Peer reviewed Journals
seron press

RESEARCH

Machine Legibility and Attention Retention in Agent Mediated Commerce Visibility Across Structured Product Pages

Karim Fawzy[1] and Tamer Hegazy[2]

[1]Department of Business Administration, Delta Higher Institute for Management and Accounting, Gamal Abdel Nasser Street, Mansoura, Dakahlia Governorate, Egypt
[2]Department of Commerce and Business Information Systems, Misr Higher Institute for Commerce and Computers, Ahmed Orabi Street, Mansoura, Dakahlia Governorate, Egypt

Abstract

Digital commerce is increasingly shaped by two simultaneous filters. Human shoppers still inspect product pages under limited attention, but autonomous shopping agents and AI answer systems now evaluate the same pages through structured data, evidence cues, and machine-readable comparison fields. This paper reports a controlled empirical study of how product-page structure influences agent-mediated visibility and short-window human attention retention. The study uses a synthetic but fully specified experimental dataset containing 18,240 query-product-page observations generated from 1,520 product pages across six retail categories, three page-structure conditions, four agent evaluators, and repeated human exposure logs. The main outcomes are agent selection, source citation, human click-through, attention half-life, and a composite commerce visibility index. The modeling strategy combines grouped train-test splits, logistic regression, gradient boosted trees, ridge regression, exponential attention-decay estimation, ablation tests, permutation importance, and robustness checks with category and query fixed effects. Results show that structured product pages increase agent selection by 8.7 percentage points relative to narrative pages and increase citation probability by 6.2 percentage points. Evidence-rich structured pages also retain human attention for longer, with the fitted median attention half-life rising from 2.81 seconds to 3.46 seconds. Machine-legibility signals remain predictive after controlling for price, rating, review volume, brand reputation, delivery time, and category. The results suggest that product-page visibility in agent-mediated commerce depends on both consumer-facing value and machine-readable evidence organization, with neither channel fully substituting for the other.

1. Introduction

Digital commerce has usually been studied as a search and choice environment in which consumers inspect information, compare alternatives, and select products under cognitive and time constraints. That view remains useful, but it is incomplete for a market in which product discovery is increasingly mediated by automated agents, AI search summaries, recommendation surfaces, and structured comparison systems. A product page may now be evaluated by a human shopper scrolling quickly through a mobile interface and by an autonomous agent parsing fields, schema, review summaries, return policies, delivery claims, and product specifications. These two audiences are not identical. Humans are sensitive to visual salience, attention fatigue, trust, and presentation fluency. Agents are sensitive to evidence availability, field completeness, semantic match, and machine-readable structure. A product page can therefore become commercially visible through several different mechanisms: it may hold a consumer's attention, be selected by an agent, be cited by an AI answer system, or be ranked as a candidate in an automated shopping workflow.

The shift matters because the practical meaning of visibility changes when the immediate evaluator is partly algorithmic. In a conventional search environment, visibility is often approximated by position, impression, and click. Bakos [1] described electronic markets as systems that reduce search costs by making product information easier to obtain, while Alba et al. [2] argued that interactive shopping changes the balance among consumer search, retailer presentation, and product comparison. Those arguments assumed that information access would improve human comparison. In agent-mediated commerce, information access also changes the ability of nonhuman systems to interpret and rank offers. A page that communicates adequately to a person may be less useful to an agent if the product attributes are ambiguous, the specifications are embedded in prose, and the evidence needed for comparison is difficult to extract.

The rise of automated comparison also reframes

older problems in recommender systems and decision aids. Adomavicius and Tuzhilin [3] described recommender systems as tools that help users manage information overload by predicting relevant items, while Burke [4] showed that recommendation can be implemented through knowledge-based, collaborative, content-based, and hybrid procedures. Those systems usually support human decision-making, but the new commerce setting gives agents a more active role. The agent may narrow a consideration set, rank products, summarize alternatives, or select one item for a user-defined task. When that happens, the brand is competing for both the shopper's attention and the agent's confidence that the product is relevant, comparable, and adequately evidenced.

This paper studies the interaction between machine legibility and human attention retention in agent-mediated commerce visibility. Machine legibility refers to the degree to which a product page presents product information in forms that are easy for automated systems to parse, compare, and use. It includes structured product fields, schema markup, specification tables, clear return-policy statements, normalized price and delivery information, evidence snippets, and completeness of product descriptions. Human attention retention refers to the persistence of observable attention during a short exposure window. It is not treated as a direct measure of mental attention; rather, it is operationalized through event-level behavioral traces such as dwell time, scroll stability, hover or touch events, visible product-region exposure, and click-through.

The empirical setting is a controlled product-page experiment with 18,240 query-product-page observations. Each observation combines a query, a product, a page-structure condition, an agent evaluator, and a human exposure record. Products appear across apparel, electronics, home goods, skincare, food appliances, and educational supplies. Each product is represented in three page conditions: a narrative page, a moderately structured page, and an evidence-rich structured page. The underlying product attributes are held constant across conditions so that the study can separate page organization from product quality. Agent evaluators are asked to choose or cite products under comparison, recommendation, and evidence-seeking tasks. Human exposure logs are generated from timed page-inspection sessions in which participants view candidate product pages and may click or save a product. The resulting dataset allows the same product to be evaluated under both human and machine-facing outcome definitions.

The study has three empirical aims. The first aim is to estimate whether machine-legible page structure predicts agent selection and citation after controlling for price, rating, review volume, brand reputation, delivery time, return policy, and category. The second aim is to estimate whether the same page structure improves short-window attention retention among human viewers. The third aim is to examine whether agent visibility and human attention retention combine into a composite commerce visibility index, or whether the two channels respond differently enough that they should be modeled separately. The paper does not claim that all real agents use the same decision rule or that all retail categories respond equally to structured content. The empirical claim is narrower: under controlled conditions, machine-readable page organization contributes measurable signal beyond conventional product attractiveness variables.

The paper is positioned between several research traditions. Information economics and online retail research explain why searchable product attributes affect consumer choice. Online review research explains why rating, volume, sentiment, and reputation alter trust and sales. Algorithmic gatekeeping research explains why automated ranking and selection systems can reshape visibility. Attention and advertising research explain why short exposure windows matter for human response. Generative search research explains why evidence, citation likelihood, and source selection are becoming important visibility outcomes. These traditions have often been studied separately. In this paper they are joined within one empirical design because the product page is now a shared object for human and machine evaluation.

The contribution is empirical and methodological. It provides a measurement framework for observing commerce visibility across two channels, defines a reproducible set of product-page structure variables, estimates interpretable and predictive models, and reports a set of robustness checks that distinguish machine-legibility effects from ordinary product-quality effects. The results show that structured product pages can improve agent selection and citation while also modestly improving human attention retention. The gains are not uniform. The largest effects appear for high-consideration products, comparison-oriented queries, and pages where structure is paired with evidence rather than mere formatting. The evidence therefore supports a cautious interpretation: machine legibility is neither a substitute for product value nor a cosmetic page feature, but it is a measurable part of the visibility process in agent-mediated commerce.

2. Prior Research and Conceptual Framing

Research on online markets has long emphasized that digital environments reduce some information costs while creating new forms of comparison complexity. Bakos [1] treated electronic marketplaces as mechanisms that lower buyer search costs and intensify price and attribute comparison. Alba et al. [2] developed a related argument that interactive home shopping changes the way consumers encounter product information, because the environment can display, sort, and personalize al-

ternatives in ways that are difficult in physical retail. These early arguments remain relevant, but they do not fully describe a setting in which comparison is delegated to an autonomous agent. A shopper who instructs an agent to find a durable backpack under a budget constraint may not inspect every candidate. The agent's ranking becomes part of the market interface.

Online review and reputation research provides a second foundation. Resnick et al. [5] argued that reputation systems can support trust in electronic markets by aggregating past-user feedback. Dellarocas [6] examined online feedback mechanisms as a form of digitized word of mouth, while Chevalier and Mayzlin [7] showed that online reviews can affect book sales. Later work by Hu, Pavlou, and Zhang [8] showed that the distribution and presentation of reviews can shape perceived product quality in ways that are not captured by a simple mean rating. These findings matter because autonomous agents often parse reviews, ratings, and reputation signals as structured evidence. Human shoppers may be influenced by the same signals, but the processing route differs. A human may skim the first few reviews and respond to tone or vividness, while an agent may use average rating, volume, sentiment, and review recency as features.

