



Marketing Science Institute Working Paper Series 2026

Report No. 26-127

Generative Search: Evidence from a Large-Scale Field Experiment

Shuang Zheng, Yuting Zhu, Xin Ye and Liang Shen

“Generative Search: Evidence from a Large-Scale Field Experiment ” © 2026

Shuang Zheng, Yuting Zhu, Xin Ye and Liang Shen

MSI Working Papers are Distributed for the benefit of MSI corporate and academic members and the general public. Reports are not to be reproduced or published in any form or by any means, electronic or mechanical, without written permission.

Generative Search: Evidence from a Large-Scale Field Experiment

Shuang Zheng[†] Yuting Zhu[‡] Xin Ye[§] Liang Shen[¶]

May 13, 2026

Abstract

Generative search is an emerging search paradigm that integrates Generative AI (GenAI) into traditional search engines by presenting users with AI-generated responses before conventional search results. Whereas keyword-based search requires consumers to translate their underlying needs into effective keyword queries, generative search lets users express those intentions directly in natural language, a shift made possible by GenAI's new mode of information presentation. This shift moves the consumer–search engine interaction upstream to the stage of problem formulation, rendering empirically observable a previously hidden phase of search behavior: the mapping from problem formulation to keyword articulation. Yet it remains empirically unknown whether generative search increases consumer purchases, and how search behavior changes when it can begin from expressed intent rather than keywords. Using a large-scale field experiment conducted on Meituan, a leading Chinese technology platform, this paper provides empirical evidence on the effectiveness of generative search. We find that generative search significantly increases consumer purchases. Further analysis reveals that generative search facilitates more effective and diverse keyword queries, reduces exploratory browsing and clicking while concentrating evaluation within relevant categories and merchants. The empirical evidence is most consistent with a mechanism in which AI-generated answers provide information about consumers' underlying needs, while also expanding awareness of relevant attributes and shifting attention toward more relevant categories. These findings provide empirical insights for future consumer search models that allow search to originate at the intention-expressing stage.

Keywords: generative search, AI search, generative AI, large language model, consumer search and choice, customer needs.

¹The first two authors are listed alphabetically and contributed equally to the papers as first authors. The authors thank Meituan for providing the empirical setting and data support. The views expressed in this paper are those of the authors and do not necessarily reflect the views of any organization. All errors are ours. The authors thank helpful comments from Xinlei Chen, Yuxin Chen, Ruobing Han, Ying Lei, Yang Li, Emaad Manzoor, Ilya Morozov, Jiwoong Shin, Tianshu Sun, Raphael Thomadsen, Juanjuan Zhang, Shuo Zhang; from seminar participants at Cheung Kong Graduate School of Business, Chongqing University, Hunan University, Google, University of British Columbia, University of Oklahoma, and Xiangjiang Laboratory; and from audiences at 2025 CES Annual Conference, 2025 China India Insights Conference, 2025 Informs MSOM Conference, 2025 Marketing Science Conference and 2025 Yale InsightsOn Conference, 2026 POMS Annual Conference, 2026 Symposium on AI in Marketing, 2026 Workshop on Empirical Operations Management.

[†]Assistant Professor of Marketing, Renmin University of China, School of Business, zhengshuang@rmbs.ruc.edu.cn.

[‡]Assistant Professor of Marketing, National University of Singapore, NUS Business School, y.zhu@nus.edu.sg.

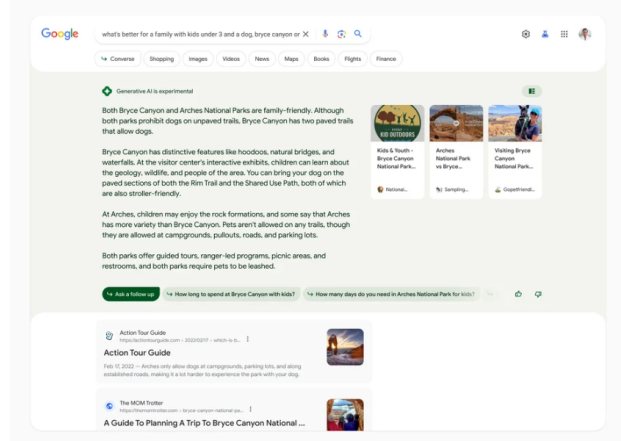
[§]Professor of Management Science and Engineering, Dean of DUT School of Economics and Management, Dalian University of Technology, DUT School of Economics and Management, yexin@dlut.edu.cn.

