Generative AI and Sales Productivity: Field Experiments in Online Retail*

Lu Fang† Zhe Yuan‡ Kaifu Zhang§
Dante Donati¶ Miklos Sarvary‖

May 29, 2026

Abstract

We quantify the short-term impact of Generative Artificial Intelligence (GenAI) on sales performance through a series of large-scale randomized field experiments involving millions of users and products at a leading cross-border online retail platform. Over 2023-2024, the platform integrated GenAI into seven consumer-facing business workflows spanning customer service, consumer-product matching, advertising, and seller services. We find that GenAI adoption increases sales in most workflows, with effects ranging from no detectable impact to 16.3%, depending on GenAI's marginal contribution relative to baseline firm practices. Across the four GenAI applications with positive sales effects, the implied annual incremental value is roughly $5—an economically meaningful impact given the retailer's scale and the early stage of GenAI adoption. The gains operate primarily through higher conversion rates rather than larger cart values, consistent with GenAI improving the shopping experience by reducing search, information, communication, and personalization frictions. Importantly, these effects are not associated with worse post-purchase outcomes, as product return rates and customer ratings do not deteriorate. Finally, we document substantial demand-side heterogeneity, with larger gains for less experienced consumers. Our findings provide novel, large-scale causal evidence on how GenAI shapes sales productivity in online retail, highlighting both its immediate value and broader potential.

Keywords: AI, Consumer Experience, Field Experiments, Productivity, Retail Platforms, Sales
JEL codes: C93, D24, L81, O33

*The authors thank Kinshuk Jerath, Carl Mela, and Andrey Simonov for valuable feedback, as well as seminar and conference participants at Duke University, Imperial Business School, BIG AI Conference, MSI AI Forum, TSE Digital Economics Conference, AIML Conference, Operational Innovation Network Summit, University of Rochester, UC Davis, TUM Workshop on Generative AI in Marketing, UCL School of Management, Columbia Business School, Business & Generative AI Conference, and Zhejiang University School of Management for helpful comments and discussions. The authors thank the partner e-commerce platform for its support with this project and are grateful to its employees for patiently answering many of our questions. The paper only reflects the opinion of the authors, and not the opinion of any organizations. All rights reserved. All remaining errors are the responsibility of the authors. Lu Fang acknowledges financial support from the National Natural Science Foundation of China [Grants 72501258 and 72192803], the National Social Science Fund of China [Grant 22&ZD081]. Zhe Yuan acknowledges financial support from the National Key Research and Development Project [Grant 2024YFB3312900], the National Natural Science Foundation of China [Grants 72141305 and 72203202], the Fundamental Research Funds for the Central Universities. Kaifu Zhang, Dante Donati, and Miklos Sarvary have no funding to report. Lu Fang and Zhe Yuan served as consultants for the partner platform during the duration of the project. Kaifu Zhang was an employee of the partner platform during the duration of the project. Dante Donati and Miklos Sarvary have no conflicts of interest to report.

†Zhejiang University of Finance & Economics, fl_fanglu@zufe.edu.cn
‡Corresponding author, Zhejiang University, yyyuanzhe@gmail.com
§Partner Company, zhangkaifu314@gmail.com
¶Columbia University, dd3137@gsb.columbia.edu
‖Columbia University, miklos.sarvary@columbia.edu

"*The next industrial revolution has begun*," Nvidia Chief Executive Officer Jensen Huang said. *"AI will bring significant productivity gains to nearly every industry and help companies be more cost- and energy-efficient, while expanding revenue opportunities." Nvidia Stock Surges as Sales Forecast Delivers on AI Hopes*, Bloomberg, May 22, 2024.

1 Introduction

The rapid diffusion of Generative Artificial Intelligence (GenAI) has sparked growing interest in its potential to improve productivity across sectors of the economy (Acemoglu, 2025). At the same time, investors and industry practitioners have increasingly questioned whether large AI investments will translate into sustained business returns.[1] A growing literature shows that GenAI can enhance performance in a range of worker-level tasks, including software development, customer support, education, law, and other professional services (e.g., Brynjolfsson et al., 2025; Noy and Zhang, 2023; Cui et al., 2026; Dell'Acqua et al., 2026; Choi et al., 2023; Kestin et al., 2025). Yet, despite rapid adoption, there is still limited causal evidence on whether GenAI improves sales performance in real-world, consumer-facing business workflows. This question is especially important in online retail platforms, which, over the past thirty years, have reduced consumers' search, information, and communication costs throughout the customer journey (Belleflamme and Peitz, 2021).

We document three main findings. First, most GenAI deployments improve sales, but the effects vary substantially across workflows, ranging from no detectable impact to gains of up to 16.3%, with the largest improvements observed in customer service and search-related applications. Aggregating across the four GenAI applications with positive sales effects, we estimate an annual incremental value of approximately $5 per consumer. Second, the evidence suggests that GenAI improves sales primarily by increasing purchase incidence rather than spending intensity: across several workflows, engagement and conversion rates rise, while average cart values remain largely unchanged, consistent with an enhanced consumer experience. We also find no evidence that these gains come at the expense of ex post consumer satisfaction, as return rates and customer ratings do not deteriorate following GenAI adoption. Third, the effects are heterogeneous across market participants, with less experienced consumers benefiting disproportionately more from GenAI-powered workflows. Taken together, these results show that GenAI can generate economically meaningful improvements in online retail performance, partly by further reducing frictions, while also revealing substantial heterogeneity in where these gains arise and who benefits from them.

A key distinction of our study is that it examines GenAI in real-world, consumer-facing retail workflows, focusing on front-office sales and related business outcomes rather than back-office worker-level task efficiency alone (e.g., Brynjolfsson et al., 2025; Noy and Zhang, 2023; Cui et al.,

[1] See, for example, recent articles by *Sequoia Capital*: www.sequoiacap.com/article/ais-600b-question; and *The Economist*: www.economist.com/leaders/2025/09/11/what-if-the-3trn-ai-investment-boom-goes-wrong.

2026). Using seven large-scale field experiments within a single online retail platform, we identify the short-run impact of GenAI under a common organizational environment, holding other observed inputs and prices fixed. Because all experiments take place within the same firm, this setting allows us to compare GenAI across workflows that differ in function and baseline conditions without conflating those differences with variation in implementation environments across studies. The results point to heterogeneous gains across applications and users, and are consistent with GenAI improving consumer experience by reducing search, information, and personalization frictions, though our design does not separately identify those channels from other possible mechanisms.

We partnered with a leading cross-border e-commerce platform to identify and quantify the sales productivity gains from GenAI adoption in online retail. The platform enables consumers worldwide to purchase directly from manufacturers at competitive prices.[2] Between September 2023 and June 2024, the firm deployed GenAI solutions across seven business workflows: (1) Pre-sale Service Chatbot, (2) Search Query Refinement, (3) Product Description Generation, (4) Marketing Push Message Creation, (5) Google Advertising Title Optimization, (6) Chargeback Defense, and (7) Live Chat Translation. These workflows span three broader functional areas: (i) consumer and seller services (1, 6, 7); (ii) consumer–product matching (2, 3); and (iii) advertising and promotion (4, 5). Each workflow corresponds to a distinct stage of the customer journey, allowing us to assess GenAI's sales productivity impact across a wide range of retail operations.

Our setting involves multiple experiments applying GenAI to distinct business workflows. Several features enhance comparability across these experiments: all applications were developed and deployed by the same technical team and operated within the same firm under similar organizational and competitive conditions. Despite this common implementation environment, the effect of GenAI on workflow performance is likely to depend on its marginal contribution relative to baseline conditions, which differ across workflows. In each case, the control group reflects the firm's standard practices prior to GenAI adoption. For example, in the Pre-sale Service Chatbot workflow, the control group received no live or interactive pre-sale customer support; in Search Query Refinement, the control condition relied on standard machine-learning-based search algorithms with only lexical translation; and in most other workflows, the benchmark was human input.

Within each workflow, the GenAI deployment was evaluated through a large-scale randomized field experiment that compared the GenAI-enhanced workflow in the treatment group to a baseline version used in the control group. Notably, baseline workflows often included automation or human input but did not incorporate GenAI technologies. The treatment condition differed solely through the integration of GenAI, while prices, other inputs and the broader workflow environment remained constant across conditions. Randomization occurred at the level of consumers—and,

[2]The platform connects hundreds of thousands of predominantly small-business sellers with hundreds of millions of active buyers across more than 100 countries and regions. The platform supports around 20 languages, providing localized services that facilitate global communication and accessibility.

in one case, products—with minimal overlap (less than 1%) across experiments, enabling causal identification of the effects of GenAI on consumer behavior. For five of the seven experiments, we obtained detailed consumer- or product-level engagement and transaction data, including views, clicks, number of orders, expenditure, conversions, product returns, and customer ratings. For analysis, we aggregate these data to the consumer level (and to the product level in one experiment) and leverage consumer, seller, and product characteristics to study treatment-effect heterogeneity. Our primary outcome is sales value, measured by total consumer expenditure. We also examine conversion rates, a standard measure of consumer experience in online retail; cart value, which captures changes in purchase intensity among consumers who make a purchase; and returns and ratings, reflecting customer decision quality and ex post satisfaction.

Our results reveal that most GenAI deployments yield economically significant short-term gains in sales, though the magnitudes differ across workflows. Across the five processes with detailed data, sales effects range from no detectable impact in the advertising workflows to gains of up to 16.3% in the Pre-sale Service Chatbot. In the Search Query Refinement and Product Description workflows, the effects are smaller, typically in the range of 2–3%, yet still substantial for a platform of this scale and maturity. In percentage terms, these effects are smaller than estimates from GenAI applications in worker-level tasks, ranging from about 15% in customer support to 55.8% in coding (Brynjolfsson et al., 2025; Peng et al., 2023). But the comparison should be interpreted cautiously: we measure downstream sales outcomes in a mature retail platform, rather than worker-level efficiency such as task completion time, so even a 2–3% increase can imply economically meaningful gains when applied at scale. Back-of-the-envelope calculations based on the four deployments with detailed transaction data and positive effects suggest that these GenAI applications generate an annual incremental value of approximately \$4.6–\$5.2 per consumer. These effects correspond to roughly 5.5–6.2% of the per-user revenue growth observed in global e-commerce between 2023 and 2024. We also document notable improvements in workflows without granular data: a 15% increase in success rate for Chargeback Defense, and a 5.2% increase in self-reported consumer satisfaction from Live Chat Translation.

Taken together, these results show that GenAI generates sizable gains in targeted workflows and meaningful effects for a large, mature retailer, with further potential as adoption broadens and targets revenue-critical processes. For example, while in 2023 the platform applied GenAI to only a handful of workflows, by 2024 it had expanded to more than 40 applications and by 2025 to over 60. Meanwhile, API calls to proprietary GenAI tools increased twentyfold between 2024 and 2025, reflecting the rapid scaling of GenAI adoption across the platform. The long-run impact will ultimately depend on equilibrium forces, specifically whether complementarities across workflows amplify these gains or industry-wide adoption offsets them through intensified competition.

We next study the potential mechanisms underlying the sales effects. Across workflows, sales

gains are accompanied by higher conversion rates—and, where available, by improvements in intermediate engagement metrics such as click-through rates—while cart values and purchase intensity conditional on purchase remain largely unchanged. Conversion rates increase by 1–22%, with little evidence of changes in spending intensity among buyers. Importantly, we also find no deterioration in the post-purchase outcomes available to us: product return rates and customer ratings do not decline following GenAI adoption, and in some cases returns fall or ratings improve. Taken together, these results suggest that GenAI primarily affects purchase incidence rather than spending conditional on purchase, while providing no evidence in our measured post-purchase outcomes of lower ex post consumer satisfaction.

The overall pattern is consistent with GenAI affecting sales through both informational and persuasive channels, with their relative importance differing across workflows. In some applications, especially Search Query Refinement and Live Chat Translation, the most natural interpretation is friction reduction: GenAI improves semantic understanding, matching, or communication in ways that help consumers and service agents interact more effectively. In others, such as Pre-sale Service Chatbot, Product Description, Marketing Push Message, and Google Advertising Title, GenAI may operate through both channels: providing richer and more relevant information on the one hand, while also making content more salient, polished, or engaging on the other hand. These channels are especially plausible in our multilingual cross-border setting, where search, information, and communication frictions are naturally high. Although our data do not allow us to separately identify the contribution of each channel, we quantify their combined effect on workflow-level sales performance.

Finally, we examine heterogeneity in treatment effects across buyers, sellers, and products. A natural hypothesis is that GenAI may generate larger gains for participants who either face greater frictions to begin with or respond more strongly to changes in consumer-facing content, such as less experienced buyers, smaller and newer sellers, and products that are harder to find or evaluate. We find statistically robust heterogeneity on the demand side: newer, less active, and lower-spending consumers benefit disproportionately more from GenAI-powered workflows. On the supply side, point estimates often suggest larger gains for smaller and newer sellers, although these differences are imprecisely estimated. Heterogeneity across product groups is more mixed and context-dependent.

This paper contributes to the emerging literature on the economic effects of GenAI. Existing work shows that GenAI can improve performance in a range of individual tasks, including writing, coding, marketing, and legal analysis, but most of that evidence comes from worker-level settings and emphasizes supply-side outcomes such as time savings or task completion (Noy and Zhang, 2023; Peng et al., 2023; Dell'Acqua et al., 2026; Cui et al., 2026). By contrast, we study large-scale GenAI deployment within a single firm and evaluate its effects using demand-side business outcomes. In doing so, we provide evidence on whether GenAI creates measurable value through

higher sales and conversion, rather than only through labor-saving or efficiency effects. Our setting also allows us to compare multiple workflows within the same organizational environment and to study heterogeneity across users with different levels of experience, complementing prior work on the uneven effects of GenAI across tasks and users (Calvino et al., 2025; Noy and Zhang, 2023; Brynjolfsson et al., 2025; Hui et al., 2024; Otis et al., 2024).

The paper also contributes to the literature on how digital platforms shape consumer decisions by reducing frictions and, in some settings, by influencing the content consumers see. A large body of work shows that platforms can improve outcomes by reducing information asymmetries, search costs, and targeting frictions through reputation systems, ranking algorithms, filtering tools, personalization, and targeted recommendations (Jin and Kato, 2006; Tadelis, 2016; Cabral and Hortaçsu, 2010; Dinerstein et al., 2018; Ursu, 2018; Chen and Yao, 2017; Fradkin, 2017; Yoganarasimhan, 2020; Blake et al., 2015; Sun et al., 2024). More broadly, some consumer-facing technologies may affect outcomes not only by improving information and matching, but also by making platform content more salient, polished, or engaging, consistent with classic informative and persuasive views of advertising (Bagwell, 2007; Goldfarb and Tucker, 2011). We extend this literature to GenAI by showing how it affects several consumer-facing retail workflows along the customer journey, increasing conversion in ways consistent with improved matching and a smoother consumer experience, while also leaving room for persuasive channels that may matter more in some applications than in others.

The remainder of the paper is structured as follows. Section 2 describes the study setting, experimental design, and data. Section 3 presents the main results. Section 4 examines heterogeneity in treatment effects. Section 5 concludes. The Appendix provides additional details and results.

2 Study Setting, Experimental Design, and Data

2.1 Study Setting

The seven field experiments analyzed in this paper were conducted over roughly nine months, from September 2023 to June 2024. The company's GenAI initiatives, however, began earlier in 2023, with initial efforts devoted to model training, implementation planning, and experimental preparation. The selection of workflows for GenAI re-engineering was not systematic but instead reflected managerial judgment, with platform managers prioritizing workflows viewed as most promising in terms of technical feasibility, organizational cost, and potential business value. The selected workflows cover several core modules of e-commerce operations, including customer service, consumer-product matching, advertising, and seller service. Table 1 provides an overview of these workflows, highlighting the business need in each case and the specific way GenAI was integrated.

Table 1 also helps conceptualize how GenAI may affect retail sales performance across work-

Table 1: Business Workflows Re-engineered with Generative AI

	Functional Area	Business Workflow	Business Needs/Objectives	GenAI Capability	Description of GenAI Application
1	Customer Service	Pre-sale Service Chatbot	Addressing each individual service request, providing unique, accurate, and content-rich answers.	AI agent	Deploying a GenAI-powered, 24/7 customer service chatbot that can respond to idiosyncratic consumer inquiries in all languages.
2	Consumer-product Matching	Search Query Refinement	Accurately decoding and translating the latent demands behind multilingual consumer search queries to improve consumer-product match.	Translation, content comprehension and generation	Using GenAI to improve consumers' demand expression by understanding, refining and translating their search queries, thus enhancing the matching accuracy of the search algorithm.
3	Consumer-product Matching	Product Description	Creating comprehensive, structured product descriptions tailored to diverse linguistic preferences and cultural norms (e.g., currently, nearly half of the self-sold products have no or limited description).	Content recognition, comprehension and generation	Using GenAI to produce comprehensive and structured textual descriptions for the product detail page's description module, adapted to each market.
4	Advertising	Marketing Push Message	Individual targeting of hundreds of millions of users with customized messages.	Content comprehension and generation	GenAI allows the generation of millions of messages, thereby enhancing the personalization of messages for precision marketing.
5	Advertising	Google Advertising Title	Creating product advertisement titles that closely match user interest and demands.	Content optimization and generation	Using GenAI to optimize product titles for Google ads for better user interest and engagement.
6	Seller Service	Chargeback Defense	Streamlining the complicated process in a cross-border context with language barriers and diverse regulations and customs (e.g., over half of chargeback disputes go unaddressed by sellers).	AI agent	Developing a GenAI-driven agent that offers a one-stop, automated solution for sellers to streamline the intricacies of chargeback defense.
7	Customer Service	Live Chat Translation	Delivering native-language customer services to a diverse, multilingual consumer base	Real-time translation	Integrating GenAI into the platform's core English customer service process to provide real-time translation for all languages.

flows. Broadly speaking, GenAI may operate through both informational and persuasive channels, consistent with classic work on information and advertising in markets (Stigler, 1961; Nelson, 1974). Bagwell (2007) provides a more recent comprehensive review. On the informational side, GenAI may improve semantic understanding, translation, responsiveness, and the richness and relevance of product information. In online retail, these improvements may lower search costs, reduce information asymmetries, improve matching, and facilitate communication between consumers, sellers, and service agents. This mechanism is especially relevant in multilingual and cross-border settings, where search, information, and communication frictions are naturally high. On the persuasive side, GenAI may affect outcomes through consumer-facing text and communication that is more salient, polished, or engaging, that increases the perceived professionalism or credibility of the platform's communication, or that generates short-run novelty effects that raise attention and purchase propensity.

These channels are not mutually exclusive, and their relative importance is likely to differ across applications. This is especially true across the seven workflows we study. For instance, Search Query Refinement and Live Chat Translation are most naturally expected to operate through friction reduction, because they improve matching or communication without materially changing the persuasive content seen by consumers. Chargeback Defense is similarly best understood as reducing seller-side frictions in a complex, language-intensive process. By contrast, Pre-sale Service Chatbot, Product Description, Marketing Push Message, and Google Advertising Title may affect outcomes through either informational or persuasive channels, since these applications directly shape the content consumers see. In these settings, GenAI may reduce search costs and information asymmetries by providing richer, more relevant, or more structured information, while also influencing behavior by making content more prominent, well-crafted, or compelling.