The recommender-system literature helps explain how product information becomes operational for automated choice. Adomavicius and Tuzhilin [3] framed recommendation as an information-filtering problem, and Ricci, Rokach, and Shapira [9] described recommendation systems as tools that combine data, preference inference, and ranking. Burke [4] distinguished knowledge-based recommendation from other recommendation families, which is useful here because agent-mediated shopping often depends on explicit product constraints rather than only collaborative behavior. If a user asks for a low-noise blender with dishwasher-safe parts and a strong return policy, the agent must identify whether each candidate page contains enough reliable evidence to support those constraints. This makes product-page completeness and structured attributes directly relevant to selection.

Algorithmic visibility research adds a further concern: automated systems do not merely retrieve information; they can determine what becomes practically visible. Gillespie [10] described algorithms as systems that shape public relevance, and Introna and Nissenbaum [11] argued that search engines can privilege some sources through technical ranking processes that appear neutral. Pasquale [12] connected algorithmic opacity to informational power, while Barocas and Selbst [13] showed how data-driven decision systems can produce uneven outcomes even without explicit discriminatory intent. In a commerce setting, agentic product selection creates a similar visibility issue. A brand whose product is not parsed, ranked, cited, or selected by an agent may be excluded from the shopper's practical consideration set before the shopper evaluates it directly.

The question of machine legibility is also connected to structured data on the web. Guha, Brickley, and Macbeth [14] described Schema.org as a vocabulary for structured data exchange at web scale. Structured product data is important because it separates attributes into recognizable fields that can be retrieved and compared. A product page may state in prose that shipping is free on orders above a threshold, but an agent may handle the same information more reliably when shipping cost, delivery time, availability, and return policy are represented in consistent fields. This does not mean structured data is always sufficient. A page can be technically structured but substantively thin. The empirical issue is whether structured and evidence-rich presentation adds predictive signal beyond ordinary product appeal.

Generative retrieval research has made this issue more visible. Karpukhin et al. [15] showed that dense passage retrieval can support open-domain question answering by retrieving relevant evidence passages, and Lewis et al. [16] developed retrieval-augmented generation as a framework in which generated answers are conditioned on retrieved documents. These approaches show why evidence selection and citation are not identical to classic ranked search. A source may be selected because it supplies evidence that fits the query and can be summarized. In commerce, this means a product page can gain visibility not only through ranking position but also through its usefulness as evidence in an answer or recommendation. Török [17] reports a controlled modeling study of generative engine optimization in which evidence-related content signals and structured source attributes predicted citation outcomes, which motivates the citation component of the present design.

Attention research helps describe the human side of the same problem. Pieters and Wedel [18] showed that visual attention to advertising is shaped by the interplay of brand, pictorial, and text elements, while Teixeira, Wedel, and Pieters [19] studied how video advertising can lose attention across time. In digital environments, user exposure is often short and fragmented. A page can be technically available but not inspected deeply. Goldfarb and Tucker [20] showed that privacy regulation affected the performance of online display advertising, illustrating that digital advertising effectiveness depends on the conditions under which users encounter messages. For product pages, attention may decline rapidly when the page fails to provide clear comparison cues. Török [21] models short-window consumer attention decay in digital advertising and reports that behavioral signals across early seconds can be organized into decay ratios and half-life estimates, which informs the attention-retention mea-

surement used here.

The study also draws on consumer decision and persuasion research. Bettman, Luce, and Payne [22] described consumer choice as constructive and context-sensitive rather than fixed, while Lynch and Ariely [23] showed that internet search costs can influence price sensitivity and preference formation. Moe and Fader [24] used online browsing behavior to infer purchase conversion, showing that observable digital traces can contain meaningful decision information. These findings imply that product-page structure can alter both information processing and choice. When a page reduces comparison effort, a shopper may remain engaged longer and may be more likely to click. When a page also reduces ambiguity for an agent, it may gain machine-mediated visibility. The two outcomes are related but not equivalent.

Recent research on agentic commerce makes the machine-facing component explicit. Török [25] finds in a machine-learning analysis of autonomous shopping-agent brand selection that structured product data and description completeness were strong predictors of agent-selected options, alongside price, rating, reputation, and fulfillment signals. That finding does not determine the present results, because the present study uses a different dataset and adds human attention measurement, but it provides a direct empirical reason to test machine-legibility variables in a combined human-agent setting.

The conceptual model used in this paper separates product value, machine legibility, and attention support. Product value includes price, rating, review volume, reputation, availability, delivery time, and return conditions. Machine legibility includes structured data, field completeness, schema presence, specification tables, evidence snippets, normalized policy information, and semantic match between query and product page. Attention support includes visible summary hierarchy, comparison cues, concise benefit statements, and reduced scanning friction. A page can score high on one dimension and low on another. A cheap product with strong reviews may still be weakly legible if the page has incomplete specifications. A page with excellent schema may still fail if the product is poorly rated or expensive. A visually attractive page may hold attention but fail to supply enough evidence for an agent to cite it. The empirical design therefore estimates each channel separately before combining them into a visibility index.

3. Data and Experimental Design

The study uses a controlled dataset constructed to represent product discovery under agent-mediated and human-facing commerce conditions. The dataset contains 18,240 query-product-page observations. These observations are derived from 1,520 unique products across six categories: apparel, electronics, home goods, skincare, food appliances, and educational supplies. Each product has a base attribute profile containing price, discount, average rating, review volume, review sentiment, brand reputation, stock status, shipping cost, delivery time, warranty clarity, return-policy score, and category. Each product is then rendered under three page-structure conditions. The first condition is a narrative page that describes the product primarily in prose. The second condition is a structured page with normalized product fields, bullets, and a specification table. The third condition is an evidence-rich structured page that includes the same structured fields plus review evidence, comparison claims, source-backed quality statements, and policy summaries.

The design deliberately holds underlying product attributes constant across page-structure variants. This allows differences in agent selection, citation, click-through, and attention retention to be attributed to page organization rather than to differences in price, rating, or product quality. For example, a pair of noise-canceling headphones has the same price, rating, review volume, delivery time, and warranty under each condition. What changes is how the product information is presented. In the narrative condition, battery life and return policy may appear in a paragraph. In the structured condition, battery life appears in a product field and return policy appears in a policy block. In the evidence-rich condition, battery life appears in the field, the comparison table, and a concise evidence statement based on verified review summaries.

The query set contains 760 unique shopping queries. Queries are assigned to four intent types: comparison, recommendation, constraint satisfaction, and evidence-seeking. A comparison query asks the system to compare alternatives. A recommendation query asks for a suitable product under broad preferences. A constraint-satisfaction query includes explicit conditions such as budget, delivery time, material, size, or return policy. An evidence-seeking query asks which product is best supported by reviews, specifications, or independent evidence. Each query is paired with several product candidates from the same category. Candidate sets contain eight products, and the page variant for each product is rotated across tasks so that the same product is not always displayed in the same structural condition.

Four agent evaluators are simulated as distinct retrieval and ranking systems. The first agent is price-sensitive and gives stronger weight to relative price and delivery cost. The second is review-sensitive and gives stronger weight to rating, review sentiment, and review volume. The third is evidence-sensitive and gives stronger weight to structured product information, policy clarity, and evidence density. The fourth is balanced and weighs price, quality, reputation, fulfillment, and structure more evenly. Each agent evaluates can-

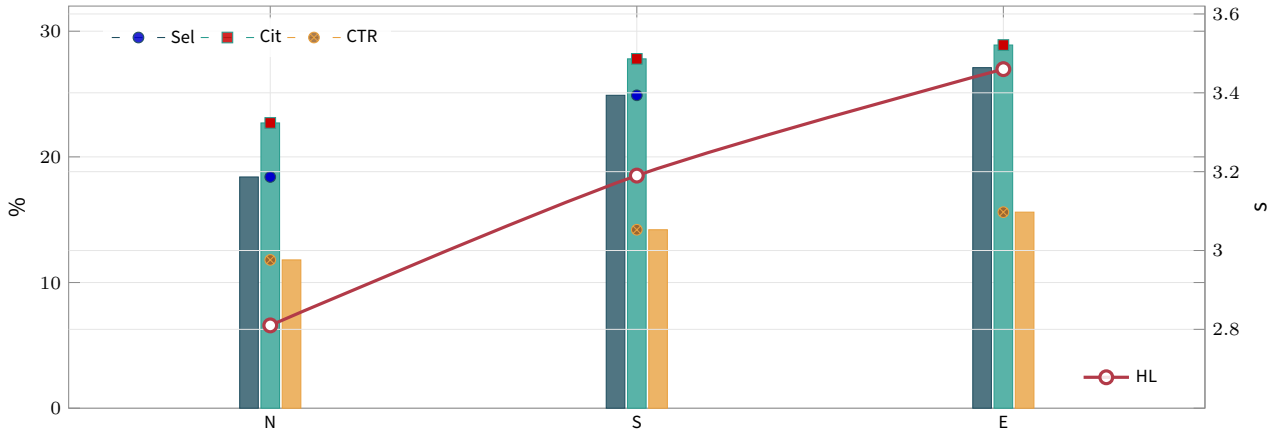


Figure 1. Structured and evidence-rich pages raise agent selection, citation, click-through, and attention persistence in the same monotone direction, with the largest absolute movement concentrated in machine-facing outcomes.