[¶]Research Scientist, Meituan, shenliang03@meituan.com.

1 Introduction

Since the launch of ChatGPT in late 2022, the adoption of generative AI (GenAI) has accelerated. In response, search-based platforms worldwide have begun integrating conversational AI into their core search products. A common design is to present a GenAI-generated response to a user's query prior to displaying the traditional search results (see Figure 1). This approach has been adopted by major platforms such as Google¹, Bing², and Xiaohongshu (Chen et al. 2025).

Figure 1. A Generative Search Example from Google.



Source: <https://blog.google/products/search/generative-ai-search/>.

In this paper, we examine this emerging search paradigm, termed as *generative search* or *AI search*, that integrates GenAI and large language models (LLMs) into the search results page³, in the context of e-commerce shopping platform. Compared with traditional keyword-based search, generative search introduces two key changes. First, the capabilities of GenAI and LLMs enable a new mode of information presentation: instead of returning a ranked list of items, the system can generate AI-driven, natural language responses that synthesize information from multiple sources in the training data. Second, this new information presentation paradigm fundamentally alters how users interact with the search engine. Traditional keyword-based search relies on well-

¹<https://blog.google/products/search/generative-ai-search/>.

²<https://blogs.bing.com/search/July-2024/generativesearch>.

³<https://www.ibm.com/architectures/hybrid/genai-generative-search>.

formed keyword queries (e.g., categories, brands, or attributes)⁴. In contrast, generative search allows consumers to express their needs or intentions directly in natural language. We define needs or intentions as context-dependent statements that capture a consumer’s underlying goals or concerns to be addressed through a product or service (e.g., Griffin and Hauser 1993; Timoshenko and Hauser 2019). As a result, generative search shifts user-search engine interaction upstream in the consumer decision process from keyword-based queries to the direct articulation of underlying needs. For example, a consumer may input, “I want to find a fun place to take my kids this weekend,” and the system generates responses tailored to this expressed intent. By contrast, under traditional keyword-based search, the consumer must first translate such intentions into specific keywords (e.g., “kids entertainment,” “parks for kids,” or “pizza restaurants”).

By enabling interaction at the stage of consumers’ original needs, generative search can increase purchases by helping consumers better understand which products or services align with their underlying needs through AI-generated answers, thereby facilitating navigation of the search and shopping process. However, prior research shows that prominent positions in search results strongly shape consumer attention and choice (e.g., Narayanan and Kalyanam 2015; Jeziorski and Moorthy 2018; Ursu 2018); by occupying a large share of these prime positions, generative search may crowd out exposure to the broader result set and reduce purchases. Ultimately, whether generative search truly enhances consumer purchases is an empirical question.

What’s more, under traditional keyword-based search, the mapping from problem formulation to keyword articulation occurs off-platform. By bringing this upper-funnel interaction onto the platform and intervening at the stage of consumer needs understanding, generative search makes a previously unobservable stage of consumer search behavior empirically observable. In doing so, it introduces new questions about how consumer search behavior evolves under this paradigm, with important implications for existing theories of consumer search.

Against this backdrop, we address two research questions. First, we provide causal evidence on whether generative search increases consumer purchases. Second, we document how generative search alters consumer search behavior, thereby illuminating the mechanisms through which

⁴In online e-commerce platforms, effective keywords are typically those that clearly describe product categories, brands, or attributes, as search algorithms are designed to match queries against merchant listings encoded with structured keywords. While such keywords may partially reflect consumers’ underlying purchase intentions, more exploratory or question-like queries do not map cleanly onto this keyword space, limiting the effectiveness of traditional retrieval-and-ranking approaches.

it affects purchase outcomes.

To address our research questions, we conducted a large-scale field experiment on Meituan, a leading Chinese technology platform for local consumer products and retail services (see Section 2.1 for company background). The experiment ran on Meituan’s flagship app from September 5 to October 5, 2024. A total of 6,451,767 consumers who had previously submitted at least one intention-expressing search query were randomly assigned to either the treatment group (generative search; 3,227,126 consumers) or the control group (keyword-based search; 3,224,641 consumers). Additional details on the experiment design are provided in Section 2.2.

Our field experiment shows that generative search can significantly boost consumer purchases for involved consumers. Specifically, compared to keyword-based search, generative search results in a 0.174% increase in the total number of orders, a 0.072% rise in the order rate, and a 0.261% growth in gross transaction value per consumer over the course of the experiment.