Importantly, these channels are suggestive interpretations rather than well-identified mechanisms. Our empirical setting does not allow us to separately and precisely identify the informational and persuasive channels through which GenAI may affect outcomes. Instead, our goal is to quantify their combined effect on sales when GenAI is added to an existing workflow. In each experiment, the treatment condition differs from the control condition only through the integration of GenAI, while prices, other inputs and the broader workflow environment remain unchanged. Although GenAI itself requires development, integration, and inference resources, there were no concurrent changes to the workflows other than the addition of GenAI that would confound the comparison. We therefore interpret our estimates as the short-run impact of GenAI on workflow-level sales performance, holding other aspects of the workflow as fixed as possible.

2.2 Experimental Design

To test the sales productivity gains of the seven GenAI applications, the firm conducted a series of large-scale, randomized field experiments. Six of these experiments were executed at the consumer level, with participating consumers randomly assigned to either treatment or control groups. The only exception was the Google Advertising Title, which was conducted at the product level, where a subset of products selected for Google ads was randomly divided into treatment or control groups. Consumer overlap across experiments was minimal (below 1%). In all cases, for each consumer or product, the treatment status remained fixed throughout the experimental period. The key distinction between treatment and control was that the treatment group was exposed to workflows re-engineered with GenAI, whereas the control group continued under the platform's standard practices without GenAI integration.

A potential concern is that treatment exposure for some users may affect the experiences or behavior of others, creating possible SUTVA violations. In our setting, such spillovers could arise if treatment-induced demand shifts alter marketplace conditions, such as the composition of search results, or if treated customers share information with control customers. We expect these concerns

to be limited in our context. Although we do not observe the exact share of platform users exposed in each workflow, our back-of-the-envelope calculations suggest that treated users represented only a small fraction of the platform's relevant user base.[3] Moreover, most experiments were short, and, except for Pre-sale Service Chatbot, AI operated mostly in the background and affected private user experiences, limiting awareness and diffusion.

The total size of the subject pool varied greatly between experiments, with the smallest experiment having 30 thousand subjects while the largest containing up to 13 million subjects. Most of the experiments featured an equal distribution between the treatment and control subjects, with each group comprising approximately half of the total sample. The exceptions are Pre-sale Service Chatbot and Live Chat Translation, where the treatment group consumers comprised two-thirds of the total sample. Below, we provide details on the experiments related to all seven business workflows, with a summary of key features presented in Table 2. Appendix A presents illustrative user interfaces and examples for each workflow.

Pre-sale Service Chatbot The experiment, conducted over a two-month period from September to October 2023, included a random sample of 44 thousand consumers who initiated pre-sale customer service requests for the platform's self-sold products during the experimental period. These consumers were randomly divided into treatment and control groups. Consumers in the control group received the platform's automated response service, which delivered a pre-programmed standardized notification indicating that customer service was unavailable. This auto-response condition reflects the platform's standard operating practice of prioritizing a limited number of human agents for post-sale rather than pre-sale support for self-sold products, given that pre-sale inquiries are generally less urgent. This setup also mirrors the constraints faced by many third-party sellers on the platform, particularly smaller-scale sellers, who lack the capacity to provide real-time multilingual customer service. By contrast, consumers in the treatment group were supported purely by GenAI-powered chatbots.[4]

Search Query Refinement The experiment comprised three sub-experiments, each targeting consumers using different languages: Arabic, Japanese, and Polish. These languages were chosen because they are less commonly used on the platform and have historically been underserved by the platform's traditional translation of search queries.[5] The sub-experiments were launched at different points between May and June 2024, each lasting nine days. During each period, a random

[3]Because we do not observe the total number of users in each workflow across the entire platform, we approximate exposure using the number of treated purchasers and publicly available information on the platform's annual active shoppers. These calculations suggest that treated purchasers account for less than 3% of the platform's active shoppers within the corresponding experimental time window across workflows.

[4]The platform did not provide chat transcripts or message-level content because of confidentiality and consumer-privacy restrictions. Our analysis therefore uses treatment assignment and transaction outcomes rather than the content of chatbot conversations.

[5]On our focal platform, the search algorithm initially translates multilingual queries into English to facilitate matching with product and seller information stored in English.

Table 2: Summary Descriptions of the Experiments

	Business Workflow	Time Frame	Sample Size	Control	Treatment	Data Availability	Product Sold By
1	Pre-sale Service Chatbot	Two months from Sep. to Oct. 2023	44,614 consumers	Pre-programmed auto response indicating no customer service	GenAI Agent	Yes	Platform
2	Search Query Refinement	Three nine-day sub-experiments from May. to Jun. 2024	1,849,382 consumers	Basic query translation with no semantic comprehension	GenAI-translated queries with semantic comprehension	Yes	Sellers & Platform
3	Product Description	Five one-week sub-experiments in Dec. 2023	4,772,937 consumers	Human-generated descriptions	AI-generated descriptions on top of those created by humans	Yes	Platform
4	Marketing Push Message	One day in Dec. 2023	13,715,528 consumers	Human-generated standardized messages	GenAI generated a large and diverse set of messages	Yes	Sellers & Platform
5	Google Advertising Title	Twelve days in Jan. 2024	1,244,016 products	Human-generated ad titles	GenAI-optimized ad titles	Yes	Sellers & Platform
6	Chargeback Defense	Two months from Oct. to Dec. 2023	About 30 thousand consumers	Human agent	GenAI agent	No	Sellers & Platform
7	Live Chat Translation	One month in Oct. 2023	About 0.2 million consumers	Filipino agent without translation assistance	Filipino agent with GenAI real-time translation assistance	No	Sellers & Platform

[1] In column "Product Sold By", "Platform" denotes products procured and sold directly by the platform—i.e., the platform's self-sold products. "Sellers & Platform" indicates that the experiment included products sold both by the platform itself and by third-party sellers on the platform.

subset of consumers conducting searches was assigned to the experiment. These consumers were randomly divided into two groups, yielding a total sample of approximately 2 million consumers across all sub-experiments. In the control group, consumer search queries were subject only to basic translation without semantic comprehension. In the treatment group, GenAI was deployed to translate queries by comprehending their underlying intent and refining them to improve semantic accuracy and clarity.

Product Description The experiment comprised five sub-experiments, each involving consumers who spoke English, Spanish, French, Portuguese, or Korean, which are among the most widely used languages on the platform. All sub-experiments ran for one week in December 2023, with staggered start dates. GenAI was employed to create multilingual, textual product descriptions for a predetermined product set of approximately 45,000 randomly selected platform self-sold products spanning a broad range of categories.[6] According to our partner company, self-sold products are primarily sourced from Chinese vendors, who typically provide image-based product presentations with limited Chinese text embedded in the images. While such image-based content aligns with Chinese consumer preferences, global consumers are more accustomed to text-based descriptive bullet points, such as the "About this item" section on Amazon. Consequently, nearly half of the self-sold products either lack textual descriptions or contain only minimal textual information. During each sub-experiment period, a random subset of consumers who clicked into the product detail pages of the selected products entered the experiment and were evenly split into treatment and control groups, resulting in a total of approximately 5 million participants. Control group consumers viewed the original human-generated descriptions, whereas treatment group consumers were shown the GenAI-created descriptions on top of the original, human-created descriptions.[7]

Marketing Push Message The experiment took place over the course of approximately one month in December 2023. A random subset of consumers who received push notifications on their mobile devices entered the experiment and were randomly assigned to either control or treatment groups. Given the large scale of this experiment, we restricted our analysis to the first day, which contained 13 million consumers. On our partner platform, push messages were traditionally created by staff, requiring 1-2 employees several hours each month to produce a few dozen messages. Given the platform's hundreds of millions of consumers, this limited volume meant that many consumers received identical content, constraining the potential for personalized marketing. Accordingly, in the control group of our experiment, consumers primarily received uniform, human-generated marketing content, totaling roughly 2,000 distinct messages. By contrast, during the platform's initial risk-managed rollout, approximately 40% of consumers in the treatment group were exposed to AI-generated messages, while the remaining treatment-group consumers continued to receive

[6] On this platform, product descriptions refer to the text content in the description module of product detail pages that summarizes key features and selling points.

[7] For products without existing human-generated descriptions, control group consumers saw no description—as was historically the case for such products, and treatment group consumers saw the AI-generated descriptions only.

human-generated messages. This partial exposure generated nearly 2.7 million unique AI-generated message variants and thus far greater differentiation across individuals.

Google Advertising Title The experiment was conducted over twelve days in January 2024, with randomization occurring at the product level. The sample included 1.2 million products selected by the retail platform for advertising in the sponsored section of Google Shopping, representing a diverse set of categories. For Google ads, the quality of the advertisement title is critical: a well-crafted title not only increases product discovery by aligning with user search keywords but also enhances user clicks by incorporating appealing terms that drive consumer interest.[8] In our experiment, the control group retained the original product titles created by the sellers, while in the treatment group, titles used in the ads were optimized by GenAI based on seller titles.[9] A key distinction of this experiment is that the GenAI model was not fine-tuned specifically for the advertising context within the e-commerce domain. As a result, the generated titles may fail to emphasize product attributes most relevant to consumer search and purchase decisions, leading us to adopt a more agnostic stance regarding the expected treatment effect in this setting.

Chargeback Defense The experiment, conducted from late October to late December 2023, included over 30 thousand consumers. During this period, a random subset of consumers who initiated chargeback requests was assigned to the experiment and then randomly divided into two groups. Contesting chargeback disputes requires a broad skill set, including claim analysis, evidence collection, and persuasive defense writing, which is especially challenging in cross-border contexts characterized by language barriers and complex regulations and customs. As a result, more than half of chargeback disputes on the focal platform were left unaddressed by sellers. In the control group, consumer claims were initially addressed by sellers. If no action was taken, approximately 3-5 outsourced workers then intervened to resolve the claims. However, these employees could only handle a small fraction of cases elaborately, while most were processed using generalized templates that proved far less effective. In contrast, claims in the treatment group were initially managed by sellers and subsequently supported by GenAI agents.

Live Chat Translation The experiment was conducted over one month in October 2023 and involved approximately 0.2 million non-English-speaking consumers who contacted the platform's customer service for issues such as clarifying details of platform-level promotions or resolving disputes with sellers when no agreement was reached. Due to cost constraints, a large share of such requests is handled by customer service agents from the Philippines who provide assistance in English, as employing native agents for each market is roughly three times more expensive. During the experiment, non-English-speaking consumers who initiated inquiries to the platform's customer

[8] Many e-commerce platforms maintain libraries of such buzzwords which, based on historical data, are known to boost consumer click-through and conversion rates.

[9] When promoting products on Google Shopping, the platform also used the pricing and image information provided by sellers, and these factors remained unchanged across treatment and control groups in our experiment.

service were randomly split between treatment and control conditions. In the treatment group, consumers interacted with Filipino agents with real-time bidirectional GenAI translation support, while those in the control group engaged with Filipino agents without GenAI translation assistance.

2.3 Data and Estimation

We obtained comprehensive granular consumer- or product-level data for the first five of the seven experiments, allowing for in-depth analysis of sales gains. For the remaining two experiments—Chargeback Defense and Live Chat Translation—the platform could not provide granular data. In these cases, we rely on analyses conducted by the platform's internal data science team. These estimates complement our direct observations, offering a broader perspective on the impact of GenAI across various business areas (see "Data Availability" in Table 2).

For experiments conducted at the consumer level, we record each consumer's treatment status and observe their complete set of activities, including the number of product views (Views), product clicks (Clicks), product orders (Orders), and total expenditure on those orders (Sales).[10] For comparability across workflows, our main analysis focuses on sales, measured as total consumer expenditure. To shed light on the mechanisms underlying the sales effects, we additionally examine outcomes along the extensive and intensive margins. In particular, we analyze conversion rates, defined as a binary indicator for whether a consumer makes at least one purchase during the experimental period, and cart value, measured as average expenditure conditional on making a purchase. Conversion rates capture changes in purchase incidence and serve as a widely used proxy for consumer experience in online retail, while cart value reflects adjustments along the intensive margin. When available, we also analyze intermediate engagement metrics, such as click-through rates. For the product-level experiment, we collect the same outcomes at the product level. Table 3 reports summary statistics for these key variables across the five experiments for which granular data are available.

To compare mean sales, conversion rates, and cart values between the treatment and control groups, we use the following general empirical specification, adapted as needed for each experiment:

$$y_i = \beta \times Treat_i + \alpha_{c(i)} + \epsilon_i, \tag{1}$$

where i denotes the randomized unit (consumer or product), and y_i is the outcome. $Treat_i$ is the treatment indicator, which equals one if the consumer or product belongs to the treatment group and zero otherwise. $\alpha_{c(i)}$ denotes the cohort fixed effects. Specifically, in the Pre-sale Service Chat-

[10] In Search Query Refinement, product views represent the number of products a consumer browses on the search results page, which displays a summarized collection of products immediately after a query search. In Google Advertising Title, product views refer to the number of views of advertised products within the Google Shopping tab. In Search Query Refinement and Product Descriptions, product clicks capture the number of times consumers clicked into product detail pages. In Marketing Push Message, product clicks reflect consumer clicks on push notifications, while in Google Advertising Title, they indicate clicks on advertised products.

Table 3: Summary Statistics of Main Outcomes

	Mean	Standard Dev.	Median	Min	Max
Pre-sale Service Chatbot					
Sales	1.86	9.75	0	0	517
Conversion Rate	0.07	0.25	0	0	1
Cart Value	27.05	26.41	18.90	1.11	517
Search Query Refinement					
Views	313.36	615.02	125	1	105,883
Clicks	8.23	16.99	3	0	2,024
Orders	0.16	0.73	0	0	85
Sales	2.24	21.41	0	0	4,960
Conversion Rate	0.09	0.28	0	0	1
Cart Value	25.39	67.83	10.42	0.01	4,960
Product Description					
Clicks	1.98	2.06	1	1	173
Orders	0.06	0.30	0	0	23
Sales	0.51	4.56	0	0	2,942
Conversion Rate	0.04	0.21	0	0	1
Cart Value	11.62	18.51	7.06	0.59	2,942
Marketing Push Message					
Clicks	0.02	0.13	0	0	1
Orders	0.002	0.05	0	0	6
Sales	0.03	1.51	0	0	501
Conversion Rate	0.002	0.04	0	0	1
Cart Value	15.52	33.97	7.81	0.01	501
Google Advertising Title					
Views	19.36	82.38	5	2	12,033
Clicks	0.22	1.69	0	0	627
Sales	0.13	2.97	0	0	322
Conversion Rate	0.004	0.06	0	0	1
Cart Value	33.39	34.39	21.42	0.54	322

[1] "Views" refers to the number of product views. "Clicks" denotes the number of product clicks. "Orders" is the number of product orders. "Sales" represents the total expenditure on product orders. "Conversion rate" measures consumers' likelihood of making a purchase. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experiment period, and 0 otherwise. "Cart value" is the expenditure per consumer, conditional on the consumer making a purchase.

[2] In the Google Advertising Title experiment, the unit of observation is the product. The conversion rate is 1 if a product is purchased at least once during the experimental period, and 0 otherwise, while cart value refers to sales per product, conditional on the product being purchased.

bot and Google Advertising Title experiments, consumers or products entered the experiments on different days, we therefore control for entry-day cohort fixed effects. In the Search Query Refinement and Product Description experiments, multiple sub-experiments were conducted across different languages at varying times, we thus include entry-day-by-language cohort fixed effects. For the Marketing Push Message experiment, the sample spans only a single day, so no cohort fixed effects are included. Details on the model specification for each experiment are provided in Appendix C.

We estimate Equation (1) via OLS, adjusting the standard errors for heteroskedasticity. Under random assignment, β recovers the average treatment effect of GenAI adoption, expressed as

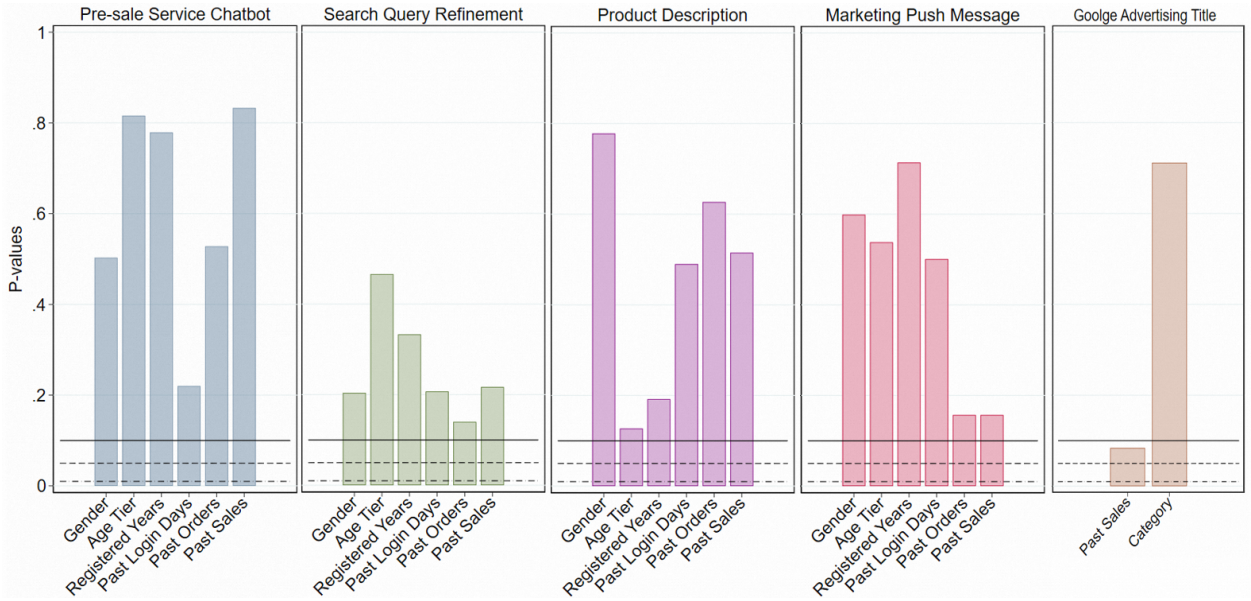
the absolute lift in outcomes. For all detailed experiments estimated with this equation, except for Marketing Push Message, assignment to the treatment group coincides with exposure to the GenAI-enabled workflow, so this estimate is also the treatment-on-the-treated (TOT) effect for the corresponding workflow. Marketing Push Message differs because only approximately 40% of consumers assigned to the treatment group were exposed to AI-generated messages during the platform's initial risk-managed rollout. For that experiment, β is therefore an intent-to-treat (ITT) effect of assignment to the treatment group; in the aggregate-gains calculation, we report this ITT estimate as a lower bound and an exposure-adjusted upper bound under the assumption that exposure within the treatment group was random. We also report results in percent lift, rescaling β by the control group mean. For sales, we use levels to address concerns regarding log transformations with zero outcomes (Chen and Roth, 2024). For conversions (a binary outcome), we primarily estimate a linear probability model and confirm that the findings are robust to logit specifications. We also estimate the model using pre-experiment covariates as controls, and the results remain consistent. Consumer overlap across experiments was minimal (less than 1%), and our findings are robust to excluding overlapping observations. This design allows treatment effects to be solely attributed to individual workflows, though it does not capture potential complementarities across GenAI applications.