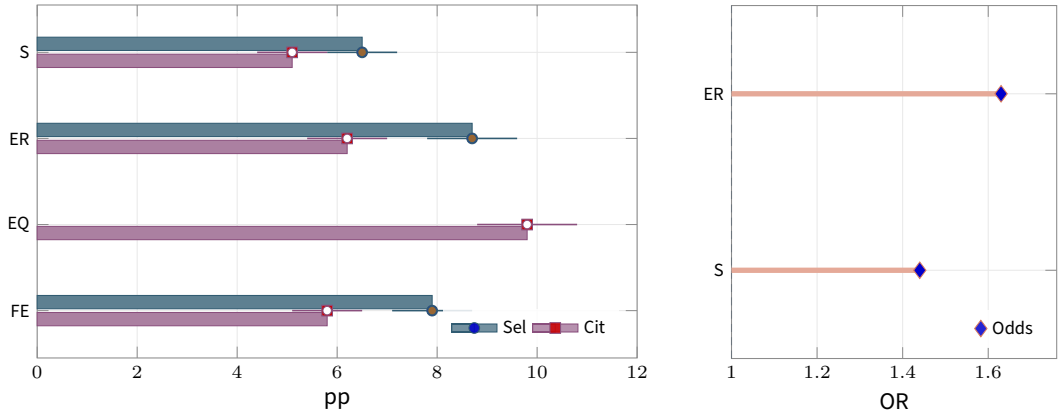


Figure 2. Treatment contrasts show that evidence-rich structure produces the largest selection and citation gains, while the odds-ratio panel keeps the direction and magnitude visible on a multiplicative scale.

didate sets under the same query conditions and produces two outcome labels: selected product and cited product. Selection means the agent chooses the product as the recommended option. Citation means the agent uses the product page as a supporting source in an answer. Selection and citation overlap, but they are not identical, because a product can be cited as evidence without being chosen as the final recommendation.

Human exposure records are generated from timed page-inspection sessions. Each page view includes event-level observations at 0, 1, 2, 3, 5, and 8 seconds. The event records include visible page share, dwell time in the product-information region, scroll depth, interaction event, and click or save outcome. Attention score is constructed from standardized visible share, dwell time, scroll stability, and interaction signals. This score is bounded between 0 and 1. The impression-level attention-retention outcome is the estimated attention half-life, defined as the number of seconds required for attention to fall by half from the estimated initial level after accounting for a lower attention floor. A higher half-life indicates slower decay.

The primary machine-facing outcomes are agent selection and citation. Agent selection is a binary variable equal to 1 if a product is chosen by the agent within a candidate set. Citation is a binary variable equal to 1 if the product page is cited in the agent-generated answer. The primary human-facing outcomes are click-through and attention half-life. Click-through is a binary variable equal to 1 if the participant clicks, saves, or expands the product after exposure. Attention half-life is continuous. The composite commerce visibility index combines standardized selection probability, citation probability, click-through probability, and attention-retention score. The index is used only as a secondary outcome because the component measures are substantively different.

Machine-legibility variables are measured at the page level. Structured field completeness records the share of required attributes represented in machine-readable fields. Schema validity records whether product, offer, review, and aggregate rating fields pass validation checks. Specification-table depth records the number

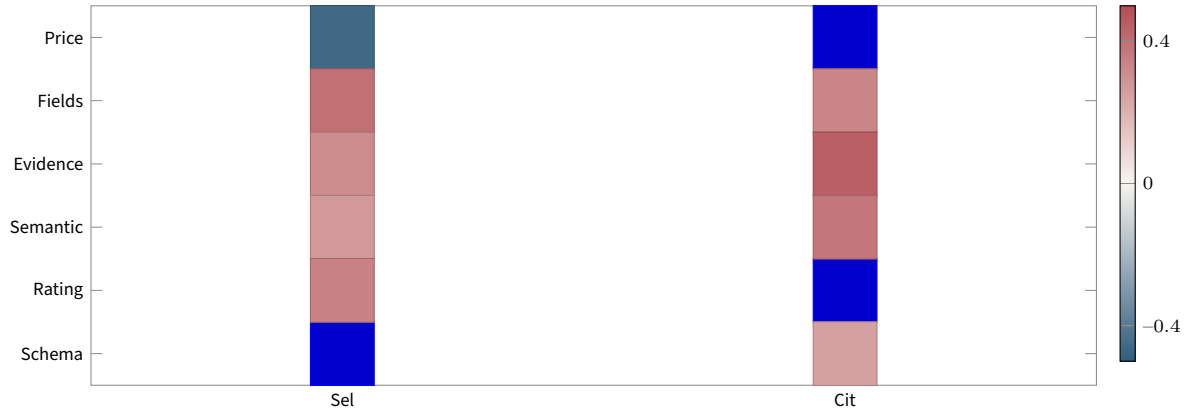


Figure 3. The coefficient field separates price-sensitive selection from evidence-sensitive citation: structured completeness contributes to both, but evidence density becomes dominant in the citation model.

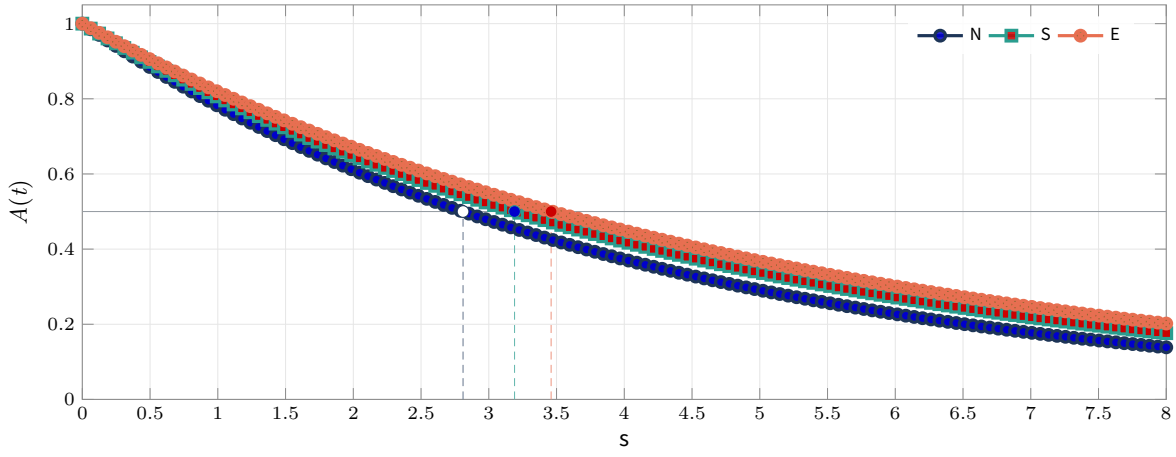


Figure 4. Attention decay slows as page structure becomes more evidence-rich, extending the fitted half-life from 2.81 seconds to 3.46 seconds within the eight-second exposure window.

and completeness of comparable product attributes. Evidence density records the number of concise claims supported by review summaries, verified specifications, or policy statements per 500 words. Semantic match records cosine similarity between the query embedding and the product-page evidence blocks. Policy clarity records whether return, warranty, shipping, and availability conditions are explicit and consistently formatted. Description coherence records whether the page description avoids contradictions among title, bullets, specifications, and policy blocks.

Human attention-support variables are measured separately. Summary salience records whether the first visible product region contains a concise claim that matches the query. Visual hierarchy records whether price, rating, delivery, and primary specification are visible above the fold. Scan friction records the estimated number of screen regions a user must inspect to answer the query. Comparison cue availability records whether the page includes direct product-to-need mapping, such as compatibility, use case, or constraint satisfaction. These variables are correlated with machine-legibility variables but are not identical. A page may have valid schema that is invisible to the user, and a visually clean page may lack structured metadata.