An analysis of the consumer search journey based on a representative subsample of 90,000 consumers shows that, relative to keyword-based search, generative search leads consumers to submit more queries with greater lexical diversity and to shift from vague, intention-expressing queries toward more concrete, keyword-based queries. Concurrently, treated consumers spend less time on the platform, browse fewer merchants and categories, and click fewer unique merchants and categories, while generating more clicks within each category and exhibiting higher click-through rates (CTR). Together, these empirical patterns may be most consistent with an information-provision mechanism: AI-generated answers help consumers learn about their underlying needs and broaden their awareness of relevant attributes, alongside a complementary shift of attention toward category-level information.

We make two empirical contributions to the literature and to industry practice. First, we provide empirical evidence in the literature that generative search can enhance purchase outcomes. This evidence also offers practical guidance to industry practitioners on the value of investing in generative search technologies. Second, by documenting how generative search reshapes consumer search behavior and translates into higher purchase incidence, we provide empirical insights for future theoretical work on consumer search models that allow search to begin at the intention-expression stage.

Our paper is closely related to the consumer search literature (see comprehensive reviews by

Honka et al. 2019 and Honka et al. 2024) and is most directly aligned with three strands of prior research. The first strand examines how information provision through platform or website design shapes consumer search behavior (e.g., Branco et al. 2012; Ke et al. 2016; Gardete and Hunter 2024). This literature typically studies how exogenously provided information affects consumers' search and purchase outcomes. In this sense, the GenAI-generated responses in our setting can be viewed as a novel and richer form of information provision. The second strand studies query refinement and query suggestion methods that infer consumers' latent intentions within keyword-based search systems (e.g., Sun 2025; Zheng et al. 2025). These approaches rely on the search engine to infer intentions from observed queries, clicks, or browsing patterns, subsequently offering refined keywords or recommendations. The third strand examines consumer search across category consideration and product consideration stages (e.g., Morozov 2023), typically modeling how consumers transition from broad category exploration to focused product evaluation within traditional search environments.

Our key distinction from these three strands of literature is that generative search shifts the locus of consumer–search engine interaction from keyword formulation to problem formulation. By allowing consumers to initiate search through the direct expression of intentions and to receive AI-generated summaries, generative search can reshape how consumers conceptualize their underlying problems from the outset, rather than offering partial guidance within keyword search frame. Consistent with this shift, we document systematic changes in query-entry behavior and reductions in browsing and clicking activity, patterns that stand in contrast to prior findings.

The remainder of the paper is organized as follows. Section 2 describes the company background and experiment design. Section 3 presents the empirical results addressing whether generative search increases consumer purchases. Section 4 examines how generative search alters consumer search behavior, shedding light on the mechanisms underlying the purchase effects. Finally, Section 5 concludes.

2 Field Experiment

2.1 Company Background

We collaborate with Meituan, a leading Chinese technology company and one of the world's largest e-commerce platforms for local products and services⁵. Meituan offers a wide range of services, including food delivery, hotel and travel bookings, entertainment ticketing, bike-sharing, and many other local lifestyle-related offerings. Meituan reported approximately USD 47 billion in annual revenue in 2024⁶.

The product we collaborated on is the company's flagship "Meituan" super app, which delivers an extensive range of local lifestyle products and services⁷. Consumers must register and log in using their mobile phone number to access the platform.

The homepage of the Meituan app (Figure A1 in Online Appendix A) contains three primary components. First, a universal search bar at the top enables consumers to search across all product and service categories (e.g., "Banu Hotpot" as illustrated in Figure A1). Second, users may navigate directly by selecting specific categories such as food delivery, group buying, or hotel bookings. Third, the homepage displays personalized promotions and product recommendations. Our collaboration is with the algorithm team responsible for the universal search function at the top of the homepage. Approximately half of the consumer traffic utilizes this search feature for relevant purchasing activities.

The consumer shopping process on Meituan's cross-category search function closely resembles that of major U.S. platforms such as Amazon, Yelp, and Airbnb. Figure A2 in Online Appendix A illustrates a representative example. When consumers search for "Peking Duck," the app automatically provides related keyword suggestions. Consumers may proceed with their initial query or select one of the recommended alternatives. Once a query is finalized, the app presents a ranked list of merchants, displaying information such as merchant name, delivery availability, location, rating volume and scores, category, average price per person, distance, promotions, featured items (such as the featured dishes in this example), and their prices. As on other search-based platforms, users may reorder the results by criteria such as distance, price, popularity, or

⁵<https://en.wikipedia.org/wiki/Meituan>.

