transparent scoring method prevents the portfolio from being dominated by technically attractive models that lack a controllable business action.

The recommended sequence begins with measurement contracts and diagnostic decisions that have clear owners. Examples include identifying zero-result search demand, separating unavailable-product abandonment from weak intent, and measuring payment failure by method and software version. Predictive models should follow only after the organization can define the intervention and comparator. Causal targeting and constrained policy optimization are later-stage capabilities because they depend on treatment logging, randomization or defensible identification, and an economic objective agreed with finance and business teams.

The framework also changes how performance is discussed with leadership. Rather than reporting model accuracy as the primary achievement, teams should report the percentage of eligible decisions covered, the share of actions that pass policy constraints, measured incremental contribution, customer and operational guardrails, and uncertainty. This creates a common language across analytics, product, merchandising, operations, finance, privacy, and customer service. It also makes explicit when a simple rule or workflow change is preferable to machine learning.

Finally, user and store analytics should be governed as one decision system. Marketing cannot optimize response independently of inventory; merchandising cannot optimize exposure without considering relevance and customer fatigue; and fulfillment cannot prioritize speed without considering margin and service commitments. A cross-

functional decision council is therefore most useful when it approves objectives and trade-offs, not when it attempts to review every model parameter.

4.4. Limitations and Future Research

The paper is conceptual and does not provide an empirical estimate of business lift. The framework must be adapted to product category, purchase cycle, market, device mix, identity availability, and regulatory jurisdiction. B2B commerce may require account, role, quote, approval, and contract-price events that differ from consumer retail. Marketplace environments also introduce seller, fulfillment, and advertising interactions.

Future research should validate the propositions on multi-category datasets and compare centralized versus federated implementations. Additional work is needed on multimodal behavior, conversational commerce, agent-mediated shopping, long-horizon effects of personalization, and the interaction between recommendation policies and user preference formation. Omnichannel research should examine how store and online signals can be integrated without creating disproportionate privacy or identity risk.

A further limitation is that the framework assumes that relevant exposures and commercial conditions can be logged with sufficient fidelity. In practice, third-party marketplaces, advertising platforms, privacy restrictions, offline interactions, and vendor systems may obscure the candidate set or treatment mechanism. Missing-not-at-random exposure can limit causal interpretation. Future research should compare privacy-preserving aggregation, federated learning, clean-room analysis, and on-device personalization for different decision classes. It should also investigate whether generative and conversational interfaces require new journey states, outcome measures, and disclosure controls.

Empirical validation should use multiple categories because purchase frequency, margin, return behavior, and decision latency differ substantially across grocery, fashion, and electronics, subscription, and B2B commerce. Useful designs include field experiments, staggered market deployment, interrupted time series with validated controls, and simulation calibrated to historical policies. Research should report both predictive quality and realized policy value, including negative outcomes such as promotion dependency, reduced assortment diversity, unfair exposure, service burden, and customer complaints.

4.5. Threats to Validity

The conceptual synthesis is subject to selection and interpretation bias because it integrates literature from several disciplines rather than following a narrowly bounded meta-analysis. Construct validity is also a concern: terms such as intent, engagement, friction, churn, and loyalty are operationalized differently across categories and organizations. The framework mitigates this risk by requiring explicit observation and outcome windows, but empirical studies must still validate whether the chosen measures represent the intended constructs.

Internal validity can be threatened by unlogged exposure, policy changes, interference among treatments, and endogenous store conditions. For example, inventory scarcity may simultaneously change ranking, delivery promise, and conversion, making a single-variable explanation misleading. External validity is limited because

results from one category, market, or identity regime may not transfer to another. Organizational capability is an additional moderator: the same analytical model may create different value depending on decision rights, deployment latency, experimentation maturity, and the willingness to discontinue ineffective policies.

Table 7: Empirical Research Agenda

Research theme	Suggested design	Primary evidence
Cross-category validity	multi-category field experiments	incremental margin, calibration, guardrails
Identity and privacy	federated or aggregated comparison	utility, privacy risk, coverage
Policy feedback loops	longitudinal policy-switch design	diversity, concentration, drift
Human-AI decisioning	staged deployment with override logging	value, override quality, accountability
Omnichannel behavior	linked online/offline natural experiment	channel lift, substitution, service cost

Table VII summarizes empirical designs that can test the framework across technical, economic, privacy, and organizational dimensions.

5. Conclusion

This study develops BDAF as a design-science framework for converting e-commerce behavioral evidence into governed business decisions. BDAF integrates behavioral instrumentation, consent-aware identity, multi-resolution journeys, time-varying store context, feature management, analytical progression, causal decisioning, and continuous feedback. It connects immediate purchase intent with longer-horizon customer value and explicitly separates predictive likelihood from treatment effect. The framework therefore provides a coherent artifact for enterprise implementation and a testable research agenda for examining how analytical maturity, data governance, store context, and causal validation influence decision quality.

By extending behavior analytics from user traces to the interaction among customer state, store state, treatment, and outcome, BDAF provides a structured basis for decision support without prescribing a single algorithm. Its contribution is the specification of data contracts, analytical selection logic, economic objectives, governance boundaries, validation requirements, and feedback evidence needed to select and operate models responsibly. Because this paper is conceptual, the framework should be interpreted as an original design artifact and set of empirically testable propositions rather than as evidence of realized business lift.