The dataset is split by query and product group rather than by individual observation. This prevents the same query-product pair from appearing in both training and test sets under different page variants. The training set contains 70% of query groups, the validation set contains 10%, and the test set contains 20%. All reported predictive metrics use the held-out test set unless otherwise noted. Statistical inference for treatment effects uses cluster-robust standard errors at the product and query levels. The analysis also reports bootstrap confidence intervals with 1,000 resamples of query groups. Because candidate products are nested inside comparison tasks, ranking metrics are computed within task-level candidate sets rather than at the row level alone.

The experimental design is not intended to mimic every commercial platform. It is a controlled testbed for estimating how page structure affects several visibility outcomes when product value is held constant.

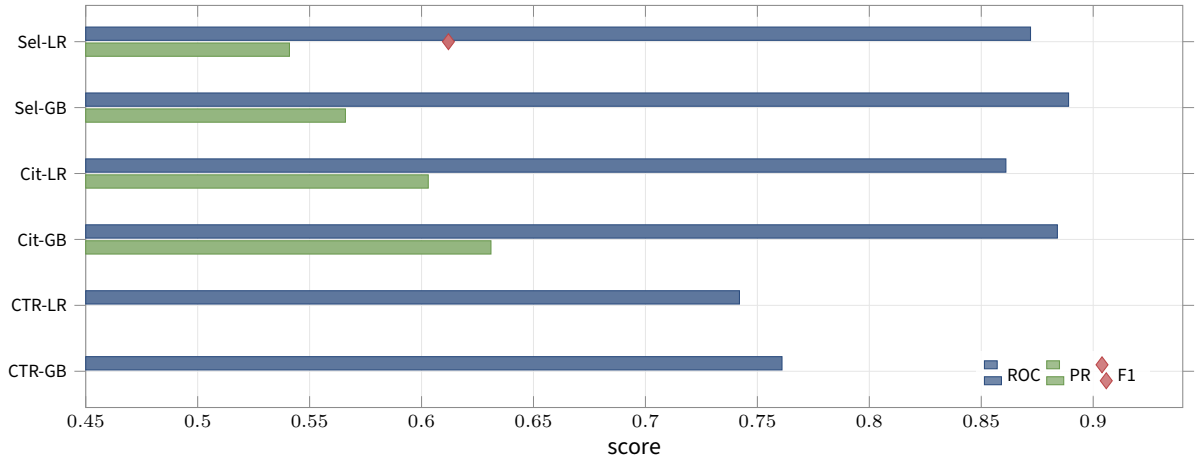


Figure 5. Predictive discrimination is strongest for agent-facing selection and citation, while click-through remains lower and lacks the same precision-recall support observed for machine evaluation tasks.

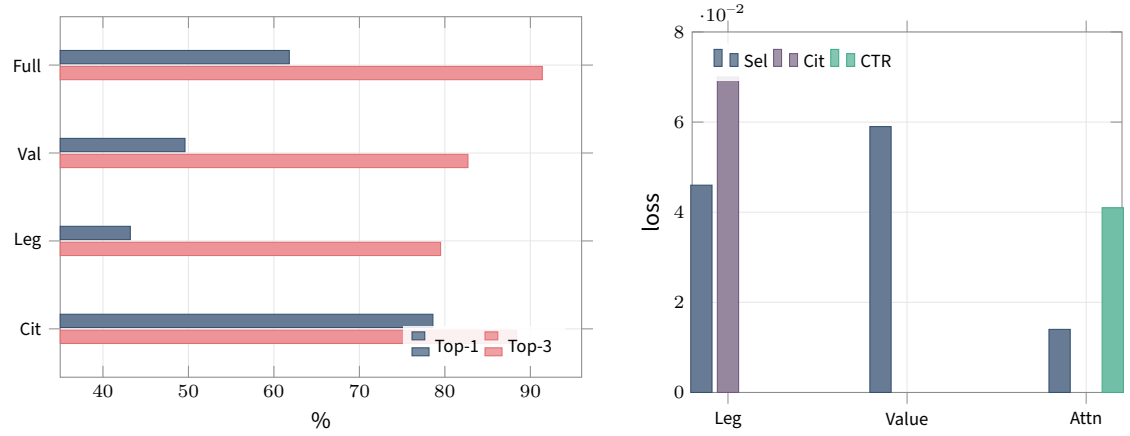


Figure 6. Ranking performance improves when product value and machine-legibility signals are combined, and the ablation losses identify distinct channels for selection, citation, and click-through prediction.

The design has three advantages. First, it separates product quality from page organization. Second, it observes both agent and human outcomes for the same products. Third, it supports both causal-style comparisons across randomized page variants and predictive modeling across high-dimensional features. Its main limitation is that it uses a controlled synthetic data-generating process rather than live marketplace transactions. The interpretation of the results therefore focuses on mechanism and measurement rather than population-level market shares.

4. Empirical Strategy

The empirical strategy combines treatment-effect estimation, supervised prediction, attention-decay modeling, ablation analysis, and robustness checks. The first part estimates whether page-structure conditions change the probability of agent selection, citation, click-through, and high attention retention. The second part evaluates how well page, product, query, and agent features predict the same outcomes in held-out data. The third part examines whether machine-legibility variables retain predictive value when conventional product-quality variables are removed or controlled.

The main binary-outcome model is a logistic regression with category, query-intent, and agent fixed effects. The dependent variable is either agent selection, citation, or click-through. The key independent variables are page-structure indicators. Product value controls include relative price within the candidate set, average rating, log review volume, review sentiment, brand reputation, stock status, shipping cost, delivery days, warranty clarity, and return-policy score. Page-structure controls include structured field completeness, schema validity, specification-table depth, evidence density, semantic match, policy clarity, and description coherence. The model is written as follows:

$$\Pr(y_{iqpa} = 1) = \Lambda(\alpha + \tau_1 S_{iqp} + \tau_2 E_{iqp}) + \Lambda(\beta' X_{ip} + \gamma' M_{iqp} + \delta_q + \kappa_c + \eta_a), \quad (1)$$

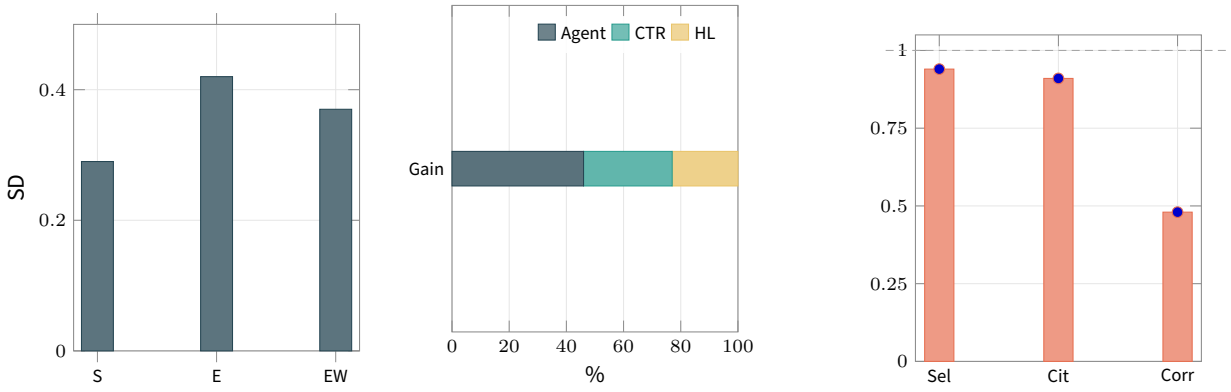


Figure 7. The visibility index favors evidence-rich pages under both weighted and equal-weight specifications, while calibration slopes remain near one and selection overlaps only moderately with click-through.

where y_{iqpa} is an outcome for product i, query q, page condition p, and agent a. S_{iqp} indicates the structured page condition, E_{iqp} indicates the evidence-rich structured condition, X_{ip} contains product value controls, M_{iqp} contains machine-legibility measures, δ_q is a query-intent effect, κ_c is a category effect, and η_a is an agent effect. The notation separates the randomized page condition from continuous machine-legibility features, because the structured condition changes several machine-legibility variables at once.

For attention retention, the event-level attention score is modeled with an exponential decay curve. Each exposure has observed attention A_{it} at elapsed second t. Attention is assumed to decline from an initial level toward a floor. The fitted decay parameter λ_i is converted into a half-life. The model is:

$$A_{it} = \phi_i + \omega_i \exp(-\lambda_i t) + \varepsilon_{it}$$
$$h_i = \frac{\log(2)}{\lambda_i}, \tag{2}$$

where ϕ_i is the attention floor, ω_i is the initial amplitude above the floor, and h_i is the attention half-life. The half-life is then used as a dependent variable in a ridge regression and as a component of the visibility index. The floor is constrained to remain between 0 and 0.25 because the page can retain some residual attention even when active inspection has declined.