To enrich our analysis, we obtained pre-experiment seller data for products included in the experiments (e.g., seller size measured by annual sales, operational years, and the number of sub-accounts linked to a seller's online store). These data enable us to examine seller-level heterogeneity in the three experiments involving products sold by both third-party sellers and the platform (see column "Product Sold By" in Table 2). This analysis is not applicable to the Pre-sale Service Chatbot and Product Description experiments, as both involve platform self-sold products sold by a limited number of platform-operated sellers, resulting in insufficient variation in seller characteristics for meaningful heterogeneity analysis. Additionally, we collect product characteristics, such as the concentration level of the associated category, price, and annual sales quantity, to explore product-level heterogeneity.

We also augmented the data with consumer demographics and pre-experiment shopping history (years of registration, activity level, purchase volume, etc.), which supported both the analysis of buyer-level heterogeneity and the verification of random group assignment in the experiments.[11] As confirmed by the covariate balance checks reported in Figure 1, we found no systematic significant pre-experiment differences across consumers between the control and treatment groups. Thus, the randomization process was effective at allocating comparable consumers and products into the two groups. Details on the covariate balance checks are presented in Appendix B.

[11] According to our research agreement with the partner platform, all consumers in our data are anonymous to ensure consumer privacy. We identify consumers by hashed IDs instead of knowing their actual names.

Figure 1: P-Values for Covariate Balance Checks Across Experiments



[1] This figure presents the p-values for the covariate balance checks across all experiments. For the first four experiments, which are conducted at the consumer level, six consumers' demographic and behavioral variables are examined. For the single experiment conducted at the product level (Google Advertising Title), the focus is on two key measurements of products, which are the product's historical sales and the distribution of product categories.

[2] The bar presents the p-values. The solid, dashed, and dash-dot lines indicate p-values of 0.1, 0.05, and 0.01, respectively.

[3] "Gender" is predicted by the platform based on consumers' past shopping behaviors. "Age Tier" is a 7-point scale from 1 (youngest) to 7 (oldest). "Registered Years" indicates the duration from the year of consumer registration to the year of experiment. "Past Login Days" represents the number of days consumers have logged into the platform in the 30 days prior to the experiment. "Past Orders" is the number of product orders in the 30 days prior to the experiment. "Past Sales" represents the total expenditure on product orders in the 30 days prior to the experiment. "Category" is the category associated with a product.

3 Main Results

3.1 Sales Impact by Workflow

Table 4 reports the estimated effects on sales, our primary outcome. Column (1) presents the absolute treatment effect, while column (2) reports the corresponding percentage change relative to the control group. As discussed in Section 2.3, these estimates correspond to TOT effects for workflows in which treatment assignment coincides with GenAI exposure, and to an ITT effect for Marketing Push Message because of partial exposure within the assigned treatment group. The results presented here do not account for differences in the frequency of consumer interactions with the workflows or in the duration of the corresponding experiments. For example, search affects nearly all consumers on a daily basis, whereas chatbot interactions occur less frequently. In Section 3.4, we provide adjusted measures that account for these differences. Additional results and robustness analyses are reported in Appendix C.

Table 4: Average Treatment Effects of GenAI Adoption on Sales Across Workflows

	Sales Productivity Impact (US $)			
	(1)	(2)	(3)	(4)
Business Workflow	Coefficient	% Change	Obs	Unit of Obs
Pre-sale Service Chatbot	0.274***	16.3%	44,614	Consumer
	(0.0970)			
Search Query Refinement	0.0648**	2.9%	1,849,382	Consumer
	(0.0314)			
Product Description	0.0104**	2.1%	4,772,937	Consumer
	(0.00417)			
Marketing Push Message	0.000402	1.6%	13,715,528	Consumer
	(0.000816)			
Google Advertising Title	-0.00602	-4.5%	1,244,016	Product
	(0.00534)			
Chargeback Defense†	15% defense success rate increase			
Live Chat Translation†	5.2% consumer satisfaction increase			

[1] "Sales" represents the total expenditure on product orders, in USD.
[2] Column (1) reports the estimated coefficients, with standard errors in parentheses. Column (2) reports % Change, calculated as the treatment effect divided by the control group mean. Column (3) reports the number of observations. *** p<0.01, ** p<0.05, * p<0.1.
[3] † For these experiments data are not available: we report findings estimated by the platform's internal data science team.

The table shows that most GenAI deployments are associated with sales gains, with substantial heterogeneity across workflows. The largest effect is observed in the pre-sale service chatbot workflow, where GenAI increases sales by 16.3% relative to the control group ($p < 0.01$; Column 2). In this setting, consumers in the treatment group interact with a GenAI-powered chatbot, whereas consumers in the control group receive an automated message indicating that real-time support is unavailable. This control condition reflects the platform's standard operating practice: human agents are prioritized for post-sale inquiries, leaving a substantial share of pre-sale inquiries for self-sold products without live assistance due to capacity constraints. This setting also mirrors the operating reality faced by many third-party sellers on the platform, particularly smaller-scale sellers who lack the capacity to provide real-time, multilingual customer support. Taken together, these features suggest that GenAI improves sales performance by performing tasks that are difficult for human agents to fulfill due to talent shortages or resource constraints.

One potential concern is that consumers in the control group may have been frustrated by the absence of assistance, mechanically reducing their likelihood of purchase and thus inflating the estimated treatment effects. To address this concern, we conducted additional experiments on the GenAI-powered chatbot. Appendix Table C1 reports a series of comparisons, including benchmarks against human agents and hybrid configurations that combine GenAI assistance with human escalation when needed. The results indicate that the GenAI chatbot delivers service quality comparable

to human customer support (Column 2). Moreover, integrating GenAI with human agents yields substantially larger effects: relative to the no pre-sale service condition, sales increase by 25% when GenAI assistance is combined with human escalation ($p < 0.01$; Column 3), pointing to strong complementarities between GenAI and human labor. Most importantly, comparing consumers who receive GenAI-assisted service with human escalation to those served exclusively by human agents shows that the former spend 11.5% more ($p < 0.1$; Column 4). We interpret this comparison as a conservative lower bound on GenAI's sales productivity impact in pre-sale customer support.

The remaining four workflows with detailed data reported in Table 4 exhibit more modest effects on sales, ranging from a negative and statistically insignificant estimate to gains of up to 3%. In particular, the Search Query Refinement application increases sales by 2.9% ($p < 0.05$), while automated Product Description generation raises sales by 2.1% ($p < 0.05$). Although smaller in magnitude, these effects are economically meaningful for a platform of this scale and maturity, especially given the near-universal exposure of consumers to search and product descriptions.

The Marketing Push Message workflow shows a positive yet not statistically significant improvement in sales (1.6%). This likely reflects the combination of a very low baseline conversion rate (only 0.16% of consumers make a purchase) and high variance in expenditures among converters, suggesting that broader implementation could provide sufficient power to detect treatment effects. Notably, however, this case exhibits significant increases in conversion rate and number of orders, which we discuss in detail later. It is also important to note that, for risk-mitigation purposes, only a subset of the treatment group was exposed to AI-generated messages in this experiment, which may attenuate the observed treatment effects under partial exposure.[12]

By contrast, the Google Advertising Title workflow exhibits a statistically insignificant negative effect. This pattern is consistent with the GenAI model not being fine-tuned to the advertising context within e-commerce, leading it to omit commercially salient keywords commonly used in ad titles. As a result, these ads may receive lower quality scores from Google, reducing their probability of winning auctions at a given bid. In addition, Google's advertising algorithm may identify and deprioritize AI-generated titles, further lowering their visibility. Consistent with this interpretation, Table C8 in the Appendix shows that ads with AI-generated titles receive fewer views and clicks than those with human-generated titles. Taken together, these results underscore the importance of domain-specific fine-tuning or retraining of foundation models when deploying GenAI in industry-specific tasks that require specialized knowledge (Deloitte, 2023).

For the final two business processes studied—Chargeback Defense and Live Chat Transla-

[12]The estimated effect is likely a lower-bound estimate. Because approximately 40% of treatment customers were exposed to AI-generated messages during the initial rollout, under the assumption that exposure within the treatment group was random, rescaling the effect by the exposure rate gives \$0.000402 / 0.40 = \$0.001005 per exposed customer, implying a 4% lift relative to the control group.

tion—we do not observe granular transaction-level data and instead rely on the platform's internal metrics and analyses, which do not report statistical significance. While these outcomes are not directly comparable to the sales effects discussed above, they nevertheless point to substantial improvements in two operational dimensions. Specifically, Chargeback Defense is associated with a 15% increase in defense success rates, and Live Chat Translation increases consumer satisfaction by 5.2%. Both effects point directionally to successful implementation of the GenAI applications.[13]

These results provide evidence on the potential of GenAI to improve front-office sales performance in online retail. A key distinction of our study is its focus on downstream revenue outcomes rather than worker-level gains, which makes our estimates not directly comparable to those from studies emphasizing labor productivity. Among the seven deployments we examine, most deliver measurable performance gains, showing that GenAI can generate substantial improvements in retail outcomes under real-world operating conditions. Overall, the evidence points to a positive but heterogeneous impact of GenAI along the customer journey.

The variation in effects across workflows is unlikely to reflect differences in implementation quality, which was comparable across applications. Instead, it appears to arise from differences in the marginal contribution of GenAI relative to baseline conditions. In each case, the control group reflects the firm's standard operating practices prior to GenAI adoption. For example, in the Pre-sale Service Chatbot workflow, the control group received no pre-sale customer support; in Search Query Refinement, it relied on standard machine-learning-based search algorithms with lexical translation into English; and in most other workflows, the benchmark was human input. The results therefore point to genuine heterogeneity in where GenAI is most effective. Customer-support applications, such as Pre-sale Service Chatbots, generate the largest improvements; search and product-discovery tasks yield more modest gains; and advertising-related applications exhibit no statistically significant effects.[14] Taken together, these findings highlight both the role of baseline conditions in shaping treatment effects and the differential effectiveness of GenAI across functional areas.

3.2 Post-purchase Outcomes

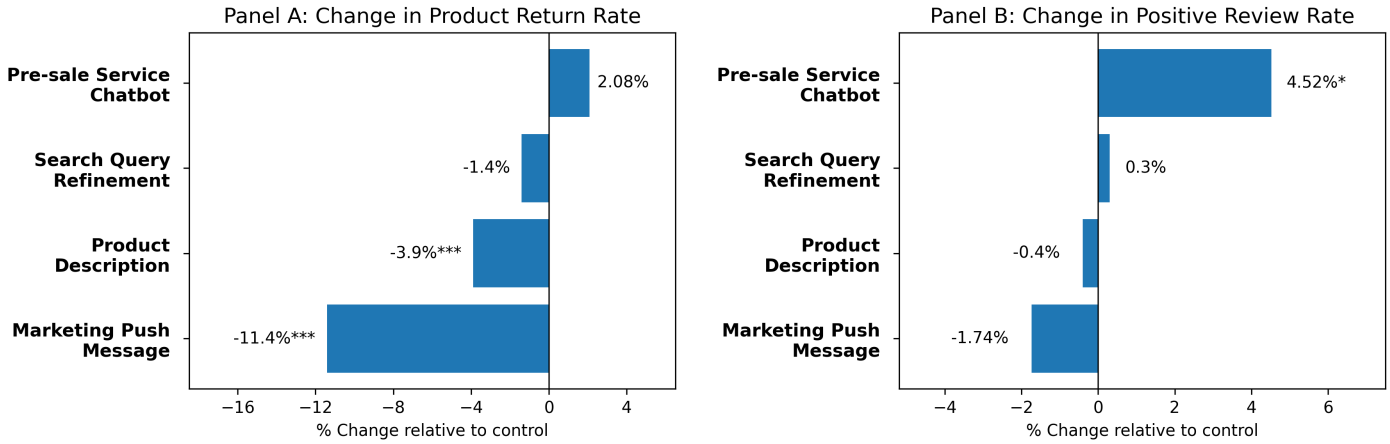
A natural concern with deploying GenAI in customer-facing processes is that, while it may provide richer, more relevant pre-purchase information and improve ex ante decision quality, it may also introduce hallucinations or provide overly persuasive content that misleads consumers, thereby reducing ex post consumer satisfaction. In our case, increases in revealed purchase behavior may not map cleanly into welfare gains, as higher sales could reflect distorted choices rather than improved consumer surplus. To assess this possibility, we examine two post-purchase outcomes that capture

[13] In the case of Chargeback Defense, additional gains not captured in our calculations may arise from cost reductions, as the GenAI-enabled workflow eliminates the need for manual intervention.

[14] In the Marketing Push Message workflow, features of the experimental design may also affect measured impacts, as only a subset of treated users was exposed to GenAI, likely attenuating estimated effects.

consumers' evaluations of their purchase decisions. The first is the product return rate, defined as the share of orders that are returned. The second is the positive review rate, defined as the share of rated orders that receive four- or five-star ratings on a five-point scale.[15]

Figure 2: Average Treatment Effects of GenAI Adoption on Product Returns and Reviews



[1] "Return Rate" is defined as the share of orders that are returned within a year after the purchase. "Positive Review Rate" is defined as the share of rated orders that receive four- or five-star ratings on a five-point scale. The estimation of Return Rate is conditional on consumers with orders, while the estimation of Positive Review Rate is conditional on consumers who provide ratings. Such data are not available for the Google Advertising Title workflow.
[2] Full regression results are reported in Table C9. *** p<0.01, ** p<0.05, * p<0.1.

As shown in Figure 2 (full regression results are provided in Table C9), we find no evidence that the observed sales gains come at the expense of measured ex post consumer satisfaction. Product return rates and customer ratings do not deteriorate following GenAI adoption. Notably, return rates decline significantly for Product Descriptions and Marketing Push Messages (by 3.9% and 11.4%, respectively; $p < 0.01$), consistent with improvements in both ex ante decision quality and ex post consumer satisfaction in these workflows. We also observe a positive effect on customer ratings for the Pre-sale Service Chatbot (a 4.5% increase; $p < 0.10$). In all other cases, the estimated effects are statistically insignificant. We therefore interpret these post-purchase outcomes as useful but imperfect evidence that the sales gains are not accompanied by lower revealed consumer satisfaction.

3.3 Potential Mechanisms

We investigate two potential drivers behind the observed impact on sales. The first is an increase in purchase incidence, captured by an increase in the probability that a consumer completes a purchase at the extensive margin. We measure this pattern using conversion rates, a widely used industry metric that reflects consumers' revealed satisfaction and perceived friction during the ex ante search

[15]The platform does not use auto-filled or pre-populated ratings; ratings are voluntarily submitted by consumers after purchase. We measure the positive review rate conditional on rated orders. For returns, we track purchases over a one-year post-purchase horizon, which is long relative to the platform's typical return window. Nevertheless, returns may still understate dissatisfaction if high transaction costs lead some consumers not to return products.

and evaluation process in e-commerce. The second pattern is a change in the behavior of consumers who make a purchase, reflected in adjustments to spending intensity and order composition. In this case, GenAI may influence consumers' willingness to pay, alter the products that consumers see, or shift demand toward products at different price points. We assess this pattern using cart value, defined as expenditure per purchasing consumer, and order intensity, defined as the number of orders per purchasing consumer, both measured conditional on consumers making at least one purchase.

Table 5: Average Treatment Effects of GenAI Adoption on Conversion Rate and Cart Value

	Conversion Rate			Cart Value ($)		
	(1)	(2)	(3)	(4)	(5)	(6)
Business Workflow	Coefficient	% Change	Obs	Coefficient	% Change	Obs
Pre-sale Service Chatbot	0.0131***	21.7%	44,614	-1.264	-4.5%	3,076
	(0.00252)			(1.036)		
Search Query Refinement	0.00101**	1.2%	1,849,382	0.370	1.5%	163,381
	(0.000411)			(0.334)		
Product Description	0.000554***	1.3%	4,772,937	0.0944	0.8%	210,155
	(0.000187)			(0.0805)		
Marketing Push Message	0.000048**	3.0%	13,715,528	-0.206	-1.3%	22,425
	(0.0000218)			(0.454)		
Google Advertising Title	-0.000090	-2.3%	1,244,016	-0.784	-2.3%	4,811
	(0.000111)			(0.992)		

[1] For workflows 1-4, "Conversion Rate" measures consumers' likelihood of making purchases. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experiment period, and 0 otherwise. "Cart Value" refers to the expenditure per consumer, conditional on the consumer making a purchase. For workflow 5, "Conversion Rate" is 1 if a product is purchased at least once during the experimental period, and 0 otherwise. "Cart Value" refers to the sales per product, conditional on the product being purchased.

[2] Columns (1) and (4) report the estimated coefficients, with standard errors in parentheses. Columns (2) and (5) report % Change, calculated as the treatment effect divided by the control group mean. Columns (3) and (6) report the number of observations. *** p<0.01, ** p<0.05, * p<0.1.

Table 5 reports the results on conversion rates and cart value, with additional evidence on order intensity provided in Appendix C. Across workflows, we document economically and statistically significant increases in conversion rates ranging from 1% to 22% (Column 2), which translate directly into higher output as measured by sales. By contrast, we find no evidence of effects along the intensive margin. Columns 4 and 5 show that cart value, as measured by the average order value among consumers who make at least one purchase (or among products purchased at least once, in the case of Google Advertising Title), remains unchanged following GenAI adoption. We likewise find no significant changes in order intensity, defined as the number of orders placed conditional on making at least one purchase (Appendix C). Taken together, these results indicate that GenAI primarily drives sales gains by increasing purchase incidence rather than by altering spending patterns among existing buyers.

The evidence on increased conversion rates in Table 5 suggests that GenAI has the potential

to reduce market frictions, both by enabling new services and by improving existing ones, thereby enhancing the shopping experience and increasing the probability of purchase. For example, GenAI can mitigate search costs and information asymmetries by providing relevant and timely assistance through a pre-sale chatbot (a 21.7% increase in conversion rates, $p < 0.01$) and by generating more comprehensive and structured product descriptions (a 1.3% increase, $p < 0.01$). It can also reduce search frictions and improve match quality by enhancing the translation and semantic understanding of consumer queries (a 1.2% increase, $p < 0.05$). In addition, GenAI enables personalization of marketing content at scale by generating customized messages across a broad product portfolio (a 3% increase, $p < 0.05$). Finally, evidence from Chargeback Defense and Live Chat Translation reported in Table 4 points to improvements consistent with enhanced experiences for sellers and consumers, respectively, including higher dispute success rates and greater customer satisfaction.