⁶<https://www.meituan.com/en-US/investor/reports>.

⁷<https://i.meituan.com/client/homepmt>.

ratings, and may apply various filters to refine the results. The search team’s objective in designing both keyword suggestions and ranking algorithms is to maximize consumer purchases⁸.

After reviewing the search ranking list, consumers may click on a merchant to access its detailed page and complete a purchase (Figure A3 in Online Appendix A). This page provides comprehensive information, including the merchant’s location, operating hours, photos and videos, recommended items, detailed customer reviews, and the full set of purchasable products or services. A primary source of revenue for Meituan is the commission fee levied on each completed purchase. Commission rates typically range from 5% to 20%.

Notice that both the pre-populated keywords (Sun 2025; Figure A1) and the search-query recommendations (Lei et al. 2025; Zheng et al. 2025; Figure A2) are the platform’s default functions. The pre-populated keywords are non-personalized and derived from trending terms, whereas the search-query recommendations are personalized to each user through auto-completion algorithm.

2.2 Experiment Design

We conducted a field experiment involving 6,451,767 users from September 5 to October 5, 2024⁹. Figure 2 illustrates the experiment design. The generative search feature was introduced only on the search results page; all other elements, including the interface design of all other pages and the back-end search and recommendation algorithms, remained unchanged across conditions. The control group used the platform’s default configuration described in Section 2.1.

There are several experiment design choices worth clarifying. The first concerns when the LLM was activated. In our experiment, the LLM was not triggered for all search queries. Instead, users in the treatment group received an LLM-generated response only when they entered a query that expressed their underlying needs, termed as *intention-expressing search query*. Figure 3 displays an example in which the query “Where to go for a couple’s date” triggered an AI answer. If a treatment-group user entered a standard keyword-based query, the platform displayed the default ranking results. Users in the control group always saw the default ranking results regardless of query type. This design is motivated by two considerations. First, it aligns with the core in-

⁸Sponsored search advertisements are also integrated into the ranking interface, with a fixed ad position. While generative search may influence the performance of these sponsored advertisements, our data do not permit an in-depth investigation, and this question lies outside the primary scope of the current study.

⁹At this time, major LLMs such as ChatGPT were not accessible in mainland China, and DeepSeek had not yet been released.

Figure 2. Experiment Design.



novations of generative search to enable user–search engine interaction at the level of consumers’ underlying needs through AI-generated natural-language answers. Second, it reflects an important practical constraint at the time of implementation: the non-trivial cost of LLM inference¹⁰.

Figure 3. Treatment Group Example.



The identification of *intention-expressing search queries* relied on a classification algorithm developed by Meituan. This model was fine-tuned on a large human-labeled corpus using the company’s proprietary LLM, LongCat, and achieved 94% accuracy. These queries are typically longer

¹⁰For discussion, see: <https://medium.com/aigenverse/why-most-enterprises-underestimate-the-real-cost-of-llm-inference-in-production-9b675ad7860>.

and often take exploratory or question-like forms, making them inherently more ambiguous from a search-algorithm perspective and difficult to handle using traditional keyword- or vector-based retrieval methods.

The second design choice concerns the population included in the experiment. We included only users who had previously submitted at least one *intention-expressing search queries*. There are two reasons for focusing on this group. First, doing so minimizes the need for product education, thereby maximizing effective usage of generative search and reducing the probability of one-sided non-compliance (e.g., Gordon et al. 2019). Second, these users are the ones most likely to benefit from generative search. Consistent with our objective of evaluating whether further investment is warranted, if we observe no meaningful benefits for this group, it would suggest that additional investment in generative search is not justified. From this qualified population, we randomly selected 6,451,767 consumers and assigned 3,227,126 to the treatment (generative search) group and the remainder to the control (keyword-based search) group.

Third, each GenAI response consisted of two components (Figure 3). The first part was a direct, natural-language answer generated by the company’s proprietary LLM, LongCat. The second part offered related search keywords generated by the platform’s keyword-recommendation algorithm based on semantic similarity. Unlike systems such as Google Search, this second component was included due to product-team considerations. To avoid confounding, our analyses exclude sessions in which user actions originated from this second component. Only about 2% of sessions involved clicks on the “Related Searches”; thus our estimates primarily reflect the effect of the direct GenAI response. We acknowledge, however, that the mere presence of the second component may influence user perception or interpretation of the generative answer.