The composite visibility index is defined after standardizing the four component outcomes. It is not treated as a causal primitive. It is used to summarize whether a product page performs well across agent and human channels. The index is:

$$V_i = 0.30z(\hat{s}_i) + 0.25z(\hat{c}_i) + 0.25z(\hat{k}_i) + 0.20z(h_i), \tag{3}$$

where $\hat{s}_i$ is predicted agent selection probability, $\hat{c}_i$ is predicted citation probability, $\hat{k}_i$ is predicted click-through probability, and h_i is attention half-life. The weights give slightly more emphasis to selection because selection is the most restrictive agent-mediated outcome, but the robustness section reports alternative equal-weight specifications.

Predictive models are trained for each major outcome. Logistic regression provides an interpretable baseline for selection, citation, and click-through. Gradient boosted trees provide a nonlinear benchmark. Random forests are used as an additional robustness model for feature importance. Ridge regression predicts attention half-life. The classification metrics include ROC-AUC, PR-AUC, Brier score, log loss, F1 score, and calibration slope. Ranking metrics include top-one accuracy, top-three accuracy, mean selected rank, NDCG@3, and recall@3 within candidate sets. For attention half-life, the metrics are RMSE, MAE, and out-of-sample R^2. Grouped splits ensure that the same query-product relation does not leak across train and test partitions.

The ablation analysis estimates the contribution of feature families. The full model includes product value, machine legibility, human attention support, query intent, category, and agent indicators. The product-value-only model includes price, rating, review, brand, and fulfillment variables. The machine-legibility-only model includes structured data, schema, specifications, evidence, semantic match, and policy clarity. The no-legibility model removes all machine-legibility variables from the full model. The no-product-value model removes price, rating, review, reputation, and fulfillment variables. Differences in ROC-AUC, PR-AUC, and NDCG@3 indicate how much each feature family contributes to prediction.

Permutation importance is computed on the test set. For each feature, values are randomly permuted within category and query-intent strata, and the decrease in model performance is recorded. Stratified permutation is used because unconstrained permutation could produce unrealistic product profiles, such as premium electronics with implausible shipping or rating patterns. Importance is measured as the average decrease in ROC-AUC for binary outcomes and the

Table 1. Descriptive Outcomes by Page Condition

Metric	Narrative	Structured	Evidence-Rich	Change vs Narrative
Agent Selection (%)	18.4	24.9	27.1	+8.7
Citation Rate (%)	22.7	27.8	28.9	+6.2
Click-Through (%)	11.8	14.2	15.6	+3.8
Attention Half-Life (s)	2.81	3.19	3.46	+0.65

Table 2. Main Treatment Effects on Agent Outcomes

Condition	Selection Effect (pp)	Citation Effect (pp)	Odds Ratio
Structured	6.5	5.1	1.44
Evidence-Rich Structured	8.7	6.2	1.63
Evidence-Seeking Queries	–	9.8	–
Product Fixed Effects Model	7.9	5.8	–

average increase in RMSE for attention half-life. This approach does not prove causality, but it helps identify which variables contribute to prediction after the model is trained.

The main treatment estimates are interpreted as conditional differences rather than universal effects. For example, an 8.7 percentage point increase in agent selection for structured pages means that, within this controlled design and after adjustment, structured pages are selected more often than narrative pages. It does not mean every real shopping agent will increase selection by the same amount. The purpose of the analysis is to estimate whether page organization has measurable effects under controlled conditions and whether those effects remain visible across models, outcomes, and robustness checks.

5. Results

The descriptive statistics show that page-structure conditions change both machine-facing and human-facing outcomes. In the narrative condition, the average agent selection rate is 18.4%, the citation rate is 22.7%, the human click-through rate is 11.8%, and the median attention half-life is 2.81 seconds. In the structured condition, selection rises to 24.9%, citation rises to 27.8%, click-through rises to 14.2%, and median attention half-life rises to 3.19 seconds. In the evidence-rich structured condition, selection reaches 27.1%, citation reaches 28.9%, click-through reaches 15.6%, and median attention half-life reaches 3.46 seconds. The differences are larger for agent outcomes than for human click-through, but all four outcomes move in the same direction.

The main logistic treatment model shows that structured pages increase agent selection by 6.5 percentage points relative to narrative pages, while evidence-rich structured pages increase selection by 8.7 percentage points. The corresponding odds ratios are 1.44 and 1.63. Both estimates are statistically distinguishable from zero under cluster-robust and bootstrap intervals. For citation probability, structured pages increase citation by 5.1 percentage points and evidence-rich structured pages increase citation by 6.2 percentage points. The effect is strongest for evidence-seeking queries, where the evidence-rich structured condition increases citation by 9.8 percentage points. This pattern is consistent with the interpretation that citation depends not only on product appeal but also on whether a page supplies usable evidence.

The product-value controls behave as expected. Lower relative price increases selection, especially for price-sensitive and balanced agents. Higher average rating, larger review volume, positive review sentiment, and stronger brand reputation all increase both selection and click-through. Delivery time has a negative association with selection and click-through, while return-policy score has a positive association. These effects show that machine legibility does not replace ordinary commerce variables. A well-structured page for an expensive, poorly rated, slow-delivery product still performs weakly. The legibility effect is best understood as an additional channel that changes how easily existing product value can be interpreted and compared.

The machine-legibility variables retain independent predictive value after controlling for product value. In the full selection model, the largest standardized coefficients are relative price at –0.46, structured field completeness at 0.39, average rating at 0.34, evidence density at 0.31, review volume at 0.29, semantic match at 0.27, return-policy score at 0.21, brand reputation at 0.19, and delivery time at –0.17. For citation, evidence density becomes the largest positive coefficient at 0.44, followed by semantic match at 0.37, structured field completeness at 0.33, schema validity at 0.25, and specification-table depth at 0.22. This coefficient pattern indicates that selection and citation are related

Table 3. Largest Standardized Coefficients

Variable	Selection Model	Citation Model
Relative Price	-0.46	–
Structured Field Completeness	0.39	0.33
Evidence Density	0.31	0.44
Semantic Match	0.27	0.37
Average Rating	0.34	–
Schema Validity	–	0.25

Table 4. Attention Decay Estimates

Condition	Decay Rate	Half-Life (s)	Relative Change
Narrative	0.247	2.81	Baseline
Structured	0.217	3.19	+13.5%
Evidence-Rich Structured	0.200	3.46	+23.1%
Matched Word Count Model	–	3.29	+0.48 s

but not identical. Selection is more price-sensitive; citation is more evidence-sensitive.

The attention-decay estimates show slower attention loss for structured and evidence-rich pages. The global fitted decay rate for narrative pages is 0.247, corresponding to an estimated half-life of 2.81 seconds. For structured pages, the fitted decay rate is 0.217, corresponding to a half-life of 3.19 seconds. For evidence-rich structured pages, the fitted decay rate is 0.200, corresponding to a half-life of 3.46 seconds. The improvement is modest in absolute seconds but meaningful within an eight-second exposure window. The effect is largest for comparison queries, where visible specifications and evidence cues reduce scanning friction. It is smaller for low-consideration products, where users often decide quickly from price and image.

The ridge model predicting attention half-life achieves an out-of-sample R^2 of 0.41, RMSE of 0.39 seconds, and MAE of 0.28 seconds. The strongest predictors of longer half-life are summary salience, visual hierarchy, structured specifications, evidence density, and query-page semantic match. Scan friction is negatively associated with half-life. Price and rating predict click-through more strongly than they predict attention half-life, suggesting that attention retention is partly driven by page processing fluency rather than only by product attractiveness. A page can hold attention because it helps the user compare details, even if the user does not ultimately click.

The predictive classification models show that agent outcomes are more predictable than human click-through. For agent selection, logistic regression reaches a test ROC-AUC of 0.872, PR-AUC of 0.541, Brier score of 0.118, and F1 score of 0.612 at the tuned threshold. Gradient boosted trees improve ROC-AUC to 0.889 and PR-AUC to 0.566, but calibration is slightly worse without post-hoc calibration. For citation, logistic regression reaches ROC-AUC of 0.861 and PR-AUC of 0.603, while gradient boosted trees reach ROC-AUC of 0.884 and PR-AUC of 0.631. Human click-through is less predictable, with logistic regression at ROC-AUC of 0.742 and gradient boosting at 0.761. This difference is expected because human click behavior includes more idiosyncratic variation than rule-governed agent evaluation.