In Appendix C, we report additional evidence on how GenAI reduces frictions along the customer journey, where such data are available. For example, Table C4 shows that GenAI-refined search queries increase the likelihood of a product click by 0.3% ($p < 0.01$; Column 1), consistent with improved search performance in combination with the previously documented increase in purchase conversion. Columns 2 and 3 further indicate that treated consumers view fewer products prior to clicking or purchasing (by 0.5% and 1.5%, respectively; $p < 0.1$), suggesting a reduction in search intensity. These patterns align with prior literature showing that improvements in search quality increase conversion while reducing search effort (Yang, 2013; Zhou et al., 2025). In addition, we observe a statistically significant 2.0% increase in the click-through rate ($p < 0.01$; Column 4), defined as the ratio of product clicks to product views, while the click-to-order conversion rate (Column 5), defined as the ratio of orders to clicks, remains statistically insignificant. Together, these findings suggest that GenAI-refined queries primarily affect the composition of products retrieved at the search stage, rather than the information displayed on product detail pages, reinforcing the proposed query-refinement mechanism.

Table C5 shows that AI-generated product descriptions increase the number of orders placed by 1.1% ($p < 0.05$; Column 2). Moreover, when we stratify products by the length of their original, human-generated descriptions, we find substantial heterogeneity in effects. Products with no or insufficient textual descriptions (fewer than 50 words) experience a 6.5% increase in sales following augmentation with AI-generated content ($p < 0.05$), whereas products with sufficiently detailed descriptions (more than 50 words) exhibit no significant increase (Columns 1 and 2 of Table C6).[16] This pattern indicates that the descriptions augmented by GenAI are most effective for products with limited baseline information.

In the marketing push workflow, GenAI enables large-scale variation in customer-facing con-

[16]Based on internal research and expert surveys conducted by our partner platform, descriptions containing fewer than 50 words are classified as providing insufficient or minimal textual information. This analysis is restricted to the English-language sub-experiment because description-length data are available only for English-language content.

tent. The treatment group includes approximately 2.7 million unique message variants, compared with only roughly 2,000 variants in the control group, which relies exclusively on human-generated content. Consistent with the expanded scope for personalization, AI-generated marketing messages increase clicks by 3.1% ($p < 0.01$) and orders by 2.8% ($p < 0.10$; columns 1 and 2 of Table C7). Taken together, these results show that GenAI improves intermediate engagement outcomes, such as clicks, orders, and click-through rates, and are consistent with a smoother and more informative shopping experience that may help explain the higher conversion and sales gains documented above.

Overall, the evidence in this section is best interpreted as suggestive evidence on channels, rather than as definitive evidence that separately identifies the mechanisms underlying the treatment effects. Consistent with the conceptual discussion in Section 2.1, the increases in conversion, clicks, and orders, together with the absence of systematic changes in cart value or order intensity, are consistent with GenAI reducing search, information, communication, and personalization frictions along the customer journey. At the same time, the experiments identify the combined effect of adding GenAI to each workflow and do not separately isolate informational channels from persuasion, attention, trust, or novelty. The evidence therefore clarifies where the sales gains arise in consumer behavior—primarily at the purchase-incidence margin—while remaining cautious about why they arise.

3.4 Aggregate Gains Across Workflows

In this section, we aggregate the sales gains across workflows to estimate the AI-driven incremental value per consumer observed in our experiments, explicitly accounting for differences in the frequency of consumer interactions across workflows. Specifically, we focus on the four workflows with positive treatment effects in sales, excluding the Google Advertising Title experiment, which serves as a testable bad case and can be readily improved in future iterations.

Table 6 summarizes the key variables used in our calculations. Column 1 reports the AI-driven incremental value per consumer for each experimental workflow, reflecting the estimated average treatment effects (the absolute sales lift per consumer) reported in Column 1 of Table 4. Because each experiment had a different duration, Column 2 presents the time multiplier used to extrapolate these effects to an annual horizon.[17] For instance, the Pre-sale Service Chatbot experiment spans two months, yielding a time multiplier of six (i.e., 12/2). For each workflow, we obtain the annualized AI-driven incremental value per consumer in Column 3 by multiplying the estimated effect in Column 1 by the corresponding time multiplier in Column 2. This calculation assumes that the treatment effects observed during the experimental period remain constant over time, abstracting from potential amplification (e.g., greater engagement and purchases as AI-generated content

[17] We assume that each experiment's duration reflects the typical interval between treatment opportunities for a representative consumer on the platform. For example, a representative consumer is assumed to interact with the Pre-sale Service Chatbot about once every two months, whereas they could receive a Marketing Push Message daily.

enhances consumer satisfaction and platform loyalty) or attenuation (e.g., consumer dissatisfaction and product returns due to potential mismatches between AI-generated content and actual product characteristics).

Table 6: Aggregate Gains Across Workflows

Business Workflow	(1) Incremental Value Per Consumer ($)	(2) Time Multiplier	(3) Annualized Incremental Value Per Consumer ($)
Pre-sale Service Chatbot	0.274 (upper-bound)	6.0	1.64
	0.218 (lower-bound)	6.0	1.31
Search Query Refinement	0.0648	40.6	2.63
Product Description	0.0104	52.1	0.54
Marketing Push Message	0.001005 (upper-bound)	365.0	0.37
	0.000402 (lower-bound)	365.0	0.15
Total (linear additivity)			**5.18** (upper-bound)
			4.63 (lower-bound)

[1] Column (1) reports the absolute lifts in sales (treatment effects of GenAI) for each workflow from Table 4. Column (2) shows the factor used to annualize the workflow-specific estimates. Column (3) reports the annualized values for each workflow.

Finally, we aggregate the annualized estimates across the four workflows to obtain the total annual incremental value per consumer attributable to GenAI (see "Total" in Column 3). This aggregation assumes that effects across workflows are linearly additive and abstracts from potential cross-workflow interactions, such as cannibalization across touchpoints or expansion through synergies. Based on the four GenAI applications with positive sales effects, we estimate an annual incremental value of up to $5.18 per consumer when using the upper-bound estimates from the Marketing Push experiment (implied full-exposure effect) and the Pre-sale Service Chatbot (no service vs. GenAI assistance), and $4.63 per consumer when using the lower-bound estimates from these experiments (Pre-sale Service Chatbot: human agents vs. GenAI assistance equipped with human escalation; Marketing Push: ITT effect with partial exposure). These effects represent roughly 5.5–6.2% of the per-user revenue growth observed in global e-commerce between 2023 and 2024 (Statista, 2024), highlighting the economic significance of these gains relative to broader industry trends.

It is worth noting that these estimates capture only a partial and early-stage view of the firm's efforts to scale up GenAI across workflows. While in 2023 the platform applied GenAI to only a handful of workflows, by 2024 it was deployed in more than 40 applications and by 2025 in over 60. This rapid expansion is also reflected in the growth of API calls to the platform's GenAI tools: in mid-2024, AI-related API requests averaged over 50 million per day, rising to more than 1 billion per day by mid-2025, representing a twentyfold increase. Hence, our point estimates should be interpreted with caution and in light of these rapid adoption trends, which nonetheless indicate that the firm anticipates substantial value from GenAI.

Taken together, these results reveal sizable gains in targeted workflows and measurable contributions to overall platform sales, with substantial potential as GenAI applications diffuse across use cases and models are further refined for domain-specific tasks. This pattern aligns with the view that the aggregate productivity effects of new general-purpose technologies materialize gradually as complementary investments and organizational adaptations accumulate (e.g., Acemoglu, 2025).

4 Heterogeneous Treatment Effects

A central question in online retail is which platform participants benefit most from GenAI-driven improvements. To address this, we examine heterogeneity in treatment effects across buyers, sellers, and products. If GenAI primarily reduces frictions on both the demand and supply sides, one would expect relatively larger gains among participants with lower baseline capabilities—namely, less experienced buyers with limited platform engagement, smaller and newer sellers, and products in the long tail of the sales distribution or in less concentrated categories.

For each dimension, we classify observations into "high" and "low" groups based on pre-experiment characteristics that capture experience, scale, or market position. For buyers, we use measures of prior online shopping experience and purchasing intensity. For sellers, we classify firms based on size and tenure on the platform. For products, we rely on indicators of category concentration, sales volume, and relative price. These groupings allow us to assess whether GenAI adoption disproportionately benefits less sophisticated buyers, smaller or newer sellers, and products in the long tail, thereby shedding light on how GenAI reshapes outcomes across segments of the marketplace.

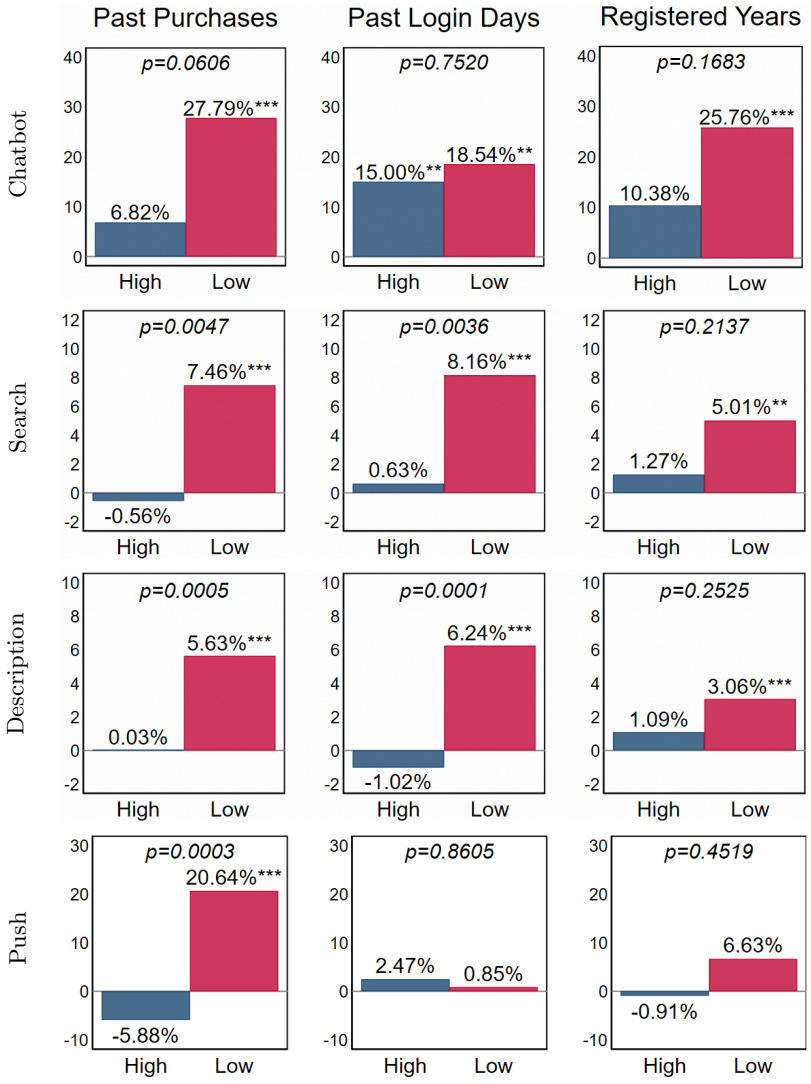
We find consistent evidence of heterogeneity on the demand side: buyers with lower prior spending and less platform experience benefit more from GenAI-powered workflows. On the supply side, the estimated effects point to larger gains for smaller and newer sellers, although these estimates are less precise. Finally, heterogeneity across product groups is more context-dependent. We note that, due to data limitations and workflow-specific features, for some workflows we could not conduct all heterogeneity analyses.

4.1 Heterogeneous Effects Across Consumers

We classify consumers into high- and low-experience groups based on three pre-experiment indicators of online shopping experience: total expenditure over the 30-day period preceding the experiment (Past Purchases), number of login days in the 30 days prior to the experiment (Past Login Days), and years since registration on the platform (Registered Years). For each indicator, consumers in the low group are defined as relatively inexperienced if they fall below the median of the corresponding distribution. We estimate treatment effects separately for high- and low-

experience consumers and test for differences in percentage treatment effects across groups using Wald tests implemented via seemingly unrelated estimation.

Figure 3: Heterogeneous Treatment Effects on Sales Across Consumers



Notes: The bars indicate the % change in treatment effects for the high (blue) and low (red) groups. The p-values are from seemingly unrelated estimations testing the equality of % treatment effects across the two groups.

Results on sales are summarized in Figure 3, with regression results and additional evidence on conversion rates and cart values reported in Appendix D.[18] Overall, we find statistically significant heterogeneity in GenAI's effects across consumer segments. We begin by analyzing heterogeneity by past purchases, our preferred proxy for consumers' prior experience on the platform. Across workflows, GenAI-driven gains are significant and systematically larger for consumers with fewer past purchases, while the effects for consumers with more purchase experience are smaller and

[18]The Google Advertising Title experiment is not suitable for consumer-level heterogeneity analysis, as it was conducted off-platform at the product level, preventing access to detailed consumer characteristics.

insignificant. Further analysis confirms that the differences between these two consumer groups are statistically significant. We find qualitatively similar patterns when stratifying consumers by past login activity and platform tenure, although the differences along these dimensions are generally less pronounced.

Our results indicate that the technology disproportionately benefits less experienced and less sophisticated users, aligning with prior evidence that GenAI can disproportionately benefit vulnerable groups, such as low-skilled workers (Brynjolfsson et al., 2025; Noy and Zhang, 2023; Cui et al., 2026), and extend this insight to the consumer domain. The results are also consistent with earlier work demonstrating that improvements in e-commerce technologies, such as enhanced information provision, search refinement, and personalization, generate greater benefits for consumers who are newer to the platform or have lower purchasing power (Sun et al., 2024).

A plausible explanation for the observed buyer heterogeneity is that less experienced consumers face greater informational and search frictions throughout the purchase journey. With limited domain knowledge and fewer alternative information sources, they rely more heavily on customer service and product descriptions to acquire relevant information. They also exhibit weaker search skills and less familiarity with online search environments, making GenAI-enhanced search particularly valuable for articulating their needs and identifying relevant products. In addition, GenAI-driven message personalization increases the perceived relevance of marketing content, which is especially effective in supporting the purchase decisions of less experienced consumers.

4.2 Heterogeneous Effects Across Sellers

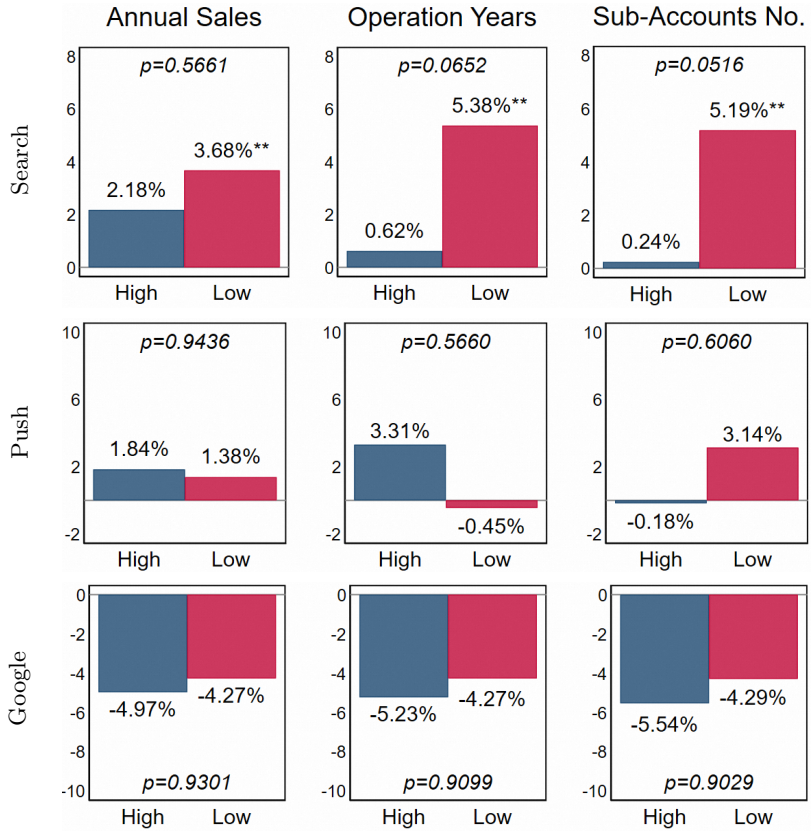
We examine heterogeneity in GenAI's effects across sellers by focusing on differences in seller size and sophistication. Sellers are classified into high and low groups based on three pre-experiment characteristics: annual sales volume (Annual Sales), years of operation on the platform (Operation Years), and the number of sub-accounts associated with the seller's store (# of Sub-Accounts). Sellers in the low group correspond to smaller and less established firms, defined by lower sales, shorter platform tenure, or fewer sub-accounts.[19] We then decompose consumer purchases into transactions involving high- and low-group sellers and estimate treatment effects separately for each group. Differences in percentage treatment effects are formally tested using Wald tests implemented via seemingly unrelated estimation.