The LLM-generated content did not follow a standardized template; additional examples are provided in Online Appendix B. Moreover, the LLM used in the experiment still had substantial room for improvement. Accordingly, our evaluation of generative search should be viewed as an assessment of the format’s effectiveness under realistic, non-ideal conditions.

3 Does Generative Search Increase Consumer Purchases?

In this section, we present the results on whether generative search increases consumer purchases. We begin in Table 1 with a randomization check to validate the experiment. Due to confi-

dentiality agreement, we can only report the p-values¹¹.

For each consumer, we observe basic demographic and platform-behavioral attributes, including gender, educational background (whether they hold a bachelor's degree), university-student status, sensitivity to discount promotions, total earnings, recent purchase frequency, and consumption level. Sensitivity to discount promotions is a continuous measure between 0 and 1, representing consumers' sensitivity to discounts on the platform. Total earnings, recent purchase frequency, and consumption level are ordinal variables constructed by Meituan. Total earnings takes four ordered values, $\{0, 1, 2, 3\}$, where higher values correspond to higher earning tiers. Recent purchase frequency is categorized into six ordered levels, $\{0, 1, 2, 3, 4, 5\}$, with larger values indicating more frequent purchases in the recent period. Consumption level, a measure of overall spending on the platform, is likewise defined on six ordered levels, $\{0, 1, 2, 3, 4, 5\}$, with higher values representing greater cumulative consumption.

The first three demographic attributes, gender, education, and university student status, are optional fields during account registration. A separate consumer analysis team calculated and inferred missing values for these variables, as well as the remaining four attributes, using comprehensive consumer behavior data available on the platform. These consumer characteristics are widely utilized across various company functions.

To assess balance across treatment and control groups, we compare the first four variables, gender, bachelor's degree, university-student status, and discount-promotion sensitivity, using Welch's *t*-test. For the three ordinal variables, we employ Pearson's Chi-squared test for distributional equality. The resulting p-values are reported in Table 1. We find no significant differences across any consumer characteristics, confirming the effectiveness of the randomization.

In Table 2, we evaluate whether generative search increases consumer purchases. We examine three outcome measures. The first is the total Number of Orders a consumer places during the experiment period. The second is the Order Rate, defined as the ratio of total orders to total search queries for each consumer. The third is Gross Transaction Value, which captures the consumer's total spending over the experiment period.

Due to confidentiality constraints, we can only report relative changes in these outcomes

¹¹Statistics related to customer characteristics and platform behavior are considered sensitive from Meituan's perspective, given the intensely competitive landscape. With a consumer base of this scale, disclosure of such statistics might enable competitors to infer proprietary user attributes and behavioral patterns.

Table 1. Randomization Check ($N = 6,451,767$).

Consumer Traits	P-value
Gender (Male = 1)	0.910
Bachelor's Degree or Higher (Has = 1)	0.587
University Student (Yes = 1)	0.405
Sensitivity to Discount Promotions	0.517
Total Earnings	0.857
Recent Purchase Frequency	0.170
Consumption Level	0.481

Notes. The table reports p-values from the randomization checks ($N = 6,451,767$). Gender, bachelor's degree, university-student status, and discount-promotion sensitivity are tested using Welch's t-tests. Total earnings, recent purchase frequency, and consumption level are ordinal measures and are evaluated using Pearson's chi-squared test for equality of distributions.

for the treatment group (generative search) relative to the control group (keyword-based search), along with the associated p-values. For each outcome, we compute the mean in the treatment and control groups and report the percentage difference, defined as the difference in means divided by the control-group mean. Statistical significance is assessed using Welch's *t*-test, which compares the two group means.

Table 2. Effect of Generative Search on Consumer Purchases.

Outcome Variable	Performance Change	P-value
Number of Orders	0.174%***	0.002
Order Rate	0.072%***	0.008
Gross Transaction Value	0.261%***	0.005

Notes. The table reports relative changes in consumer purchase outcomes between the treatment group (generative search) and the control group (keyword-based search), along with the associated p-values. Relative change is computed as the difference in group means divided by the control-group mean. The sample size in the treatment group is $N_{treatment} = 3,227,126$, and the sample size in the control group is $N_{control} = 3,224,616$. P-values are obtained using Welch's t-test. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Across all outcomes, generative search significantly increases consumer purchases. Relative to keyword-based search, generative search raises the total number of orders by 0.174%, the order rate by 0.072%, and gross transaction value by 0.261%, with all effects statistically significant¹².