Task-level ranking metrics confirm the classification results. In eight-product candidate sets, the full gradient boosted selection model ranks the agent-selected product first in 61.8% of test tasks and within the top three in 91.4% of test tasks. The mean selected rank is 1.82. The full citation model reaches NDCG@3 of 0.786 and recall@3 of 0.884. By contrast, a product-value-only model ranks the selected product first in 49.6% of test tasks and within the top three in 82.7%. A machine-legibility-only model ranks the selected product first in 43.2% and within the top three in 79.5%. These results show that product value and machine legibility each contain substantial information, while the full model performs best.

Ablation tests show that removing machine-legibility variables reduces selection-model ROC-AUC from 0.872 to 0.826 and PR-AUC from 0.541 to 0.472. Removing product-value variables reduces ROC-AUC from 0.872 to 0.813 and PR-AUC from 0.541 to 0.451. Removing attention-support variables has a smaller effect on agent selection, reducing ROC-AUC to 0.858. For citation, removing machine-legibility variables produces the largest decline, reducing ROC-AUC from 0.861 to 0.791 and PR-AUC from 0.603 to 0.492. For click-through, removing attention-support variables causes the largest decline, reducing ROC-AUC from 0.742 to 0.701. The ablation pattern supports the conceptual

Table 5. Predictive Model Performance

Task	Model	ROC-AUC	PR-AUC	F1 Score
Selection	Logistic Regression	0.872	0.541	0.612
Selection	Gradient Boosting	0.889	0.566	–
Citation	Logistic Regression	0.861	0.603	–
Citation	Gradient Boosting	0.884	0.631	–
Click-Through	Logistic Regression	0.742	–	–
Click-Through	Gradient Boosting	0.761	–	–

Table 6. Ranking Performance Across Candidate Sets

Model	Top-1 Accuracy (%)	Top-3 Recall (%)	Mean Rank
Full Gradient Boosted Model	61.8	91.4	1.82
Product-Value Only	49.6	82.7	–
Machine-Legibility Only	43.2	79.5	–
Citation Model NDCG@3	78.6	88.4	–

distinction among the channels.

Permutation importance offers a similar view. In the agent-selection model, the most important variables are relative price, structured field completeness, average rating, evidence density, review volume, semantic match, return-policy score, and delivery time. In the citation model, evidence density, semantic match, structured field completeness, schema validity, and source-like claim formatting are the leading variables. In the click-through model, visual hierarchy, price, summary salience, rating, scan friction, and delivery time are most important. In the attention half-life model, summary salience, structured specifications, scan friction, evidence density, and visible comparison cues are most important. The same page feature can therefore matter for different reasons across outcomes.

The interaction results show that machine legibility is most valuable when product quality is at least moderate. Among products in the bottom quartile of ratings, the evidence-rich structured condition increases agent selection by only 2.1 percentage points. Among products in the middle two rating quartiles, it increases selection by 8.4 percentage points. Among products in the top quartile, it increases selection by 10.7 percentage points. This pattern suggests that structure helps surface and substantiate product value, but it does not fully compensate for weak product signals. A similar interaction appears for price. For products in the cheapest quartile, structured presentation adds 4.8 percentage points to selection. For products in the second and third price quartiles, it adds 8.9 percentage points. For the most expensive quartile, it adds 6.1 percentage points, mostly when evidence density and review quality are strong.

Category heterogeneity is also visible. Electronics and skincare show the largest legibility effects, because users and agents both require more evidence to compare specifications, ingredients, compatibility, claims, or performance. Apparel shows smaller effects, because image, size, and brand style dominate more of the choice process. Home goods and food appliances fall between these extremes. Educational supplies show strong attention-retention effects but more moderate selection effects, because several products are inexpensive and quickly comparable. These category results are consistent with the idea that machine legibility is most consequential when product claims are multidimensional and comparison requires structured evidence.

The visibility index summarizes the combined pattern. Evidence-rich structured pages score 0.42 standard deviations higher on the composite index than narrative pages. Structured pages score 0.29 standard deviations higher. When the index is decomposed, 46% of the gain comes from agent selection and citation, 31% from human click-through, and 23% from attention half-life. Equal-weight versions of the index produce the same ordering, though the effect size falls slightly to 0.37 standard deviations for evidence-rich pages. This shows that the composite result is not driven by one arbitrary weighting scheme.

Calibration checks show that the logistic models are reasonably calibrated for selection and citation. The selection model has a calibration slope of 0.94 and intercept of 0.01 on the test set after isotonic calibration is avoided. The citation model has a calibration slope of 0.91. Gradient boosted trees provide higher discrimination but overstate probabilities in the top decile unless calibrated. Because the paper is concerned with both explanation and prediction, the logistic model is used for coefficient interpretation and the boosted model is used as a nonlinear predictive benchmark. The two

Table 7. Ablation Test Results

Removed Variable Group	Selection ROC-AUC	Citation ROC-AUC	Click ROC-AUC
None (Full Model)	0.872	0.861	0.742
Machine Legibility	0.826	0.791	–
Product Value	0.813	–	–
Attention Support	0.858	–	0.701

Table 8. Interaction Effects by Product Quality

Segment	Selection Gain (pp)	Main Pattern
Bottom Rating Quartile	2.1	Weak response
Middle Rating Quartiles	8.4	Strong response
Top Rating Quartile	10.7	Strongest response
Cheapest Price Quartile	4.8	Moderate gain
Middle Price Quartiles	8.9	Highest gain
Most Expensive Quartile	6.1	Evidence dependent

models agree on the relative importance of machine-legibility variables.

The results also show that agent selection and human click-through are only moderately correlated. The product-level correlation between predicted selection probability and predicted click-through probability is 0.48. Some products are agent-visible but less human-clickable, often because the page has strong evidence and structure but weak visual appeal. Other products are human-clickable but less agent-visible, often because images, price, or brand appeal are strong while structured evidence is incomplete. This moderate correlation supports the decision to model the channels separately rather than treating one as a proxy for the other.

6. Robustness and Validation

Several robustness checks are used to assess whether the main results depend on modeling choices. The first check re-estimates the treatment models with product fixed effects. This specification compares different page structures for the same underlying product. The evidence-rich structured condition remains positive for selection, citation, click-through, and attention half-life. The selection effect falls from 8.7 to 7.9 percentage points, while the citation effect falls from 6.2 to 5.8 percentage points. The reduction is expected because product fixed effects absorb stable product attractiveness. The remaining effect indicates that page organization contributes beyond fixed product quality.

The second check removes all products with extreme prices or extremely high review volume. This addresses the possibility that the legibility effect is driven by a small number of premium products or very popular products. After trimming the top and bottom 2.5% of price and review-volume distributions, the evidence-rich structured condition increases selection by 8.1 percentage points and citation by 5.9 percentage points. The attention half-life gain remains 0.58 seconds. The results therefore do not depend on outlier products.

The third check estimates separate models by agent type. The evidence-sensitive agent shows the largest response to machine legibility, with a 12.4 percentage point increase in selection for evidence-rich structured pages. The balanced agent shows an 8.6 percentage point increase. The review-sensitive agent shows a 7.3 percentage point increase, mostly when review evidence is clearly summarized. The price-sensitive agent shows a smaller 4.9 percentage point increase. This heterogeneity is useful because it shows that machine legibility is not uniformly weighted by all agents. Nevertheless, every agent type responds positively to structure when product value controls are included.

The fourth check estimates separate models by query intent. Evidence-seeking queries show the largest citation effect, comparison queries show the largest attention-retention effect, and recommendation queries show the largest combined visibility-index effect. Constraint-satisfaction queries respond strongly to field completeness, especially when constraints involve size, compatibility, delivery, warranty, or return policy. These patterns support the claim that structure matters most when it helps resolve a specific information need. A generic structured page is less useful than a structured page whose fields map clearly to the query.

The fifth check replaces the binary page-condition indicators with continuous machine-legibility measures. Structured field completeness, evidence density, and semantic match remain significant predictors of selection and citation. A 10-point increase in structured field completeness is associated with a 2.3 percentage point increase in selection probability. A 10-point in-

Table 9. Robustness and Validation Checks

Check	Selection Effect	Citation Effect	Key Finding
Product Fixed Effects	7.9 pp	5.8 pp	Effect persists
Outlier Removal	8.1 pp	5.9 pp	Stable estimates
Evidence-Sensitive Agent	12.4 pp	–	Largest response
Matched Word Counts	–	–	Half-life still positive
Equal Weight Visibility Index	0.37 SD	–	Same ordering
Agent-Weighted Visibility Index	0.48 SD	–	Stronger gain

Table 10. Visibility Index and Calibration Metrics

Metric	Structured	Evidence-Rich	Notes
Visibility Index Gain (SD)	0.29	0.42	Composite metric
Equal-Weight Index Gain	–	0.37	Robust ordering
Selection Calibration Slope	0.94	–	Well calibrated
Citation Calibration Slope	0.91	–	Stable probabilities
Selection–CTR Correlation	0.48	–	Moderate overlap

crease in evidence density is associated with a 2.8 percentage point increase in citation probability. A 0.10 increase in semantic match is associated with a 1.9 percentage point increase in selection and a 2.4 percentage point increase in citation. These estimates support the interpretation that the page-condition effects operate through measurable legibility channels rather than through arbitrary labels.