Summary results are reported in Figure 4, with regression results and additional details on conversion rates and cart values provided in Appendix E. Overall, we find limited statistical evi-

[19]Sellers are classified into the low group if they (i) account for the bottom 50% of cumulative annual sales, (ii) have operated on the platform for fewer than five years (the platform treats five years as a key performance benchmark), or (iii) maintain fewer than three sub-accounts (following platform guidance, small stores are typically individually operated or mom-and-pop businesses, whereas stores with more sub-accounts generally employ additional staff). Platform-operated sellers are classified as large.

dence of differential effects across seller types. While point estimates often suggest larger gains for smaller and less established sellers, most differences between high- and low-group sellers are statistically insignificant, reflecting imprecise estimates and limited power. This pattern holds across the workflows for which seller-level heterogeneity can be assessed.[20]

Figure 4: Heterogeneous Treatment Effects on Sales Across Sellers



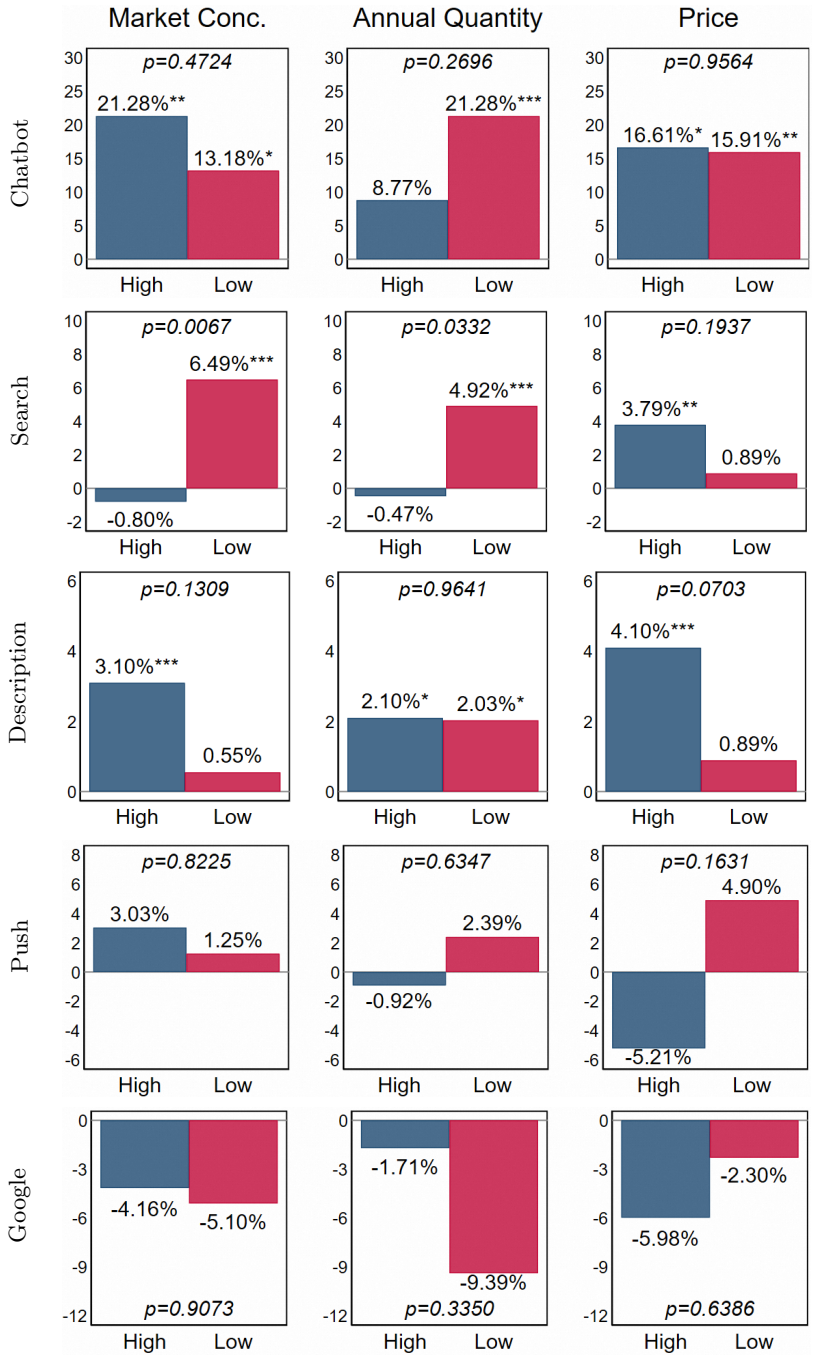
Notes: The bars indicate the % change in treatment effects for the high (blue) and low (red) groups. The p-values are from seemingly unrelated estimations testing the equality of % treatment effects across the two groups.

4.3 Heterogeneous Effects Across Products

Finally, we study heterogeneity in GenAI's effects across products by classifying them into high and low groups along three pre-experiment dimensions. First, category market concentration, measured by the sales share of the top 1% of products within each category, captures the degree of product standardization and demand homogeneity. Second, annual sales quantity within category distinguishes head from tail products. Third, product price within category proxies for purchase stakes and potential information asymmetries. For each dimension, products in the low group correspond

[20]We cannot assess seller heterogeneity in the Pre-sale Service Chatbot and Product Description experiments, as both were conducted exclusively on platform self-sold products with only a handful of platform-operated sellers, leaving insufficient variation for meaningful analysis.

Figure 5: Heterogeneous Treatment Effects on Sales Across Products



Notes: The bars indicate the % change in treatment effects for the high (blue) and low (red) groups. The p-values are from seemingly unrelated estimations testing the equality of % treatment effects across the two groups.

to less concentrated categories, lower-sales (tail) products, or lower-priced items.[21] We decompose consumer purchases into transactions involving high- and low-group products and estimate

[21]Products are classified into the low group if they (i) belong to categories with concentration below the platform average, (ii) account for the bottom 50% of cumulative sales quantity within their category, or (iii) are priced below the category median.

treatment effects separately by group, testing for differences in percentage treatment effects using seemingly unrelated estimation.

Summary results are reported in Figure 5, with regression results and additional details provided in Appendix F. Overall, we find no uniform or sharply defined pattern of product-level heterogeneity across workflows. While several point estimates suggest larger GenAI-driven gains for products in less concentrated categories, tail products or high-priced items—depending on the specific application—most differences between high- and low-group products are statistically insignificant. The direction of the effects is often consistent with GenAI reducing search and information frictions in settings with greater product differentiation, limited sales history or higher decision stakes, but these patterns are not robust across all workflows. Overall, the evidence points to limited and context-dependent product-level heterogeneity in the effects of GenAI.

5 Discussion and Conclusions

The rapid advances in GenAI have generated widespread expectations among investors and business leaders, driving unprecedented investment in infrastructure and applications. Yet important questions remain about where GenAI creates business value, how large those gains are in real-world enterprise settings, and whether they extend beyond improvements in narrowly defined worker-level tasks. This paper provides large-scale experimental evidence on these questions in online retail, studying the deployment of GenAI across multiple consumer-facing workflows within a single global platform. The setting allows us to examine how GenAI affects front-office sales and related business outcomes under a common organizational environment.

Our findings yield three main insights. First, GenAI can deliver measurable short-run improvements in sales across several business workflows, holding other observed inputs and prices fixed. Although the magnitude of these gains varies widely, from negligible effects to double-digit increases, the evidence shows that GenAI can generate economically meaningful value when deployed in consumer-facing retail processes. Second, the pattern of results is consistent with GenAI improving the consumer experience by reducing marketplace frictions along the customer journey. Across workflows, we observe higher conversion rates but little evidence of changes in spending intensity, and post-purchase outcomes such as return rates and customer ratings do not deteriorate. By enriching pre-sale communication, refining search queries, generating richer product descriptions, and personalizing marketing messages, GenAI appears to improve matching efficiency and mitigate information asymmetries, although our data do not allow us to uniquely separate these channels from others such as persuasion, trust, or novelty. Third, the benefits of GenAI adoption are heterogeneous across market participants. On the demand side, less experienced consumers derive disproportionately larger gains. On the supply side, point estimates suggest larger benefits for smaller and newer sellers, though these effects are less precisely estimated. Heterogeneity across

products is more context-dependent. Taken together, these patterns suggest that GenAI may be especially valuable for participants facing greater baseline frictions.

Back-of-the-envelope calculations provide a sense of the economic magnitude of these short-run sales impacts. Annualizing workflow-specific gains and assuming linear additivity, the four GenAI applications with positive sales effects generate an estimated annual incremental value of approximately \$4.6–\$5.2 per consumer. These effects correspond to roughly 5.5–6.2% of global per-user e-commerce revenue growth in 2023–2024. These figures should be interpreted cautiously: they assume that gains from different workflows can be summed without accounting for synergies or overlap, reflect the platform's partial adoption of GenAI in selected workflows during a short experimental period, and capture top-line sales gains rather than bottom-line profitability or full return on investment. Nevertheless, they indicate that even a small number of GenAI applications can generate substantial value for a large, mature retailer. By 2025, the partner platform had deployed GenAI in more than 60 workflows, with usage rising twentyfold as API calls to its proprietary GenAI tools increased between 2024 and 2025. Although our data do not allow us to estimate ROI directly—since we do not observe the full capital, engineering, inference, and organizational costs of GenAI deployment—this accelerating expansion is consistent with the platform internally realizing returns that more than offset these costs at scale.

Our study has several limitations that inform the interpretation of the results and point to avenues for future research. First, the adoption horizon in our experiments was short, spanning days to months. Therefore, our analysis reflects only the immediate, short-run effects of GenAI. The impacts of sustained GenAI use may differ as sellers and consumers adapt their behavior over time and as platforms refine model deployment. Relatedly, we lack data to assess longer-term outcomes such as consumer retention and repeat purchasing. For example, while our results show no deterioration in post-purchase outcomes, as measured by return rates and customer ratings, they do not speak to longer-run consumer responses, such as changes in trust, loyalty, or engagement with the platform.

Second, our analysis is limited to seven workflows that were selected by the platform based on managerial assessments of technical feasibility, organizational costs, and expected sales productivity improvements, rather than representing the full spectrum of business processes where the technology could be deployed. Other business processes, including logistics, inventory management, or dynamic pricing, remain unexplored and could yield distinct productivity effects.

Third, our estimates do not account for potential changes in labor and capital inputs. Many of the studied processes, including customer service, content creation, and chargeback defense, are currently staffed or supported by human labor. Over time, GenAI adoption could displace or augment these functions, yielding additional labor efficiency improvements not captured in our current

analysis.

Fourth, we cannot uniquely attribute the observed sales gains to market expansion, because some gains may reflect intertemporal or cross-category substitution. While we lack the data to assess intertemporal substitution directly, we examined cross-category substitution within the platform through additional, non-tabulated analyses of consumers' other purchases outside the focal experimental context. For example, in Search Query Refinement, we separately examine purchases made outside the search context, while in Product Description, we examine purchases of products not included in the description experiment. The estimated effects on these other purchases are statistically insignificant across experiments, suggesting limited cross-category cannibalization during the experimental period.[22]

Other limitations concern external validity and general-equilibrium effects. Our experiments were conducted on a single, albeit large, global retail platform and may not generalize to other marketplaces. Moreover, as GenAI adoption spreads and competing platforms deploy similar technologies, the estimated effects may be different. For instance, our experiments abstract from strategic responses by competitors, such as changes in pricing or advertising strategy, which could either amplify or attenuate realized sales gains. However, if GenAI improves consumer experience and reduces frictions more broadly, there may still be scope for market expansion even in a more competitive environment.

Taken together, our results show both the current scope and the future potential of GenAI adoption in online retail. In the short run, GenAI delivers measurable gains by creating demand-side value within selected workflows, partly through reductions in market frictions and improvements in the consumer shopping experience. These gains are economically meaningful given the scale and maturity of the partner platform. Over the longer run, GenAI's impact may expand as firms move beyond early use cases, capture cost-reduction opportunities, and adapt organizational structures to integrate the technology more effectively. Continued advances in computational speed, accuracy, and domain coverage may further amplify these effects. At the aggregate level, however, widespread adoption raises open questions about equilibrium effects, competitive dynamics, and the persistence of observed gains. Our study provides a first step by offering causal evidence on how GenAI can reshape core retail workflows to improve business outcomes, while pointing to important directions for future research on general-equilibrium effects, cost-side adjustments, and long-term productivity growth.

[22]We cannot report the detailed tables because these analyses use purchase data outside the focal workflows or product categories, which are commercially sensitive.

References

Acemoglu, Daron (2025). "The simple macroeconomics of AI". In: *Economic Policy* 40.121, pp. 13–58.

Bagwell, Kyle (2007). "The economic analysis of advertising". In: *Handbook of industrial organization* 3, pp. 1701–1844.

Belleflamme, Paul and Martin Peitz (2021). *The economics of platforms*. Cambridge University Press.

Blake, Thomas, Chris Nosko, and Steven Tadelis (2015). "Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment". In: *Econometrica* 83.1, pp. 155–174.

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond (2025). "Generative AI at Work". In: *The Quarterly Journal of Economics*.

Cabral, Luis and Ali Hortaçsu (2010). "The Dynamics of Seller Reputation: Evidence from eBay". In: *Journal of Industrial Economics* 58.1, pp. 54–78.

Calvino, Flavio, Jelmer Reijerink, and Lea Samek (2025). "The effects of generative AI on productivity, innovation and entrepreneurship". In: *OECD Artificial Intelligence Papers 39*.

Chen, Jiafeng and Jonathan Roth (2024). "Logs with zeros? Some problems and solutions". In: *The Quarterly Journal of Economics* 139.2, pp. 891–936.

Chen, Y. and S. Yao (2017). "Sequential search with refinement: Model and application with click-stream data". In: *Management Science* 63.12, pp. 4345–4365.

Choi, Jonathan H., Amy Monahan, and Daniel Schwarcz (2023). "Lawyering in the age of artificial intelligence". In: *SSRN* 4626276.

Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz (2026). "The effects of generative ai on high skilled work: Evidence from three field experiments with software developers". In: *Management Science*.

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Hila Lifshitz, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani (2026). "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality". In: *Organization Science* 37.2, pp. 403–423.

Deloitte (2023). "A new rontier in artificial intelligence: Implications of Generative AI for businesses". In: *Deloitte AI Institute*.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan (2018). "Consumer price search and platform design in internet commerce". In: *American Economic Review* 108.7, pp. 1820–1859.

Fradkin, Andrey (2017). "Search, matching, and the role of digital marketplace design in enabling trade: Evidence from Airbnb". In: *Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb (March 21, 2017)*.

Goldfarb, Avi and Catherine Tucker (2011). "Online Display Advertising: Targeting and Obtrusiveness". In: *Marketing Science* 30.3, pp. 389–404.

Hui, Xiang, Oren Reshef, and Luofeng Zhou (2024). "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market". In: *Organization Science* 35.6, pp. 1977–1989.

Jin, Ginger Zhe and Andrew Kato (2006). "Price, Quality, and Reputation: Evidence from an Online Field Experiment". In: *RAND Journal of Economics* 37.4, pp. 983–1005.

Kestin, G., K. Miller, A. Klales, T. Milbourne, and G. Ponti (2025). "AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting". In: *Scientific Reports* 15.1, p. 17458.

Nelson, Phillip (1974). "Advertising as Information". In: *Journal of Political Economy* 82.4, pp. 729–754.

Noy, Shakked and Whitney Zhang (2023). "Experimental evidence on the productivity effects of generative artificial intelligence". In: *Science* 381, pp. 187–92.

Otis, Nicholas G., Rowan Clarke, Solène Delecourt, David Holtz, and Rembrand Koning (2024). "The uneven impact of generative AI on entrepreneurial performance". In: *SSRN* 4671369.

Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer (2023). "The impact of ai on developer productivity: Evidence from github copilot". In: *arXiv preprint arXiv:2302.06590*.

Statista (2024). *eCommerce: Market Data & Analysis*. Statista report. Accessed: 2025-10-08. URL: `https://www.statista.com/study/42335/ecommerce-report/`.

Stigler, George J. (1961). "The Economics of Information". In: *Journal of Political Economy* 69.3, pp. 213–225.

Sun, Tianshu, Zhe Yuan, Chunxiao Li, Kaifu Zhang, and Jun Xu (2024). "The value of personal data in internet commerce: A high-stakes field experiment on data regulation policy". In: *Management Science* 70.4, pp. 2645–2660.

Tadelis, Steven (2016). "Reputation and feedback systems in online platform markets". In: *Annual review of economics* 8.1, pp. 321–340.

Ursu, Raluca M. (2018). "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions". In: *Marketing Science* 37.4, pp. 530–552.

Yang, Huanxing (2013). "Targeted search and the long tail effect". In: *The RAND Journal of Economics* 44.4, pp. 733–756.

Yoganarasimhan, Hema (2020). "Search Personalization Using Machine Learning". In: *Management Science* 66.3, pp. 1045–1070.

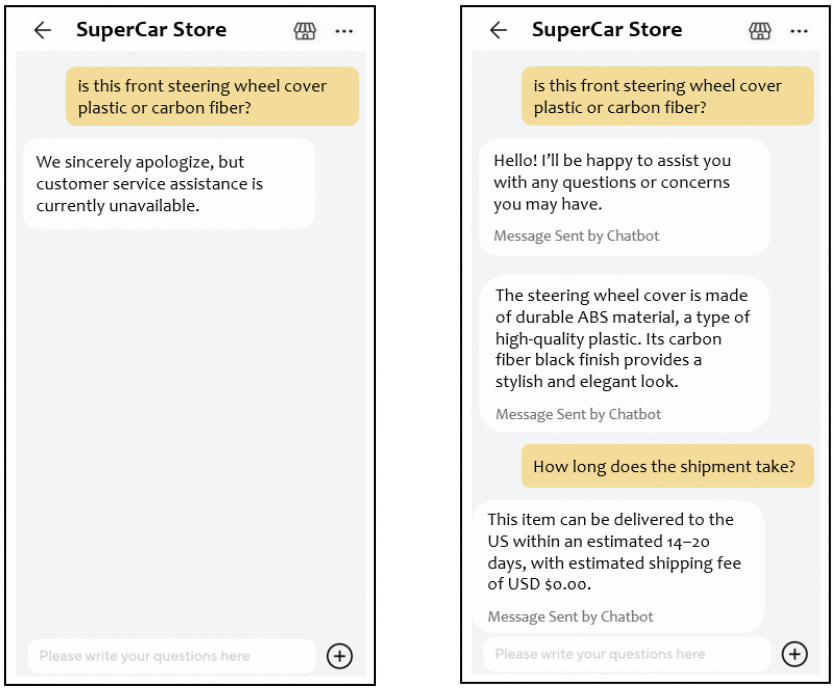
Zhou, Wei, Mingfeng Lin, Mo Xiao, and Lu Fang (2025). "Higher Precision is Not Always Better: Search Algorithm and Consumer Engagement". In: *Management Science* 71.7, pp. 6204–6226.

Appendix

A Illustrative User Interfaces and Examples of Each Experiment

Pre-sale Service Chatbot In Figure A1, Panel (a) depicts the control condition, where a pre-programmed auto response indicates that no service is available, while Panel (b) presents the treatment condition, featuring support from a GenAI-driven chatbot. The chatbot acts as a virtual sales assistant, available 24/7 to provide instant responses to pre-sale inquiries, covering product features, pricing, availability, and delivery options across multiple languages.

Figure A1: Illustration of Pre-sale Service Chatbot

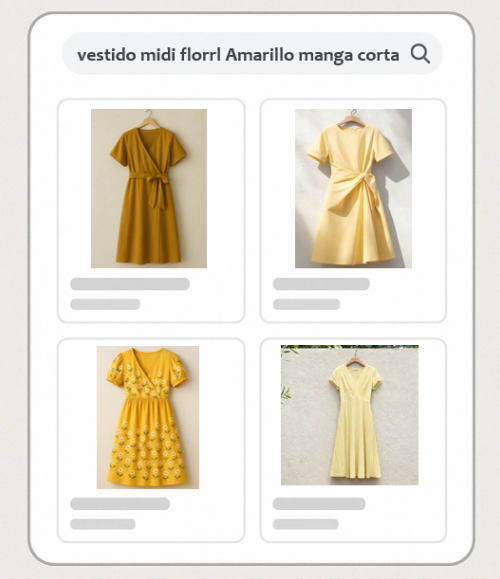


(a) Auto-Response (No Service) (b) GenAI-Driven Chatbot

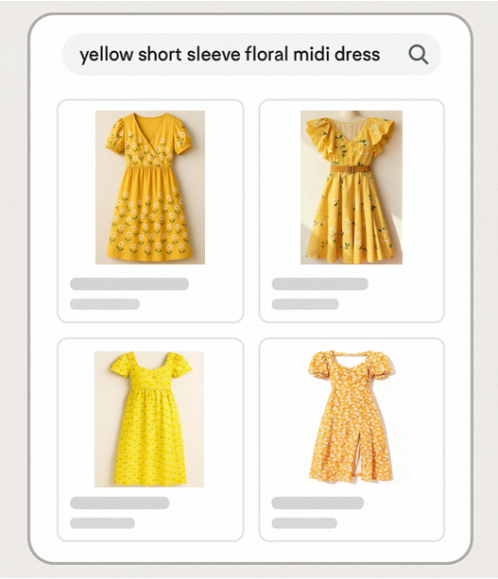
Search Query Refinement In Figure A2, Panel (a) displays the consumer's original search query in Spanish, along with the corresponding results, while Panel (b) presents the structured English query translated and refined by GenAI, as well as the corresponding search results retrieved. The GenAI-powered query refinement can improve the expression of consumer demand and facilitate the matching efficiency of search engines.