We offer two clarifications. First, our estimates should be interpreted as intent-to-treat (ITT) effects (Angrist et al. 1996). Not all users assigned to the treatment group necessarily engaged with generative search. For example, some may never have submitted any *intention-expressing search*

¹²While modest in percentage terms, these effects may be economically meaningful: given Meituan's approximately 770 million annual active users in 2024, the estimates imply roughly 16 million additional orders and about 24 million in incremental gross transaction value per year.

queries during the experiment period. From the firm’s perspective (e.g., Meituan), ITT is also the relevant estimand when evaluating whether generative search is worth investing in, because such features are typically deployed at the system level rather than targeted to specific individuals (e.g., Liu et al. 2026), and firms must account for the realistic possibility that some users will not use the feature¹³. Second, caution is warranted when extrapolating the magnitude of our effects to the broader consumer population, given our targeted consumers. Relative to Meituan’s overall user base, our sample is more highly educated and therefore likely to have greater prior exposure to AI. This may make these users more inclined to have submitted *intention-expressing search queries* and be included in our experiment sample¹⁴.

4 How Does Generative Search Change Consumer Search Behavior?

In this section, we present empirical evidence that addresses our second research question: how generative search reshapes consumer search behavior and thereby sheds light on the mechanisms underlying the observed increase in purchases. We examine behavioral changes along four margins: session-level engagement metrics, query entry, browsing, and clicking. Analyzing these intermediate search behavior responses allows us to trace how generative search alters the consumer decision environment and to empirically document the pathways through which it affects purchasing behavior.

Due to confidentiality and data privacy concerns, we do not have access to these search behavior data for the full experiment sample. Instead, Meituan provided a subset consisting of 45,000 randomly selected consumers from the treatment group and 45,000 from the control group (90,000 consumers in total). The corresponding randomization checks and purchase-effect estimates for this subsample are reported in Online Appendix C¹⁵. The randomization check also passes for this subsample, and consumer purchases increase as well.

Table 3 reports the effects of generative search on consumer search behavior. All analyses are conducted at the user level using the provided sub-sample ($N = 90,000$). We estimate the

¹³If, instead, a firm were interested in selectively enabling generative search for certain users, then the average treatment effect on the treated (ATT) would be more relevant and may exceed our ITT estimates, which includes both exposed and unexposed users.

¹⁴As AI adoption increases, our findings may become more representative of the broader population. Nonetheless, assessing external validity across other user segments remains an important direction for future research.

¹⁵We also conduct exposure-related analyses using this subsample. Notably, 39,305 of the 45,000 users in the treatment group never submitted any *intention-expressing search queries* during the experiment period. In Online Appendix C, we further report ATT estimates that instrument actual exposure with treatment assignment (Gordon et al. 2019).

following specification:

$$y_i = \gamma + \theta \text{GenSearch}_i + \epsilon_i, \quad (1)$$

where y_i denotes a measure of consumer i 's search behavior. The coefficient θ captures the effect of generative search relative to keyword-based search. For each outcome, Table 3 reports the estimated θ along with robust standard errors.

Table 3. Effect of Generative Search on Consumer Search Behavior.

Outcome Variables	Effect (θ)	Standard Error
Number of Sessions	0.0090	(0.0066)
Total Stay Time	-0.0371**	(0.0146)
Number of Search Queries	0.0132*	(0.0077)
Number of Distinct Search Queries	0.0135*	(0.0076)
Number of Intention-Expressing Search Queries	-0.0084***	(0.0030)
Number of Keyword-Based Search Queries	0.0131*	(0.0078)
Number of Sessions with Intention-Expressing Search Queries	-0.0068***	(0.0023)
Number of Browsed Merchants	-0.0189*	(0.0104)
Number of Browsed Merchant Categories	-0.0159*	(0.0090)
Number of Browsed Merchants Per Category	-0.0186	(0.0132)
Number of Total Clicks	-0.0074	(0.0097)
Number of Unique Clicked Merchants	-0.0228**	(0.0093)
Number of Clicked Merchant Categories	-0.0198**	(0.0086)
Number of Clicked Merchants Per Category	0.0590***	(0.0122)
Number of Unique Clicked Merchants Per Category	-0.0053	(0.0040)
CTR	0.0061***	(0.0015)
Unique CTR	0.0005	(0.0006)

Notes. This table reports the effects of generative search on various consumer search behavior outcomes ($N = 90,000$). For each outcome variable, we present the estimated treatment effect (θ) from Equation (1), along with the corresponding robust standard error. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Next, we first introduce definitions for each outcome variable and discuss the magnitude of the corresponding effects. We then synthesize the evidence across all behavioral margins to propose plausible, while distinguishing less likely, mechanisms underlying the observed patterns.