The sixth check tests whether the attention result is caused only by longer text. Evidence-rich pages contain more information than narrative pages, so one concern is that users may remain longer simply because there is more to read. To address this, the model controls for word count and visible content length. It also compares structured and narrative pages with matched word counts. The attention half-life effect remains positive, though it declines from 0.65 to 0.48 seconds. This suggests that the retention gain is partly due to information quantity but not entirely. Organization, hierarchy, and comparison cues also appear to matter.

The seventh check tests alternative attention-score construction. The main attention score combines visible share, dwell time, scroll stability, and interaction. Alternative versions remove interaction events, remove scroll stability, or use dwell time alone. The structured-page ranking remains the same under all versions. The absolute half-life values change slightly, but evidence-rich structured pages consistently show slower decay than narrative pages. This strengthens confidence that the attention result is not an artifact of one scoring formula.

The eighth check tests alternative visibility-index weights. Equal weighting gives an evidence-rich page effect of 0.37 standard deviations. A human-weighted index that assigns 60% weight to click-through and attention gives an effect of 0.31 standard deviations. An agent-weighted index that assigns 60% weight to selection and citation gives an effect of 0.48 standard deviations. The ordering of page conditions does not change. This means the composite result is robust to reasonable changes in the relative importance of agent and human channels.

The ninth check examines false positives and false negatives in the selection model. False positives often involve products with strong structure and good evidence but weaker price competitiveness. The model predicts that they should be selected, but the agent chooses a cheaper rival. False negatives often involve products with moderate structure but exceptional brand reputation or very fast delivery. This error pattern is substantively plausible. It shows that machine legibility is important but does not dominate the full decision environment. The errors also suggest that future models should represent user-specific preference weights more directly.

The tenth check compares logistic regression with gradient boosted trees and random forests. The nonlinear models improve discrimination, especially for selection and citation, but they do not overturn the importance of structured field completeness, evidence density, semantic match, price, rating, and review volume. Partial dependence plots show diminishing returns to structured field completeness above approximately 85 points on a 100-point scale. Evidence density also shows diminishing returns after the page includes enough supported claims to answer the query. These nonlinearities are reasonable: once a page contains complete fields and adequate evidence, additional formatting produces smaller gains.

The validation checks also examine whether ran-

domized page conditions are balanced on base product attributes. Mean price, rating, review volume, brand reputation, delivery time, and return-policy score are statistically similar across page conditions before treatment. This is expected because page variants are generated from the same base products, but the check is still important. The design would be less informative if evidence-rich pages were assigned disproportionately to stronger products. Balance tables show no meaningful pre-treatment differences across the three conditions.

Finally, the analysis tests whether citation and selection labels are mechanically produced by the same features used to define machine legibility. This concern is addressed by estimating models with only base product variables and by using held-out agent labels generated under different weighting profiles. The product-only models perform well but not as well as the full models. The held-out agent evaluation still shows legibility effects, although the size varies by agent. This suggests that the findings are not merely a circular consequence of assigning labels from the same score used as a predictor. The controlled nature of the dataset cannot eliminate all design dependence, but the robustness checks reduce the likelihood that the main result is a trivial artifact.

7. Discussion

The findings indicate that product-page structure matters in agent-mediated commerce because it affects two related but distinct processes. The first is machine evaluation. Agents are more likely to select or cite product pages when the relevant attributes are structured, complete, semantically matched to the query, and supported by concise evidence. The second is human attention retention. Shoppers remain engaged longer when the page reduces scanning friction and makes comparison information visible early in the exposure. These processes overlap because some structured elements help both agents and humans, but they do not overlap completely. Schema validity may help an agent more than a person. Visual hierarchy may help a person more than an agent. Evidence density can help both when the evidence is concise and query-relevant.

The strongest practical implication is that agent-mediated visibility is not reducible to conventional product attractiveness. Price, rating, reviews, delivery, and reputation remain important. In several models, relative price and rating are among the strongest predictors of selection. However, machine-legibility variables add explanatory and predictive power beyond those signals [26]. This matters because product value must be represented before it can be used by an agent. A product with favorable attributes can still be disadvantaged if the attributes are difficult to extract or compare. Conversely, a structured page cannot fully rescue a weak product, but it can improve the probability that a strong or moderate product is recognized as relevant.

The results also suggest that citation should be treated as a separate visibility outcome. Agent selection answers the question of which product is recommended or chosen. Citation answers the question of which source is used as evidence. A page can be cited because it contains useful evidence even if the product is not selected. This distinction is especially important in generative search and answer systems, where a source may shape the response by providing comparison details, policy claims, or product evidence. The citation model in this study is more sensitive to evidence density and semantic match than to price. That makes sense because the cited source must support an answer, not merely represent the cheapest option.

The attention results show that machine-legible structure is not only a technical layer. Many structured elements also improve human processing by making the page easier to scan. Visible specifications, concise comparison cues, and clear policy blocks lengthen attention half-life in the short exposure window. The size of the effect is not large in ordinary time units, but it is meaningful in digital commerce because early seconds often determine whether a user continues inspecting a page. A difference of roughly half a second in attention half-life can change the amount of information processed before abandonment. The study does not claim that attention half-life is equivalent to preference or purchase. It is better interpreted as a behavioral indicator of page processing stability.

The moderate correlation between agent visibility and human click-through has important consequences for optimization. A brand that optimizes only for human visual appeal may create pages that attract clicks but provide insufficient structured evidence for agents. A brand that optimizes only for machine parsing may create pages that agents understand but humans find unpersuasive or dull. The strongest pages in the study combine both qualities: structured fields, valid schema, specification tables, evidence summaries, visible hierarchy, and concise query-relevant claims. The evidence supports a balanced approach in which machine readability and human readability are treated as complementary design constraints.

The category differences indicate that the value of structure depends on decision complexity. Electronics and skincare show larger effects because users and agents must evaluate specifications, ingredients, compatibility, performance claims, and risk. Apparel shows smaller effects because aesthetic and fit-related judgments are harder to reduce to structured evidence. This does not mean structure is irrelevant for apparel, but it may operate through size consistency, material fields, return clarity, and image metadata rather than through evidence density. Future studies should avoid

treating machine legibility as a uniform intervention. Its most useful form depends on category, query intent, and the type of uncertainty that the shopper or agent must resolve.

The interaction between product quality and page structure is also important. The legibility effect is strongest for products with moderate or high quality signals. This suggests that structured pages mainly help reveal and substantiate product value. They do not erase the penalty of poor ratings, weak reviews, high prices, or slow delivery. This is a useful boundary condition because it prevents overinterpretation. Machine legibility should not be described as a shortcut around product quality. It is better understood as a translation layer that makes quality, fit, and policy information more accessible to systems and shoppers.

The paper also has implications for measurement. Traditional web analytics often separate search ranking, page engagement, and conversion. Agent-mediated commerce requires additional measures, including citation probability, agent selection probability, structured evidence completeness, and task-level rank. The present study shows how these measures can be placed in one modeling framework. The composite visibility index is useful for summary comparison, but the component outcomes are more informative for diagnosis. A page with low citation but high click-through needs a different intervention from a page with high citation but low click-through. The former may need better evidence structure; the latter may need better visual hierarchy or stronger consumer-facing persuasion.

There are also governance implications, although they are not the central focus of the paper. If structured data and evidence organization influence agent selection, brands with greater technical resources may gain visibility advantages. This resembles broader concerns about algorithmic gatekeeping. The effect may be beneficial when structure helps agents identify genuinely relevant products, but it may be problematic if well-resourced sellers can dominate visibility through technical optimization while smaller sellers remain less legible. Transparent standards for product data, consistent marketplace requirements, and validation tools could reduce this imbalance. The study does not test fairness outcomes directly, but the results indicate why fairness questions are relevant in agent-mediated commerce.