Product Description In Figure A3, Panel (a) illustrates the control condition with human-generated descriptions, while Panel (b) shows the treatment condition, where GenAI-created de-

Figure A2: Illustration of Search Query Refinement



(a) Search Results of Consumer Query



(b) Search Results of GenAI-Refined Query

scriptions are layered on top of those written by humans. The AI-generated content provides more comprehensive and structured bullet-point-style descriptions that highlight product features, benefits, and typical use cases.

Figure A3: Illustration of Product Description

Description

Waterproof; 3D Floating Blade System; Intelligent LED Display; Rechargeable & Cordless Design; Long-Lasting Battery Life; Universal Voltage

(a) Human-Created Description

Description

1. Waterproof & Whole Body Washable: Directly use and easily rinse under running water, ensuring effortless cleaning and long-lasting performance.

2. 3D Floating Blade System: Adapts to facial contours for a close, smooth, and accurate shave, reaching every hard-to-reach area.

3. Intelligent LED Display: Provides real-time updates on power status, charging reminders, and usage information, adding a touch of modern convenience to your grooming routine.

4. Built-In USB Rechargeable & Cordless Design: Ensures convenient charging and flexible use, whether at home, in the office or on the go.

5. Long-Lasting Battery Life: Ensures uninterrupted shaving for a more comfortable and efficient experience.

6. Universal Voltage: Compatible with 100-240V, suitable for worldwide use, making it a great choice for international travel.

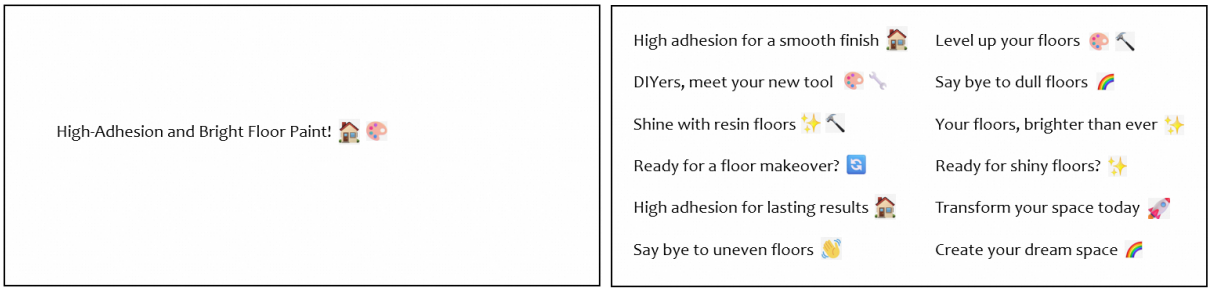
Waterproof; 3D Floating Blade System; Intelligent LED Display; Rechargeable & Cordless Design; Long-Lasting Battery Life; Universal Voltage

(b) GenAI-Created Description Above Original Human Input

Marketing Push Message In Figure A4, Panel (a) illustrates a human-generated marketing push message, whereas Panel (b) displays multiple messages produced by GenAI for the same product. Generative AI enables large-scale creation of diverse marketing content, increasing the likelihood that consumers encounter differentiated messages and thereby allowing platforms to

leverage the benefits of personalized marketing.

Figure A4: Illustration of Marketing Push Message

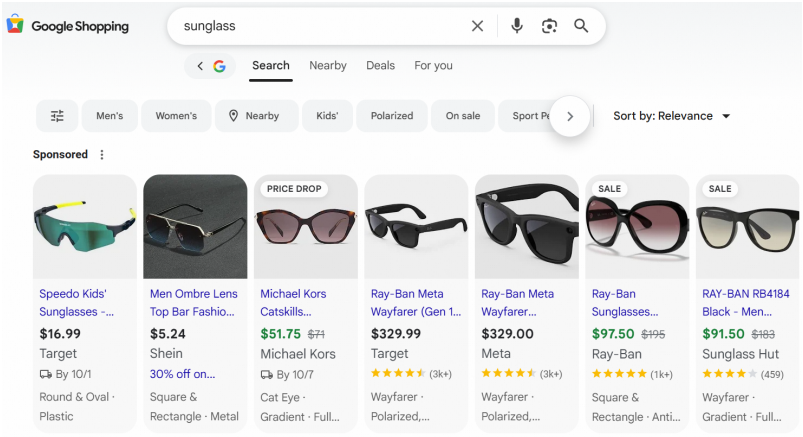


(a) Human-Created Push Message

(b) GenAI-Created Push Messages

Google Advertising Title In this case, GenAI is applied to optimize human-generated product titles for Google advertising. Since the model was not fine-tuned with e-commerce domain knowledge, the performance of GenAI-optimized titles is lower than that of human-generated titles. For example, the original human-generated title for a pair of sunglasses is: "2024 New Arrival Polarized Pitboss 2 Sunglasses Men Cycling Eyewear Goggles Bicycle Glasses". The GenAI-optimized version is: "Men's Polarized Pitboss 2 Sunglasses - Polycarbonate Frame for Cycling, Sports, Bike Goggles Bicycle". On Google Shopping, the first few words of a product title are the most prominent, as shown in Figure A5. Thus, by removing popular keywords such as "New Arrival," GenAI may reduce consumer attention.

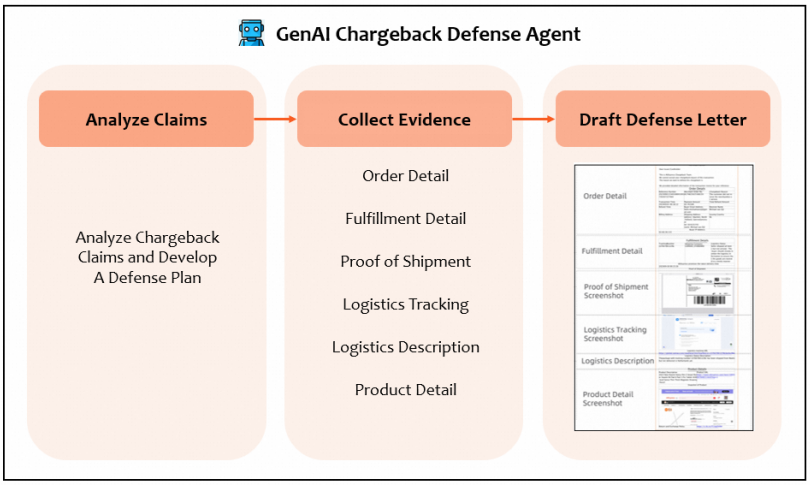
Figure A5: Illustration of Google Shopping Interface



Chargeback Defense Figure A6 illustrates the process by which the chargeback defense agent interprets and analyzes consumer claims, gathers relevant evidence—including transaction records, product details, and shipping information—and drafts persuasive defense letters. By automating

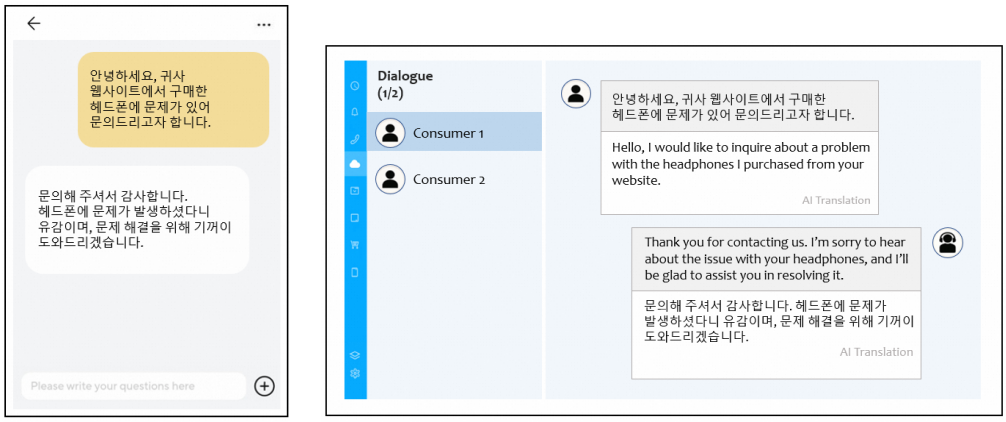
this complex workflow, GenAI enables sellers to respond to chargeback claims more quickly and consistently, thereby improving win rates while reducing compliance burdens and financial losses.

Figure A6: Illustration of Chargeback Defense



Live Chat Translation Figure A7 depicts the live chat translation system. Panel (a) shows the consumer interface, where a Korean consumer submits a query. Panel (b) presents the service agent interface, where a Filipino agent receives the query translated in real time from Korean to English by GenAI. The system supports bidirectional translation: the agent's English response is simultaneously translated into Korean, allowing the consumer to receive the reply in their native language. This functionality enables English-speaking Filipino agents to communicate seamlessly with consumers across multiple languages on the platform.

Figure A7: Illustration of Live Chat Translation



(a) Consumer Interface (b) Customer Service Agent Interface

B Covariate Balance Checks for the Field Experiments

In this section, we present detailed covariate balance checks for each of the five experiments for which granular data are available.

Table B1: Covariate Balance and Randomization Checks

	Control	Treatment	*p-value* (C=T)
Pre-sale Service Chatbot			
Gender	1.000 (0.600)	1.004 (0.598)	0.503
Age Tier	1.000 (0.280)	1.001 (0.277)	0.816
Registered Years	1.000 (0.988)	0.997 (0.957)	0.779
Past Login Days	1.000 (1.071)	1.014 (1.131)	0.220
Past Orders	1.000 (3.108)	0.981 (2.863)	0.528
Past Sales	1.000 (6.441)	1.017 (9.097)	0.833
N. of Consumers	15,457	29,157	
Search Query Refinement			
Gender	1.000 (0.799)	1.001 (0.799)	0.204
Age Tier	1.000 (0.289)	1.000 (0.289)	0.466
Registered Years	1.000 (0.860)	1.001 (0.861)	0.333
Past Login Days	1.000 (1.051)	1.002 (1.054)	0.207
Past Orders	1.000 (3.214)	1.007 (3.468)	0.140
Past Sales	1.000 (6.850)	1.014 (8.206)	0.217
N. of Consumers	929,188	920,194	
Product Description			
Gender	1.000 (0.912)	1.000 (0.912)	0.778
Age Tier	1.000 (0.283)	1.000 (0.283)	0.127
Registered Years	1.000 (0.942)	1.001 (0.942)	0.192
Past Login Days	1.000 (0.942)	0.999 (0.944)	0.490
Past Orders	1.000 (2.437)	1.001 (2.516)	0.627
Past Sales	1.000 (3.613)	0.998 (3.613)	0.515
N. of Consumers	2,392,803	2,380,134	
Marketing Push Message			
Gender	1.000 (0.894)	1.000 (0.894)	0.599
Age Tier	1.000 (0.275)	1.000 (0.277)	0.538
Registered Years	1.000 (1.009)	1.000 (1.008)	0.714
Past Login Days	1.000 (2.386)	0.999 (2.371)	0.501
Past Orders	1.000 (3.304)	1.003 (3.062)	0.157
Past Sales	1.000 (6.115)	1.004 (5.278)	0.157
N. of Consumers	6,869,558	6,845,970	
Google Advertising Title			
Past Sales	1.000 (2.261)	0.993 (2.238)	0.084
Industry ID	1.000 (0.414)	1.000 (0.416)	0.712
N. of Products	621,133	622,883	

[1] Mean (Std. Dev.) are shown with all values normalized by the corresponding variable's control group mean. For the definitions of each variable, refer to the notes in Figure 1.

[2] The first four experiments are conducted at the consumer level and thus the unit of observation is the consumer. The Google Advertising Title experiment is conducted at the product level and thus the unit of observation is the product.

C Main Results: Model, Estimation, and Additional Outcomes

Pre-sale Service Chatbot Pre-sale inquiries regarding product and seller information (e.g., product attributes, promotions, and logistics) play a critical role in shaping consumer purchase decisions. To reduce search costs, support decision-making, improve customer service, and enhance the overall consumer experience, the platform introduced a GenAI-powered chatbot capable of delivering accurate, content-rich responses tailored to a diverse consumer base and available around the clock.

We conduct our analysis using the following regression model:

$$y_i = \beta \times Treat_i + \alpha_{c(i)} + \epsilon_i, \tag{2}$$

where i indicates the consumer, y_i stands for a consumer's outcome (e.g., conversion rate or sales), $Treat_i$ is an indicator for whether the consumer is assigned to the treatment group. Since consumers entered the experiments on different days, we control for their entry-day cohort fixed effects using $\alpha_{c(i)}$.

In addition to the main experiment comparing an auto-response indicating service unavailability ("No Service") with a GenAI chatbot service ("GenAI Reply"), we also studied three supplementary experiments: (1) "No Service" versus "GenAI+Human Reply", where consumers initially interacted with a GenAI chatbot and unresolved issues were escalated to human agents; (2) "Human Reply", where consumers were exclusively served by human agents, versus "GenAI Reply"; (3) "Human Reply" versus "GenAI+Human Reply".

Table C1 presents the impact of the GenAI chatbot on sales, while Table C2 reports its effects on conversion rates and cart value. In both tables, Columns (1)–(2) report effects when the treatment is GenAI Reply, and Columns (3)–(4) report effects for GenAI+Human Reply. Within each set, Columns (1) and (3) use No Service as the control, while Columns (2) and (4) use Human Reply as the control.

Focusing on the comparison between the No Service control and the GenAI Reply treatment (Column 1), we continue to find sizable sales productivity improvements: sales increase by 16.3% and conversion rate rises by 21.7% (both significant at the 1% level). Using Human Reply as the control (Column 2), the GenAI Reply treatment shows no statistically significant differences in either sales or conversion, suggesting that the GenAI chatbot matches the quality of human service but does not outperform it. When the No Service control is compared with the GenAI+Human Reply treatment (Column 3), the gains are even larger—sales improve by 25.0% and conversion by 29.0%—indicating complementarities between GenAI and human agents. By contrast, relative to the Human Reply control (Column 4), the GenAI+Human Reply treatment yields a marginally

significant 11.5% increase in sales, with no statistically significant change in conversion (4.8%), implying that the hybrid approach can enhance revenue even when benchmarked against human agents. Finally, Table C2 shows that all estimated effects on cart value are statistically insignificant.

Table C1: Effects of Pre-sale Service Chatbot on Sales Across Experiments

Treatment:	GenAI Reply	GenAI Reply	GenAI+Human Reply	GenAI+Human Reply
Control:	No Service	Human Reply	No Service	Human Reply
	(1)	(2)	(3)	(4)
Treat	0.274***	0.0701	0.422***	0.218*
	(0.0970)	(0.0967)	(0.115)	(0.114)
%Change	16.25%	3.71%	25.03%	11.53%
Observations	44,614	44,736	30,345	30,467
R-squared	0.000	0.000	0.000	0.000

[1] "Sales" represents the total expenditure on product orders.
[2] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Table C2: Conversion Rate and Cart Value in Pre-sale Service Chatbot Experiments

Treatment:	GenAI Reply	GenAI Reply	GenAI+Human Reply	GenAI+Human Reply
Control:	No Service	Human Reply	No Service	Human Reply
	(1)	(2)	(3)	(4)
Extensive margin: Conversion Rate				
Treat	0.0131***	-0.000768	0.0175***	0.00358
	(0.00252)	(0.00259)	(0.00291)	(0.00304)
%Change	21.70%	-1.03%	28.99%	4.82%
Observations	44,614	44,736	30,345	30,467
R-squared	0.001	0.000	0.001	0.000
Intensive margin: Cart Value (conditional on purchase)				
Treat	-1.264	1.220	-0.859	1.624
	(1.036)	(0.929)	(1.203)	(1.078)
%Change	-4.52%	4.79%	-3.08%	6.38%
Observations	3,076	3,300	2,092	2,316
R-squared	0.000	0.001	0.000	0.001

[1] The first block reports the *Conversion Rate* (binary indicator equal to 1 if consumer makes at least one order during the experimental period). The second block reports *Cart Value*, expenditure per consumer conditional on making a purchase.
[2] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Search Query Refinement Consumers arrive at e-commerce platforms with diverse needs. The search engine serves as the primary channel to facilitate consumers' discovery of desired products, allowing them to express preferences through search queries. Our focal platform seeks to accurately decode the latent demands behind consumers' multilingual queries, translate the queries, and retrieve products that align with their underlying needs. The effectiveness of this process is crucial

in determining match quality, which in turn impacts consumer purchase decisions and platform revenues. GenAI is well-positioned to improve the search algorithm's capabilities in translating consumer queries based on semantic understanding and refinement.

We conduct our analysis using the following regression model:

$$y_i = \beta \times Treat_i + \alpha_{cl(i)} + \epsilon_i, \tag{3}$$

where i denotes a consumer, y_i stands for a consumer's outcome variables, $Treat_i$ is an indicator of a consumer's treatment status. Consumers are classified into various cohorts based on their language groups and their first day of entry in the experiment. As multiple sub-experiments were conducted across consumers in different languages at varying dates, we include entry-day-by-language cohort fixed effects, $\alpha_{cl(i)}$.

The results are summarized in Table C3. Column 1 indicates no significant differences in product views between the two groups. However, treatment group consumers generated 1.1% more clicks (Column 2), spent 2.9% more (Column 4), and were 1.2% more likely to make a purchase (Column 5). We find no significant impact on cart value and order intensity (Columns 6 and 7).

We also explore further mechanism analysis of the Search Query Refinement application. In Table C4, we find that the likelihood of a product click increases by 0.3% ($p < 0.01$; Column 1), consistent with improved search performance in combination with the documented purchase conversion increase in our main effect. Columns 2 and 3 further indicate that treated consumers view fewer products prior to clicking or purchasing, suggesting a reduction in search intensity. In addition, we observe a statistically significant 2.0% increase in the click-through rate ($p < 0.01$; Column 4), defined as the ratio of product clicks to product views, suggesting that consumers found the exposed products more appealing and chose to seek additional details after viewing the summarized search results. The click-to-order conversion rate (Column 5), defined as the ratio of orders to clicks, remains statistically insignificant. This pattern echoes the fact that query refinement influenced only the composition of products retrieved immediately after a query search, not the information displayed on product detail pages. Overall, GenAI-facilitated query refinement enhanced the search algorithm's ability to satisfy consumers' demand, resulting in more effective matching and improved consumer purchases.