The first two outcome variables capture session-level engagement metrics. The first variable measures the total number of sessions a user initiated during the experiment period. The second variable measures the total time a user spent on the platform during the experiment period. Both outcome variables are log-transformed. The results reveal that generative search did not induce users to visit the platform more frequently, as evidenced by the statistically non-significant change in the total number of sessions. However, generative search changed the user experience within

each session. The significant negative coefficient on total stay time ($\theta = -0.0371$) indicates that users exposed to generative search spent less time on the platform, corresponding to a 3.64% reduction in total stay time relative to the keyword-based search condition¹⁶. Notice that this is consistent with Meituan’s role as an e-commerce platform, which seeks to streamline users’ search paths and reduce search costs to facilitate faster transaction completion, contrasting with social media platforms that monetize user attention¹⁷.

The second set of outcome variables captures changes in consumers’ query-entry behavior. We consider five measures. The first is the total number of search queries a consumer entered during the experiment period, and the second is the number of distinct search queries, where two queries are considered distinct unless they are identical in every character. The third and fourth measures capture the total number of intention-expressing queries and keyword-based queries, respectively. These query types were identified using the classification algorithm described in Section 2.2. The final measure is the total number of sessions that contained at least one intention-expressing search query. All outcome variables are log-transformed. The results reveal notable shifts in query-entry behavior following the introduction of generative search. First, treated users submitted 1.33% more total search queries ($\theta = 0.0132$) and 1.36% more distinct queries ($\theta = 0.0135$), suggesting a modest increase in query volume and variety. Second, users shifted away from intention-expressing search queries toward keyword-based search queries along both the intensive and extensive margins. Specifically, we observe a significant reduction of 0.84% in the total number of intention-expressing search queries ($\theta = -0.0084$), as well as a 0.68% decrease in the number of sessions containing such queries ($\theta = -0.0068$). Taken together, these patterns indicate that the observed increase in query volume and variety was primarily driven by more frequent and more varied use of keyword-based search queries.

To examine changes in consumer browsing behavior, we consider three outcome measures. The first is the total number of merchants browsed by a user during the experiment. A merchant is defined as browsed if it appears on the consumer’s screen, reflecting the mobile app’s interface design and the assumption that displayed merchants are generally visible to attentive users.

¹⁶Given that the outcome variable is log-transformed, the percentage difference between generative search and keyword-based search is calculated as $(e^{\theta} - 1) \times 100\%$.

¹⁷On e-commerce platforms, balancing revenue generated from organic purchases and sponsored advertisements is crucial (e.g., Lu et al. 2025). From a long-term perspective, especially in highly competitive markets, streamlining users’ search paths can improve platform efficiency and contribute to the platform’s long-term health.

Although this definition may modestly overstate actual attention, it is applied uniformly across treatment and control groups and therefore does not bias between-group comparisons. We additionally report the number of categories associated with browsed merchants and the average number of browsed merchants per category. The platform employs a hierarchical merchant category classification system, and our analysis focuses on the finest-grained category level. The first two outcome variables are log-transformed. The empirical results show that treated users browsed 1.87% fewer merchants in total ($\theta = -0.0189$), driven primarily by a 1.58% reduction in the number of merchant categories browsed ($\theta = -0.0159$)¹⁸.