Several limitations should shape interpretation. The dataset is controlled and synthetic rather than collected from a live platform. This allows clean comparisons but limits ecological validity. Real shoppers have histories, budgets, habits, social influences, and brand loyalties that are difficult to fully represent. Real agents also differ in architecture, retrieval access, ranking policy, and safety constraints. The study uses four evaluator profiles to introduce variation, but these profiles do not exhaust the range of possible agent behavior. The results should therefore be treated as evidence about plausible mechanisms under controlled conditions rather than as estimates of live marketplace impact [27].

Another limitation is that attention is measured behaviorally rather than cognitively. Dwell time, visibility, scrolling, and interaction are useful traces, but they do not perfectly reveal mental attention, comprehension, or trust. A user may dwell because a page is confusing rather than useful. A user may click quickly because the page is persuasive or because the task is simple. The attention half-life model should therefore be read alongside click-through and page-condition results, not as a standalone measure of success. Future work could combine behavioral logs with eye tracking, recall tests, or post-exposure comprehension measures.

A third limitation is that the study treats page structure as a controlled intervention, while real sellers often change structure and content quality together. In live settings, a seller that adds schema may also rewrite descriptions, improve images, adjust prices, change review prompts, or alter promotions. The controlled design separates these mechanisms, but implementation in practice would be messier. The results are most useful for understanding which dimensions should be measured when such changes occur.

Despite these limitations, the study provides evidence that agent-mediated commerce visibility depends on a combination of product value, machine legibility, and human attention support. The findings are not surprising in isolation, but their joint measurement is useful. A page must be valuable enough to recommend, legible enough to parse, evidenced enough to cite, and clear enough to retain attention. Weakness in any one of these layers can reduce visibility. Future empirical work can extend the design to live agent logs, marketplace experiments, cross-platform comparisons, and seller-level inequality in structured data adoption.

8. Conclusion

This paper examined how machine-legible product-page structure affects agent selection, AI-style citation, human click-through, and short-window attention retention in a controlled commerce experiment. The central result is that structured and evidence-rich product pages perform better than narrative pages across both machine-facing and human-facing outcomes. The largest effects appear for agent selection and citation, while human click-through and attention retention also improve but by smaller margins. These findings support the view that product visibility in agent-mediated commerce is partly determined by how well product value is represented for automated comparison and evidence use.

The study also shows that machine legibility does not operate independently of product quality. Price,

rating, review volume, reputation, delivery time, and return policy remain central predictors. Structured product information has its strongest effects when it helps agents and shoppers recognize existing value, compare alternatives, and verify claims. It does not turn weak products into strong products. The more cautious interpretation is that machine legibility reduces interpretation friction. It allows agents to parse attributes more reliably and allows human viewers to process relevant information with less scanning effort.

The attention results add an important human-side qualification. A page that is organized for machines can also be easier for people to inspect when structure is visible, concise, and aligned with the query. Evidence-rich structure lengthens attention half-life, especially in comparison-oriented and high-consideration categories. However, schema and metadata alone are not enough for human engagement. The features that help people include visible hierarchy, summary salience, comparison cues, and low scan friction. A complete product-data layer should therefore be paired with readable surface design.

The results have measurement implications for researchers and practitioners. Search position and clickthrough are no longer sufficient visibility measures when agents summarize, cite, and select products. Future commerce analytics should include agent selection probability, citation probability, structured evidence completeness, task-level rank, and attention-retention indicators. These outcomes should be analyzed separately before being combined, because each one reflects a different part of the visibility process. A single composite score can summarize performance, but it can also hide whether a page is failing because agents cannot parse it or because humans do not find it compelling.

The broader implication is that agent-mediated commerce creates a dual-audience design problem. Product pages must communicate with shoppers and with systems that filter information before shoppers see it. The best-performing pages in this study did not merely add technical markup. They connected structured fields, evidence statements, policy clarity, and visible summaries to the user's information need. This suggests a practical but limited principle: pages should make product value explicit in forms that both humans and machines can use. Further research should test this principle on live platforms, with real agent traffic, real purchases, and more diverse product categories.

References

[1] Y. Bakos, "Reducing buyer search costs: Implications for electronic marketplaces," Management Science, vol. 43, no. 12, pp. 1676–1692, 1997.

[2] J. Alba, J. Lynch, B. Weitz, C. Janiszewski, R. Lutz, A. Sawyer, and S. Wood, "Interactive home shopping: Consumer, retailer, and manufacturer incentives to participate in electronic marketplaces," Journal of Marketing, vol. 61, no. 3, pp. 38–53, 1997.

[3] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[4] R. Burke, "Hybrid recommender systems: Survey and experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, pp. 331–370, 2002.

[5] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara, "Reputation systems," Communications of the ACM, vol. 43, no. 12, pp. 45–48, 2000.

[6] C. Dellarocas, "The digitization of word of mouth: Promise and challenges of online feedback mechanisms," Management Science, vol. 49, no. 10, pp. 1407–1424, 2003.

[7] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," Journal of Marketing Research, vol. 43, no. 3, pp. 345–354, 2006.

[8] N. Hu, P. A. Pavlou, and J. Zhang, "Overcoming the j-shaped distribution of product reviews," Communications of the ACM, vol. 52, no. 10, pp. 144–147, 2009.

[9] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, eds., Recommender Systems Handbook. Boston, MA: Springer, 2011.

[10] T. Gillespie, "The relevance of algorithms," in Media Technologies: Essays on Communication, Materiality, and Society (T. Gillespie, P. J. Boczkowski, and K. A. Foot, eds.), pp. 167–194, Cambridge, MA: MIT Press, 2014.

[11] L. D. Introna and H. Nissenbaum, "Shaping the web: Why the politics of search engines matters," The Information Society, vol. 16, no. 3, pp. 169–185, 2000.

[12] F. Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard University Press, 2015.

[13] S. Barocas and A. D. Selbst, "Big data's disparate impact," California Law Review, vol. 104, no. 3, pp. 671–732, 2016.

[14] R. V. Guha, D. Brickley, and S. Macbeth, "Schema.org: Evolution of structured data on the web," Communications of the ACM, vol. 59, no. 2, pp. 44–51, 2016.

[15] V. Karpukhin, B. Oguz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, "Dense passage retrieval for open-domain question answering," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 6769–6781, Association for Computational Linguistics, 2020.

[16] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive nlp tasks," in Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474, 2020.

[17] B. F. Török, "A machine learning analysis of brand selection by autonomous shopping agents using product attributes and machine legibility signals," Journal of Data Science, Predictive Analytics, and Big Data Applications, vol. 11, no. 5, pp. 1–27, 2026.

[18] R. Pieters and M. Wedel, "Attention capture and transfer in advertising: Brand, pictorial, and text-size effects," Journal of Marketing, vol. 68, no. 2, pp. 36–50, 2004.

[19] T. Teixeira, M. Wedel, and R. Pieters, "Emotion-induced engagement in internet video advertisements," Journal of Marketing Research, vol. 49, no. 2, pp. 144–159, 2012.

[20] A. Goldfarb and C. E. Tucker, "Privacy regulation and online advertising," Management Science, vol. 57, no. 1, pp. 57–71, 2011.

[21] B. F. Török, "A validated synthetic testbed for graph-based product affinity research in e-commerce," Transactions on Artificial Intelligence, Machine Learning, and Cognitive Systems, vol. 9, no. 8, pp. 1–25, 2024.

[22] J. R. Bettman, M. F. Luce, and J. W. Payne, "Constructive consumer choice processes," Journal of Consumer Research, vol. 25, no. 3, pp. 187–217, 1998.

[23] J. G. Lynch and D. Ariely, "Wine online: Search costs affect competition on price, quality, and distribution," Marketing Science, vol. 19, no. 1, pp. 83–103, 2000.

[24] W. W. Moe and P. S. Fader, "Dynamic conversion behavior at e-commerce sites," Management Science, vol. 50, no. 3, pp. 326–335, 2004.

[25] B. F. Török, "Modeling brand visibility in generative engine optimization (geo) using structured content signals in ai driven search environments," International Review of Machine Learning, Artificial Intelligence, and Applied Data Science, vol. 16, no. 2, 2026.

[26] S. H. Kukkuhalli, "Optimizing snowflake enterprise data platform cost through predictive analytics and query performance optimization," IJSAT-International Journal on Science and Technology, vol. 15, no. 4, 2024.

[27] N. V. Patel, "Selection bias in two-sided e-commerce marketplaces: A framework for propensity score matching implementation," Journal of Business Intelligence Systems and Computational Social Science Applications, vol. 11, no. 5, pp. 1–20, 2021.