Product Description Well-crafted product descriptions are essential for informing consumers about product features, benefits, and uses, thereby reducing information asymmetry, facilitating consumer decision-making and driving platform sales. Despite its importance, our studied platform shows that nearly half of the self-sold products either lack a textual description or contain only a minimal description. GenAI's strengths in content recognition, comprehension, and generation can offer an effective solution to create comprehensive and structured product descriptions for a global

Table C3: Main Effect of Search Query Refinement

	(1) Views	(2) Clicks	(3) Orders	(4) Sales	(5) Conversion Rate	(6) Cart Value	(7) Order Intensity
Treat	-0.549	0.0901***	0.00154	0.0648**	0.00101**	0.370	-0.00315
	(0.887)	(0.0245)	(0.00107)	(0.0314)	(0.000411)	(0.334)	(0.00842)
%Change	-0.18%	1.10%	0.94%	2.93%	1.15%	1.47%	-0.17%
Observations	1,849,382	1,849,382	1,849,382	1,849,382	1,849,382	163,381	163,381
R-squared	0.038	0.041	0.019	0.004	0.030	0.010	0.004

[1] "Views" stands for the number of product views in the summarized search results pages. "Clicks" stands for the number of product clicks into product detail pages. "Orders" is the number of product orders. "Sales" represents the total expenditure on product orders. "Conversion Rate" measures consumers' likelihood of making purchases. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experimental period, and 0 otherwise. "Cart Value" refers to the expenditure per consumer, conditional on making a purchase. "Order Intensity" refers to the number of orders per consumer, conditional on making a purchase.

[2] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Table C4: More Exploration of Search Query Refinement

	(1) Is Click	(2) Views Conditional on Click	(3) Views Conditional Order	(4) Click-through Rate	(5) Click-to-Order Conversion Rate
Treat	0.00267***	-1.728*	-10.34*	0.000767***	-0.000088
	(0.00057)	(1.034)	(5.362)	(0.000079)	(0.000182)
%Change	0.33%	-0.46%	-1.48%	2.02%	-0.35%
Observations	1,849,382	1,508,873	163,381	1,849,382	1,508,873
R-squared	0.013	0.039	0.069	0.003	0.014

[1] "Is Click" measures consumers' likelihood of making a click. It is a binary indicator for click, which equals 1 if a consumer makes at least one click during the experimental period, and 0 otherwise. "Views Conditional on Click" refers to the product views per consumer, conditional on making a click. "Views Conditional on Order" refers to the product views per consumer, conditional on making a purchase. "Click-through Rate" is the ratio of the number of product clicks to the number of product views, measuring the degree of conversion from views to clicks. "Click-to-Order Conversion Rate" stands for the ratio of the number of product orders to the number of product clicks, measuring the degree of conversion from clicks to purchases.

[2] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

audience.

We conduct our analysis using the following regression model:

$$y_i = \beta \times Treat_i + \alpha_{cl(i)} + \epsilon_i, \tag{4}$$

where i denotes a consumer, y_i stands for a consumer's outcome variables, $Treat_i$ is an indicator for whether the consumer is assigned to the treatment group. Because multiple sub-experiments were implemented across language groups on different start dates, we include entry-day-by-language cohort fixed effects, $\alpha_{cl(i)}$.

Column 1 of Table C5 shows no statistically significant differences in product clicks between the control and treatment groups. However, conditional on consumers clicking through to product detail pages—where the descriptions are displayed—treated consumers place 1.1% more orders and spend 2.1% more on those orders (Columns 2 and 3). This improvement is also reflected in a 1.3% increase in the conversion rate (Column 4), indicating a higher likelihood of purchase following exposure to AI-generated descriptions. By contrast, we detect no statistically significant changes in cart value or order intensity. Taken together, these results suggest that augmenting human-generated product descriptions with AI-generated content primarily affects purchase incidence rather than spending intensity, leading to higher overall sales.

Moreover, we stratify products based on the length of their original, human-generated descriptions, distinguishing between those with no or insufficient text (fewer than 50 words) and those with sufficiently detailed descriptions (more than 50 words). Products in the former group experience a 6.5% increase in sales following augmentation with AI-generated content ($p < 0.05$), whereas products in the latter group show no significant effect (Columns 1 and 2 of Table C6), indicating that richer descriptions facilitate consumer decision-making.[23]

Marketing Push Message The platform leverages marketing push notifications, direct messages sent to consumers via their platform app, to draw traffic and boost transactions. Because creating a large volume of diverse and targeted marketing messages manually was challenging, the number of marketing messages is far smaller than the hundreds of millions of consumers, often resulting in many consumers receiving identical content. With the introduction of GenAI, the platform can produce millions of distinct messages, enabling highly personalized marketing strategies through push notifications.

[23]Based on internal research and expert surveys conducted by our partner platform, descriptions containing fewer than 50 words are classified as providing insufficient textual information. This analysis is restricted to the English-language sub-experiment, as description-length data are available only for English-language content.

Table C5: Main Effect of Product Description

	(1) Clicks	(2) Orders	(3) Sales	(4) Conversion Rate	(5) Cart Value	(6) Order Intensity
Treat	0.00233	0.000600**	0.0104**	0.000554***	0.0944	-0.00186
	(0.00183)	(0.000278)	(0.00417)	(0.000187)	(0.0805)	(0.00320)
%Change	0.12%	1.08%	2.05%	1.27%	0.82%	-0.15%
Observations	4,772,937	4,772,937	4,772,937	4,772,937	210,155	210,155
R-squared	0.055	0.008	0.002	0.008	0.010	0.021

[1] "Clicks" stands for the number of product clicks into product detail pages. "Orders" is the number of product orders. "Sales" represents the total expenditure on product orders. "Conversion Rate" measures consumers' likelihood of making purchases. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experimental period, and 0 otherwise. "Cart Value" refers to the expenditure per consumer, conditional on making a purchase. "Order Intensity" refers to the number of orders per consumer, conditional on making a purchase.
[2] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Table C6: Product Description: Heterogeneity by Length of Original, Human-generated Description

	(1) Sales ≤50 words	(2) Sales >50 words	(3) Conversion Rate ≤50 words	(4) Conversion Rate >50 words	(5) Cart Value ≤50 words	(6) Cart Value >50 words
Treat	0.0115**	0.0001	0.00047**	0.00044*	0.367	-0.188
	(0.0050)	(0.0064)	(0.00022)	(0.00026)	(0.235)	(0.198)
%Change	6.50%	0.03%	2.43%	1.60%	3.99%	-1.50%
P-value	0.0536		0.5490		0.0653	
Observations	1,628,409	1,628,409	1,628,409	1,628,409	31,759	45,500
R-squared	0.000	0.001	0.003	0.005	0.001	0.004

[1] "Sales" represents the total expenditure on product orders. "Conversion Rate" measures consumers' likelihood of making purchases. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experimental period, and 0 otherwise. "Cart Value" refers to the expenditure per consumer, conditional on making a purchase.
[2] "≤50 words" refers to descriptions containing fewer than 50 words, while ">50 words" refers to descriptions containing more than 50 words.
[3] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. The p-values are from seemingly unrelated estimations testing the equality of % treatment effects across the two groups. *** p<0.01, ** p<0.05, * p<0.1.

We conduct our analysis using the following regression model:

$$y_i = \alpha + \beta \times Treat_i + \epsilon_i \tag{5}$$

where i denotes a consumer, y_i stands for a consumer's outcome variables, $Treat_i$ is an indicator for whether the consumer is assigned to the treatment group.

Table C7 shows that the use of AI-generated marketing messages leads to increases in consumer engagement. Specifically, clicks increase by 3.1%, orders by 2.8%, and the point estimate for total

purchase amount increases by 1.6%, though it is not statistically significant (Columns 1–3). In addition, the probability that a consumer makes a purchase rises by 3%, while cart value and order intensity remain statistically unchanged (Columns 4–6). Taken together, these results suggest that GenAI enhances the effectiveness of marketing content primarily by expanding consumer participation, consistent with its ability to unlock personalization at scale in settings where human-generated content is constrained by limited resources.

Table C7: Main Effect of Marketing Push

	(1) Clicks	(2) Orders	(3) Sales	(4) Conversion Rate	(5) Cart Value	(6) Order Intensity
Treat	0.000529***	0.000052*	0.000402	0.000048**	-0.206	-0.00138
	(0.000071)	(0.000028)	(0.000816)	(0.0000218)	(0.454)	(0.00854)
%Change	3.08%	2.83%	1.60%	2.95%	-1.32%	-0.12%
Observations	13,715,528	13,715,528	13,715,528	13,715,528	22,425	22,425
R-squared	0.000	0.000	0.000	0.000	0.000	0.000

[1] "Clicks" stands for the number of clicks on the marketing messages. "Orders" is the number of product orders. "Sales" represents the total expenditure on product orders. "Conversion Rate" measures consumers' likelihood of making purchases. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experimental period, and 0 otherwise. "Cart Value" refers to the expenditure per consumer, conditional on making a purchase. "Order Intensity" refers to the number of orders per consumer, conditional on making a purchase.

[2] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Google Advertising Title The platform buys advertising slots in the sponsored section of Google Shopping to promote its products on Google and attract traffic to its site. To maximize purchases derived from Google advertisements, a critical operational decision for the platform is how to design product titles to increase both product discoverability and the likelihood of user clicks. The platform leveraged GenAI to refine titles based on the original seller-created titles, optimizing them for better visibility and engagement.

We conduct our analysis employing the following regression model:

$$y_i = \beta \times Treat_i + \alpha_{c(i)} + \epsilon_i, \tag{6}$$

where i denotes a product, y_i stands for the outcome variables for a product, $Treat_i$ is an indicator for whether the product is assigned to the treatment group. We control for product entry-day cohort fixed effect $\alpha_{c(i)}$.

Table C8 presents the main findings. Columns 1 and 2 indicate a 7.7% decrease in ad views and a 10.3% decrease in ad clicks for the treatment group, respectively. Columns 3, 4, and 5 report a non-significant reduction in sales, conversion rate, and cart value, respectively. As we discussed in

Section 3.1, the null effect on sales can be attributed to the lack of fine-tuning using e-commerce domain knowledge when setting up the GenAI model.

Table C8: Main Effect of Google Advertising

	(1) Views	(2) Clicks	(3) Sales	(4) Conversion Rate	(5) Cart Value
Treat	-1.547***	-0.0247***	-0.00602	-0.000090	-0.784
	(0.148)	(0.00304)	(0.00534)	(0.000111)	(0.992)
%Change	-7.68%	-10.26%	-4.56%	-2.29%	-2.32%
Observations	1,244,016	1,244,016	1,244,016	1,244,016	4,811
R-squared	0.000	0.000	0.000	0.000	0.000

[1] "Views" represents the number of times the product advertisement is viewed on Google. "Clicks" refers to the number of times the product advertisement is clicked on Google. "Sales" represents the total expenditure on product orders. "Conversion Rate" is 1 if a product is purchased at least once during the experiment period, and 0 otherwise. "Cart Value" refers to the expenditure per product, conditional on the product being purchased.

[2] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Chargeback Defense Online sellers often struggle to defend against chargebacks due to reasons such as non-receipt of goods. Chargebacks can lead to significant financial losses and jeopardize the long-term sustainability of sellers' businesses on the platform. The whole process of contesting chargeback disputes includes analyzing claims, collecting necessary evidence (e.g., order details, fulfillment records, proof of shipment and logistics tracking through ways like interfacing with diverse APIs), and crafting compelling chargeback defense letters. The rapid advancement of GenAI enabled the platform to develop a chargeback defense agent that offers a one-stop solution to streamline the intricacies of chargeback disputes.

As the data was not available for us to review, we report findings estimated by the platform's internal data science team. Their estimates indicate that the adoption of the GenAI agent helps increase the success rate of chargeback defense by 15%.

Live Chat Translation E-commerce platforms must provide robust consumer services for consumers seeking consultation or negotiation with the platform, such as addressing inquiries about the platform's promotional details and resolving disputes when consensus with sellers is not reached. For our focal platform, delivering native-language consumer services to a diverse, multilingual consumer base incurs significant costs. Thus, a significant portion of non-English consumer inquiries were supported by Filipino consumer service agents in English. A straightforward application of GenAI allows the platform to equip low-cost Filipino agents with robust real-time translation support, aiming to provide more effective communication between consumers and agents.

Similarly, due to the unavailability of raw data, the effect estimate for this process is from the platform's internal data science team and could not be independently verified. During the experiment, consumers were queried whether they were satisfied or not with the service via a survey question immediately after service completion. As a result, there was a 5.2% increase in consumer satisfaction, suggesting that GenAI helped Filipino agents to better serve non-English-speaking consumers.

Post-purchase Outcomes In Table C9, we provide evidence on two important post-purchase measures of consumer satisfaction with the products, namely product return rates and customer ratings. We do not find evidence that GenAI applications deteriorate such outcomes.

Table C9: Average Treatment Effects of GenAI Adoption on Product Returns and Reviews

	Product Return Rate			Positive Review Rate		
	(1)	(2)	(3)	(4)	(5)	(6)
Business Workflow	Coefficient	% Change	Obs	Coefficient	% Change	Obs
Pre-sale Service Chatbot	0.0208	2.08%	3,076	0.0452*	4.52%	1,008
	(0.0582)			(0.0247)		
Search Query Refinement	-0.0141	-1.41%	163,381	0.0031	0.31%	26,209
	(0.0152)			(0.0040)		
Product Description	-0.0386***	-3.86%	210,155	-0.0042	-0.42%	31,916
	(0.0142)			(0.0034)		
Marketing Push Message	-0.114***	-11.40%	22,425	-0.0174	-1.74%	2,853
	(0.0394)			(0.0122)		

[1] "Return Rate" is defined as the share of orders that are returned within a year after the purchase. "Positive Review Rate" is defined as the share of rated orders that receive four- or five-star ratings on a five-point scale. The estimation of Return Rate is conditional on consumers with orders, while the estimation of Positive Review Rate is conditional on consumers with rated orders. Such data are not available for the Google Advertising Title workflow. To protect proprietary business information, both outcome variables are normalized by the control-group mean. Consequently, coefficients can be interpreted as percentage changes relative to the control condition.

[2] Columns (1) and (4) report the estimated coefficients, with standard errors in parentheses. Columns (2) and (5) report % Change, calculated as the treatment effect divided by the control group mean. Columns (3) and (6) report the number of observations. *** p<0.01, ** p<0.05, * p<0.1.

D Details on Consumer Treatment Effect Heterogeneity

Pre-sale Service Chatbot In this experiment, we focus on the comparison between consumers who are served only by the GenAI chatbot and those who are served by an auto-response indicating service unavailability. Table D1 reports the heterogeneous treatment effects across consumer groups. Our analysis reveals that the gains in sales are more pronounced among newer consumers with shorter registration histories (Column 2), less active consumers with fewer login days (Column 4), and consumers with less past spending (Column 6). However, differences across consumer groups are generally not statistically significant, with the exception of sales across past purchases, for which the difference is significant at the 10% level.

Search Query Refinement Table D2 reports the heterogeneous treatment effect of refining search query with GenAI on different consumer groups. For sales, the benefits are significant and larger for inexperienced consumers (Columns 2, 4, and 6), while the effects for more experienced consumers are smaller and insignificant (Columns 1, 3, and 5). The differences between these two consumer groups are statistically significant when consumers are classified by past purchases and past login days. When focusing on conversion rates, we also find statistically significant differences between consumer groups defined by years since registration. Since consumers with limited online experience often struggle to effectively articulate their needs through query-based searches, they tend to benefit more from enhanced match quality achieved by applying GenAI to the semantic translation of consumer queries.

Product Description In Table D3, we observe a similar pattern to Search Query Refinement. There is a significant and more pronounced sales increase for inexperienced consumers (Columns 2, 4, and 6). The differences between high- and low-group consumers are statistically significant for consumers classified by past purchases and past login days. Augmenting product descriptions with AI-generated content substantially enhances the sufficiency and clarity of product information, thereby motivating consumer purchase decisions, particularly among less experienced consumers.

Marketing Push Message In Table D4, we find that the effects of AI-generated marketing push messages are concentrated among consumers with fewer past purchases. For this group, treatment significantly increases both sales and conversion rates (Column 6), while the corresponding estimates for high-purchase consumers are negative and statistically insignificant (Column 5); the high-low differences are statistically significant. For registered years and past login days, conversion-rate gains are larger for lower-experience consumers, but the differences across groups are not statistically significant. These patterns suggest that AI-generated push messages are particularly effective for consumers with less purchase history, likely by making promotional content more relevant or salient.

Table D1: Consumer HTE for Pre-sale Service Chatbot

	(1)	(2)	(3)	(4)	(5)	(6)
	Registered Years		Past Login Days		Past Purchases	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	0.217	0.329***	0.318**	0.230**	0.159	0.343***
	(0.153)	(0.118)	(0.156)	(0.114)	(0.180)	(0.105)
%Change	10.38%	25.76%	15.00%	18.54%	6.82%	27.79%
P-value	0.1683		0.7520		0.0606	
Panel B: Conversion Rate						
Treat	0.0131***	0.0131***	0.0128***	0.0135***	0.0128***	0.0130***
	(0.00377)	(0.00333)	(0.00386)	(0.00321)	(0.00434)	(0.00300)
%Change	18.77%	25.77%	17.11%	29.62%	16.41%	27.09%
P-value	0.4000		0.1212		0.1955	
Panel C: Cart Value						
Treat	-2.130	0.0141	-0.516	-2.331	-2.470	0.157
	(1.432)	(1.471)	(1.387)	(1.525)	(1.568)	(1.320)
% Change	-7.11%	0.06%	-1.82%	-8.56%	-8.27%	0.61%
P-value	0.3257		0.3702		0.2170	
Observation (Panel A&B)	22,461	22,153	22,570	22,044	18,610	26,004
Observation (Panel C)	1,761	1,315	1,877	1,199	1,609	1,467

[1] We classify consumers into high and low groups and capture their online shopping experience based on registration duration, 30-day pre-experiment login days and 30-day pre-experiment purchase. For each indicator, consumers in the low group are defined as relatively inexperienced if they fall below the median of the corresponding distribution.

[2] "Sales" represents the total expenditure on product orders. "Conversion Rate" measures consumers' likelihood of making purchases. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experimental period, and 0 otherwise. "Cart Value" refers to the expenditure per consumer, conditional on making a purchase.