References

- [1] Google, “Measure ecommerce,” Google Analytics Developer Documentation, 2026.
- [2] Google, “Recommended events by business vertical: Retail and ecommerce,” Google Analytics Developer Documentation, 2026.
- [3] A. A. Tokuç and T. Dağ, “Predicting user purchases from clickstream data: A comparative analysis of clickstream data representations and machine learning models,” *IEEE Access*, vol. 13, pp. 43796–43817, 2025, doi: 10.1109/ACCESS.2025.3548267.
- [4] M. Hendriksen, E. Kuiper, P. Nauts, S. Schelter, and M. de Rijke, “Analyzing and predicting purchase intent in e-commerce: Anonymous vs. identified customers,” in *Proc. SIGIR Workshop on eCommerce*, 2020.
- [5] Y. Özyurt, T. Hatt, C. Zhang, and S. Feuerriegel, “A deep Markov model for clickstream analytics in online shopping,” in *Proc. ACM Web Conf.*, 2022, pp. 3071–3081, doi: 10.1145/3485447.3512027.
- [6] S. Roychowdhury, E. Alareqi, and W. Li, “OPAM: Online purchasing-behavior analysis using machine learning,” in *Proc. Int. Joint Conf. Neural Networks*, 2021.
- [7] L. Bigon, G. Cassani, C. Greco, L. Lacasa, M. Pavoni, A. Polonioli, and J. Tagliabue, “Prediction is very hard, especially about conversion: Predicting user purchases from clickstream data in fashion e-commerce,” *arXiv:1907.00400*, 2019.
- [8] P. S. Fader, B. G. S. Hardie, and K. L. Lee, “Counting your customers’ the easy way: An alternative to the Pareto/NBD model,” *Marketing Science*, vol. 24, no. 2, pp. 275–284, 2005.
- [9] S. Borle, S. S. Singh, and D. C. Jain, “Customer lifetime value measurement,” *Management Science*, vol. 54, no. 1, pp. 100–112, 2008, doi: 10.1287/mnsc.1070.0746.
- [10] Z. Qin, S. J. Chen, D. Metzler, Y. Noh, J. Qin, and X. Wang, “Attribute-based propensity for unbiased learning in recommender systems: Algorithm and case studies,” in *Proc. 26th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2020.
- [11] L. Ma et al., “Improving sequential recommender systems with online and in-store user behavior,” in *Proc. IEEE Int. Conf. Big Data*, 2024.
- [12] Y. Huang et al., “Entire chain uplift modeling with context-enhanced learning for intelligent marketing,” in *Companion Proc. ACM Web Conf.*, 2024, pp. 226–234.
- [13] Y. Saito, Q. Ren, and T. Joachims, “Off-policy evaluation for large action spaces via conjunct effect modeling,” in *Proc. 40th Int. Conf. Machine Learning*, vol. 202, 2023.
- [14] European Parliament and Council, “Regulation (EU) 2016/679 (General Data Protection Regulation),” *Official Journal of the European Union*, 2016.
- [15] NIST, “NIST Big Data Interoperability Framework: Volume 4, Security and Privacy,” NIST SP 1500-4r2, 2019, doi: 10.6028/NIST.SP.1500-4r2.
- [16] J. P. Near et al., “Guidelines for evaluating differential privacy guarantees,” NIST SP 800-226, 2025.

- [17] Y. Su, L. Wang, M. Santacatterina, and T. Joachims, "Continuous adaptive blending for policy evaluation and learning," in Proc. Int. Conf. Machine Learning, 2019, pp. 6005–6014.
- [18] S. Mittameedi, "Customer experience in e-commerce: A systematic review of metrics, models, and the role of AI," IEEE Access, 2025.
- [19] R. Kohavi and F. Provost, "Applications of data mining to electronic commerce," Data Mining and Knowledge Discovery, vol. 5, pp. 5–10, 2001, doi: 10.1023/A:1009840925866.
- [20] S. Akter and S. F. Wamba, "Big data analytics in e-commerce: A systematic review and agenda for future research," Electronic Markets, vol. 26, pp. 173–194, 2016, doi: 10.1007/s12525-016-0219-0.
- [21] K. N. Lemon and P. C. Verhoef, "Understanding customer experience throughout the customer journey," Journal of Marketing, vol. 80, no. 6, pp. 69–96, 2016, doi: 10.1509/jm.15.0420.
- [22] M. Holmlund, Y. Van Vaerenbergh, R. Ciuchita, A. Ravald, P. Sarantopoulos, F. Villarroel Ordenes, and M. Zaki, "Customer experience management in the age of big data analytics: A strategic framework," Journal of Business Research, vol. 116, pp. 356–365, 2020, doi: 10.1016/j.jbusres.2020.01.022.
- [23] Y. Tueanrat, S. Papagiannidis, and E. Alamanos, "A review of the customer journey literature," Journal of Business Research, vol. 125, pp. 336–353, 2021.
- [24] W. M. P. van der Aalst et al., "Process mining manifesto," in Business Process Management Workshops, Lecture Notes in Business Information Processing, vol. 99, pp. 169–194, 2012, doi: 10.1007/978-3-642-28108-2_19.
- [25] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, and X. He, "Bias and debias in recommender system: A survey and future directions," ACM Transactions on Information Systems, vol. 41, no. 2, Art. no. 31, 2023, doi: 10.1145/3564284.