[3] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Table D2: Consumer HTE for Search Query Refinement

	(1)	(2)	(3)	(4)	(5)	(6)
	Registered Years		Past Login Days		Past Purchases	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	0.0360	0.0867**	0.0203	0.104***	-0.0216	0.106***
	(0.0526)	(0.0380)	(0.0566)	(0.0303)	(0.0757)	(0.0291)
%Change	1.27%	5.01%	0.63%	8.16%	-0.56%	7.46%
P-value	0.2137		0.0036		0.0047	
Panel B: Conversion Rate						
Treat	-0.000104	0.00186***	0.000500	0.00141***	0.000651	0.00115***
	(0.000697)	(0.000487)	(0.000675)	(0.000482)	(0.000907)	(0.000421)
%Change	-0.09%	2.79%	0.43%	2.32%	0.44%	1.93%
P-value	0.0023		0.0421		0.1122	
Panel C: Cart Value						
Treat	0.329	0.440	-0.0145	1.126**	-0.355	1.224***
	(0.431)	(0.526)	(0.452)	(0.460)	(0.484)	(0.450)
%Change	1.34%	1.69%	-0.05%	5.36%	-1.35%	5.13%
P-value	0.8965		0.0363		0.0139	
Observation (Panel A&B)	813,317	1,036,065	890,044	959,338	596,721	1,252,661
Observation (Panel C)	93,362	70,019	104,531	58,850	88,135	75,246

[1] Refer to Table D1 for detailed notes.

Table D3: Consumer HTE for Product Description

	(1)	(2)	(3)	(4)	(5)	(6)
	Registered Years		Past Login Days		Past Purchases	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	0.00671	0.0130***	-0.00606	0.0263***	0.000224	0.0206***
	(0.00761)	(0.00455)	(0.00655)	(0.00522)	(0.00718)	(0.00423)
%Change	1.09%	3.06%	-1.02%	6.24%	0.03%	5.63%
P-value	0.2525		0.0001		0.0005	
Panel B: Conversion Rate						
Treat	0.000575*	0.000520**	0.000325	0.000785***	0.000568*	0.000533**
	(0.000312)	(0.000229)	(0.000284)	(0.000245)	(0.000292)	(0.000233)
%Change	1.10%	1.39%	0.65%	2.06%	1.05%	1.59%
P-value	0.7441		0.1194		0.5235	
Panel C: Cart Value						
Treat	0.003	0.184*	-0.180	0.438***	-0.127	0.447***
	(0.126)	(0.098)	(0.112)	(0.115)	(0.115)	(0.098)
%Change	0.03%	1.62%	-1.50%	3.96%	-1.06%	4.09%
P-value	0.2496		0.0001		0.0001	
Observation (Panel A&B)	2,035,278	2,737,659	2,322,437	2,450,500	2,386,336	2,386,601
Observation (Panel C)	107,098	103,057	115,708	94,447	129,462	80,693

[1] Refer to Table D1 for detailed notes.

Table D4: Consumer HTE for Marketing Push Message

	(1)	(2)	(3)	(4)	(5)	(6)
	Registered Years		Past Login Days		Past Purchases	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	-0.000269	0.00128	0.000694	0.000195	-0.00250	0.00238***
	(0.00115)	(0.00112)	(0.00136)	(0.00100)	(0.00161)	(0.000712)
% Change	-0.91%	6.63%	2.47%	0.85%	-5.88%	20.64%
P-value	0.4519		0.8605		0.0003	
Panel B: Conversion Rate						
Treat	0.0000314	0.0000681**	0.0000122	0.0000737***	-0.0000590	0.000111***
	(0.0000301)	(0.0000312)	(0.0000345)	(0.0000281)	(0.0000416)	(0.0000210)
% Change	1.76%	4.98%	0.71%	4.82%	-2.22%	14.17%
P-value	0.4372		0.2810		0.0000	
Panel C: Cart Value						
Treat	-0.432	0.225	0.285	-0.573	-0.600	0.828
	(0.578)	(0.731)	(0.714)	(0.583)	(0.558)	(0.761)
% Change	-2.62%	1.60%	1.75%	-3.80%	-3.76%	5.62%
P-value	0.4956		0.3551		0.1286	
Observations (Panel A&B)	7,959,851	5,755,677	5,796,995	7,918,533	6,085,076	7,630,452
Observations (Panel C)	14,351	8,074	10,026	12,399	16,025	6,400

[1] Refer to Table D1 for detailed notes.

E Details on Seller Treatment Effect Heterogeneity

Search Query Refinement Table E1 reports the heterogeneous impact on high- versus low-type sellers classified based on three metrics: annual past sales, tenure on the platform, and number of sub-accounts. Small sellers with lower transaction volumes (Column 2), shorter operational histories (Column 4), and fewer sub-accounts (Column 6) experience a significant and larger increase in sales from treated consumers. By contrast, sales on larger, tenured, and scaled sellers show no significant change. The differences between these two seller groups are statistically significant when sellers are divided by operation years and the number of sub-accounts. When using conversion rates, we also find statistically significant differences between seller groups defined by annual sales. Thus, the GenAI-powered search query refinement generates greater value for smaller sellers.

Marketing Push Message In Table E2, we find limited evidence of heterogeneity across seller groups. Most high-low differences are statistically insignificant across sales, conversion rates, and cart values. The main exception is conversion rates by the number of sub-accounts: sellers with fewer sub-accounts experience a larger conversion-rate gain, and the high-low difference is marginally significant at the 10% level.

Google Advertising Title Table E3 shows that the differences between high- and low-group sellers are all statistically insignificant.

F Details on Product Treatment Effect Heterogeneity

Pre-sale Service Chatbot Table F1 reports the product-level heterogeneous treatment effects of the Pre-sale Service Chatbot experiment. When focusing on conversion rates, we detect statistically significant differences between product groups classified by annual sales quantity, indicating that the GenAI-driven chatbot service is more effective at converting consumers for tail products.

Search Query Refinement In Table F2, we find that for sales, the gains are larger and statistically significant only for products in low-concentration categories (Column 2), products with lower sales volume (Column 4), and products with high price level (Column 5). The differences between these two product groups are statistically significant when products are classified by market concentration and annual quantity, indicating that GenAI reduces search frictions in settings with greater product differentiation or limited sales history.

Product Description Table F3 shows that, for sales, statistically significant differences across product groups arise only when products are classified by price level. This pattern is consistent with GenAI being more effective at reducing information asymmetries in settings with higher decision stakes.

Table E1: Seller HTE for Search Query Refinement

	(1)	(2)	(3)	(4)	(5)	(6)
	Annual Sales		Operation Years		# of Sub-Accounts	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	0.0241	0.0407**	0.0070	0.0578**	0.00238	0.0624**
	(0.0243)	(0.0180)	(0.0182)	(0.0237)	(0.0174)	(0.0246)
%Change	2.18%	3.68%	0.62%	5.38%	0.24%	5.19%
P-value	0.5661		0.0652		0.0516	
Panel B: Conversion Rate						
Treat	0.000085	0.00105***	0.000444	0.000673**	0.000242	0.00100***
	(0.000291)	(0.000352)	(0.000343)	(0.000306)	(0.000317)	(0.000330)
%Change	0.21%	1.69%	0.76%	1.47%	0.49%	1.85%
P-value	0.0573		0.3493		0.0738	
Panel C: Cart Value						
Treat	0.550	0.238	-0.042	0.773	-0.054	0.613
	(0.554)	(0.267)	(0.287)	(0.486)	(0.325)	(0.426)
%Change	2.06%	1.34%	-0.22%	3.29%	-0.26%	2.75%
P-value	0.7711		0.1484		0.2069	
Observation (Panel A&B)	1,849,382	1,849,382	1,849,382	1,849,382	1,849,382	1,849,382
Observation (Panel C)	76,536	115,804	108,790	85,195	91,774	100,643

[1] We classify sellers into high and low groups based on three different proxies. Low-group sellers are generally small sellers, as defined by meeting the following pre-experiment criteria: (1) accounting for the bottom 50% cumulative share of total sales; (2) having operated on the platform for fewer than five years; or (3) maintaining fewer than three sub-accounts for their online store. Platform-operated sellers are classified as large.

[2] "Sales" represents the total expenditure on product orders. "Conversion Rate" measures consumers' likelihood of making purchases. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experimental period, and 0 otherwise. "Cart Value" refers to the expenditure per consumer, conditional on making a purchase.

[3] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Table E2: Seller HTE for Marketing Push Message

	(1)	(2)	(3)	(4)	(5)	(6)
	Annual Seller Sales		Operation Years		# of Sub-Accounts	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	0.000216	0.000186	0.000453	-0.0000514	-0.0000206	0.000422
	(0.000594)	(0.000559)	(0.000579)	(0.000575)	(0.000517)	(0.000631)
% Change	1.84%	1.38%	3.31%	-0.45%	-0.18%	3.14%
P-value	0.9436		0.5660		0.6060	
Panel B: Conversion Rate						
Treat	0.0000240	0.0000236	0.0000166	0.0000310**	0.00000409	0.0000435***
	(0.0000147)	(0.0000161)	(0.0000170)	(0.0000137)	(0.0000157)	(0.0000152)
% Change	3.28%	2.69%	1.68%	4.97%	0.48%	5.68%
P-value	0.8299		0.2370		0.0557	
Panel C: Cart Value						
Treat	-0.221	-0.195	0.223	-0.951	-0.0912	-0.420
	(0.734)	(0.564)	(0.530)	(0.813)	(0.553)	(0.729)
% Change	-1.38%	-1.27%	1.61%	-5.16%	-0.66%	-2.40%
P-value	0.9852		0.2488		0.7633	
Observation (Panel A&B)	13,715,528	13,715,528	13,715,528	13,715,528	13,715,528	13,715,528
Observation (Panel C)	10,212	12,213	13,649	8,776	11,618	10,807

[1] Refer to Table E1 for detailed notes.

Table E3: Seller HTE for Google Advertising Title

	(1)	(2)	(3)	(4)	(5)	(6)
	Annual Seller Sales		Operation Years		# of Sub-Accounts	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	-0.00605	-0.00610	-0.00656	-0.00578	-0.00759	-0.00562
	(0.00711)	(0.00797)	(0.00935)	(0.00650)	(0.0125)	(0.00589)
% Change	-4.97%	-4.27%	-5.23%	-4.27%	-5.54%	-4.29%
P-value	0.9301		0.9099		0.9029	
Panel B: Conversion Rate						
Treat	-0.000186	0.0000057	-0.0000884	-0.0000906	-0.0000464	-0.000100
	(0.000155)	(0.000159)	(0.000189)	(0.000137)	(0.000232)	(0.000127)
% Change	-4.82%	0.14%	-2.46%	-2.23%	-1.33%	-2.49%
P-value	0.4048		0.9713		0.8736	
Panel C: Cart Value						
Treat	-0.0528	-1.588	-0.990	-0.697	-1.672	-0.602
	(1.372)	(1.425)	(1.896)	(1.162)	(2.520)	(1.071)
% Change	-0.17%	-4.41%	-2.84%	-2.09%	-4.27%	-1.85%
P-value	0.4551		0.9014		0.7378	
Observations (Panel A&B)	622,083	621,933	397,339	846,677	258,157	985,859
Observation (Panel C)	2,343	2,468	1,411	3,400	897	3,914

[1] Refer to Table E1 for detailed notes.

Marketing Push Message Table F4 shows that, in most cases, we do not observe statistically significant differences between the two product groups.

Google Advertising Title Table F5 shows no statistically significant differences between the two product groups.

Table F1: Product HTE for Pre-sale Service Chatbot

	(1)	(2)	(3)	(4)	(5)	(6)
	Market Concentration		Annual Quantity		Price	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	0.136**	0.138*	0.0597	0.214***	0.137*	0.137**
	(0.0532)	(0.0821)	(0.0564)	(0.0799)	(0.0782)	(0.0589)
% Change	21.28%	13.18%	8.77%	21.28%	16.61%	15.91%
P-value	0.4724		0.2696		0.9564	
Panel B: Conversion Rate						
Treat	0.00647***	0.00666***	0.00265	0.0105***	0.00404**	0.00910***
	(0.00166)	(0.00196)	(0.00162)	(0.00198)	(0.00158)	(0.00202)
% Change	26.60%	18.48%	10.32%	30.28%	17.44%	24.46%
P-value	0.3443		0.0182		0.4139	
Panel C: Cart Value						
Treat	-1.108	-1.300	-0.366	-1.990	-0.258	-1.584*
	(1.182)	(1.551)	(1.370)	(1.472)	(2.182)	(0.926)
% Change	-4.22%	-4.47%	-1.38%	-6.86%	-0.72%	-6.84%
P-value	0.9704		0.4498		0.3981	
Observation (Panel A&B)	44,614	44,614	44,614	44,614	44,614	44,614
Observation (Panel C)	1,274	1,802	1,223	1,853	1,151	1,925

[1] We classify products into high and low groups based on three key pre-experiment dimensions. (1) Category market concentration, measured by the sales share of the top 1% of products (ranked by annual sales) within each category. Products in the low group belong to categories with concentration levels below the platform average. (2) Annual sales quantity, also defined within the category. Products in the low group—referred to as tail products—are those comprising the bottom 50% cumulative share of total sales quantity within each category. (3) Product price, defined within each category to control for category-level pricing variation. Low-priced products are those priced below the median of their respective category.

[2] "Sales" represents the total expenditure on product orders. "Conversion Rate" measures consumers' likelihood of making purchases. It is a binary indicator for purchase, which equals 1 if a consumer makes at least one order during the experimental period, and 0 otherwise. "Cart Value" refers to the expenditure per consumer, conditional on making a purchase.

[3] Standard errors are in parentheses. % Change is calculated by dividing the treatment effect by the control group mean. *** p<0.01, ** p<0.05, * p<0.1.

Table F2: Product HTE for Search Query Refinement

	(1)	(2)	(3)	(4)	(5)	(6)
	Market Concentration		Annual Quantity		Price	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	-0.00862	0.0734***	-0.00381	0.0686***	0.0590**	0.00586
	(0.0174)	(0.0253)	(0.0156)	(0.0258)	(0.0296)	(0.00864)
% Change	-0.80%	6.49%	-0.47%	4.92%	3.79%	0.89%
P-value	0.0067		0.0332		0.1937	
Panel B: Conversion Rate						
Treat	0.000247	0.000823***	0.0000005	0.00121***	0.000526*	0.000548
	(0.000332)	(0.000302)	(0.000305)	(0.000343)	(0.000301)	(0.000342)
% Change	0.45%	1.85%	0.00%	2.07%	1.19%	0.94%
P-value	0.0888		0.0064		0.7493	
Panel C: Cart Value						
Treat	-0.292	1.106**	-0.098	0.587	0.717	-0.032
	(0.291)	(0.534)	(0.319)	(0.411)	(0.615)	(0.132)
%Change	-1.49%	4.35%	-0.55%	2.46%	2.04%	-0.28%
P-value	0.0208		0.2052		0.2583	
Observations (Panel A&B)	1,849,382	1,849,382	1,849,382	1,849,382	1,849,382	1,849,382
Observation (Panel C)	101,975	82,912	84,799	109,265	82,272	108,485

[1] Refer to Table F1 for detailed notes.

Table F3: Product HTE for Product Description

	(1)	(2)	(3)	(4)	(5)	(6)
	Market Concentration		Annual Quantity		Price	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	0.00932***	0.00113	0.00497*	0.00547*	0.00760***	0.00285
	(0.00299)	(0.00284)	(0.00281)	(0.00299)	(0.00291)	(0.00286)
% Change	3.10%	0.55%	2.10%	2.03%	4.10%	0.89%
P-value	0.1309		0.9641		0.0703	
Panel B: Conversion Rate						
Treat	0.000384**	0.000141	0.000456***	0.000166	0.000393***	0.000209
	(0.000151)	(0.000120)	(0.000134)	(0.000141)	(0.000100)	(0.000164)
% Change	1.38%	0.81%	2.07%	0.69%	3.32%	0.63%
P-value	0.5014		0.0845		0.0047	
Panel C: Cart Value						
Treat	0.186**	-0.019	0.018	0.138	0.092	0.032
	(0.089)	(0.141)	(0.108)	(0.104)	(0.204)	(0.071)
%Change	1.72%	-0.16%	0.16%	1.24%	0.59%	0.33%
P-value	0.1895		0.4258		0.8605	
Observations (Panel A&B)	4,772,937	4,772,937	4,772,937	4,772,937	4,772,937	4,772,937
Observation (Panel C)	133,779	83,687	106,252	115,810	57,458	159,492

[1] Refer to Table F1 for detailed notes.

Table F4: Product HTE for Marketing Push Message

	(1)	(2)	(3)	(4)	(5)	(6)
	Market Concentration		Annual Quantity		Price	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	0.000146	0.000255	-0.0000554	0.000457	-0.000429	0.000831
	(0.000339)	(0.000742)	(0.000349)	(0.000737)	(0.000511)	(0.000636)
% Change	3.03%	1.25%	-0.92%	2.39%	-5.21%	4.90%
P-value	0.8225		0.6347		0.1631	
Panel B: Conversion Rate						
Treat	0.0000213**	0.0000263	0.0000115	0.0000361*	0.0000130	0.0000346*
	(0.00000921)	(0.0000198)	(0.0000108)	(0.0000190)	(0.00000976)	(0.0000195)
% Change	7.59%	1.98%	2.91%	2.97%	4.06%	2.68%
P-value	0.1190		0.9813		0.6838	
Panel C: Cart Value						
Treat	-0.730	-0.108	-0.566	-0.0896	-2.292*	0.285
	(1.038)	(0.504)	(0.772)	(0.546)	(1.381)	(0.444)
% Change	-4.24%	-0.71%	-3.71%	-0.57%	-8.92%	2.17%
P-value	0.6095		0.6106		0.0825	
Observations (Panel A&B)	13,715,528	13,715,528	13,715,528	13,715,528	13,715,528	13,715,528
Observations (Panel C)	3,993	18,432	5,498	16,927	4,483	17,942

[1] Refer to Table F1 for detailed notes.

Table F5: Product HTE for Google Advertising Title

	(1)	(2)	(3)	(4)	(5)	(6)
	Market Concentration		Annual Quantity		Price	
	High	Low	High	Low	High	Low
Panel A: Sales						
Treat	-0.00564	-0.00651	-0.00247	-0.0109	-0.00831	-0.00282
	(0.00709)	(0.00811)	(0.00736)	(0.00766)	(0.00758)	(0.00727)
% Change	-4.16%	-5.10%	-1.71%	-9.39%	-5.98%	-2.30%
P-value	0.9073		0.3350		0.6386	
Panel B: Conversion Rate						
Treat	-0.0000708	-0.000113	-0.0000432	-0.000154	-0.0000788	-0.000116
	(0.000151)	(0.000165)	(0.000156)	(0.000155)	(0.000135)	(0.000187)
% Change	-1.73%	-3.09%	-0.98%	-4.72%	-2.37%	-2.46%
P-value	0.8177		0.5241		0.9891	
Panel C: Cart Value						
Treat	-0.819	-0.723	-0.242	-1.733	-1.546	0.0417
	(1.258)	(1.608)	(1.215)	(1.714)	(1.567)	(1.159)
% Change	-2.48%	-2.07%	-0.74%	-4.87%	-3.69%	0.16%
P-value	0.9452		0.5023		0.4771	
Observations (Panel A&B)	714,872	529,144	714,642	529,374	714,642	529,374
Observations (Panel C)	2,906	1,905	3,126	1,685	2,347	2,464

[1] Refer to Table F1 for detailed notes.