Lastly, we report seven outcome measures to characterize clicking behavior. These include the total number of clicks and the number of unique merchants clicked by a consumer during the experiment period. We also report the number of categories associated with clicked merchants, the average number of clicked merchants per category, and the average number of unique clicked merchants per category. In addition, we examine changes in click-through rates (CTR), including the overall CTR, defined as total clicks divided by total merchants browsed, and the unique CTR, defined as the number of unique merchants clicked divided by the number of unique merchants browsed. The first three outcome variables are log-transformed. The empirical results show that, although total clicks declined modestly and insignificantly, the number of unique merchants clicked decreased by 2.25% ($\theta = -0.0228$). Decomposing this effect, treated users clicked 1.96% ($\theta = -0.0198$) fewer merchant categories, while generative search increased the number of merchants clicked per category by 0.059 on average; both effects are statistically significant. Consistent with the browsing results, the overall pattern was primarily driven by changes at the category level, with only modest changes within each category. Notably, despite reductions in both browsing and clicking, the CTR increased by 0.61 percentage points, whereas the change in unique CTR was small and statistically insignificant¹⁹.

Several plausible mechanisms may explain the empirical findings in Table 3. We discuss each in turn, highlighting both the patterns they can account for and those they cannot. These

¹⁸Reduced browsing is not necessarily detrimental to the health of an e-commerce marketplace, because a substantial portion of browsing in search environments reflects search frictions and consumers' difficulty in identifying relevant options, rather than productive discovery.

¹⁹CTR is a useful measure for e-commerce platforms because it captures whether the platform successfully matches user needs with relevant products. In our setting, the combination of reduced browsing depth and higher CTR is consistent with improved search quality: users achieve better outcomes while browsing fewer merchants, suggesting that the generative search helps consumers identify more relevant options with less unnecessary exploration.

limitations, in turn, motivate the specific insights we offer for consumer search models that allow search to begin at the intention-expressing stage.

The most natural explanation is that AI-generated answers are *informative*: they tell users something about the distribution of options that match their needs. In the search and learning literature (e.g., Rothschild 1974; Dzyabura and Hauser 2019), this corresponds to consumers updating beliefs about the distribution of match values, here simultaneously across categories, brands, and products due to the multi-dimensional information provided by AI-generated answers²⁰. If learning becomes more efficient, we should see more purchases and a higher CTR alongside less browsing, less clicking, and shorter time on the platform, which is what we find.

Query-entry behavior is harder to pin down because consumer search models say little about how keyword queries, that is, specific categories, brands, or attributes in e-commerce search, are formed²¹. A standard learning model would predict fewer queries as uncertainty falls, yet we observe more keyword queries and greater query variety. One plausible reconciliation is that richer, multi-dimensional information broadens consumers' *awareness* of relevant attributes (Honka et al. 2017). As for the shift from broad intention-expressing queries toward more precise keyword-based queries, this may arise because consumers often begin one or multiple sessions with a relatively stable underlying need (e.g., "finding a dating venue"). As AI-generated answers accelerate learning about relevant categories and attributes, the need to repeatedly express that underlying intention to the platform declines along both the extensive and intensive margins.

What remains unexplained under the information-provision perspective is why category-level reduction accounts for most of the decline in browsing and clicking behavior. A complementary mechanism may lie in attention reallocation: even when AI-generated answers continue to supply information, users with bounded rationality may shift how they allocate cognitive resources in response (Caplin 2016; Bordalo et al. 2022). Consistent with this, Bhui and Jiao (2023) documents that consumers economize attention by prioritizing category-level information.

Other mechanisms are possible, but they are less consistent with the full set of empirical patterns. One alternative is persuasion. AI-generated answers might directly increase the attrac-

²⁰Notice that this multi-dimensional information presentation forms contrast with traditional keyword-level interventions (e.g., Sun 2025), which document increased clicking behavior.

²¹Unlike filtering, queries allow consumers to actively define a personalized search space without relying on pre-specified attributes.

tiveness of certain options, similar to persuasive advertising (Bagwell 2007), or they might distort beliefs (Kamenica and Gentzkow 2011). Neither fits cleanly. AI answers typically span multiple dimensions rather than pushing a single product or category, making it unclear what form of utility shift could jointly produce the changes we observe in query entry, browsing, and clicking. Moreover, strategic belief distortion alone is hard to reconcile with the increase in purchases, since distorted beliefs need not raise demand in any systematic direction.

AI avoidance (Dietvorst et al. 2015) and interface crowding are also possibilities, but neither, on its own, predicts higher purchases. They could matter if the discomfort prompts users to think more deliberately and translate their needs into more precise search terms (Guo and Zhang 2012). This deliberation-based interpretation matches much of the evidence, but two pieces don't fit: deliberation usually takes time, yet total stay time falls, and the mechanism gives no particular reason for deliberation to focus on categories.

A few other channels can also be ruled out. Novelty effects typically raise engagement; we see the opposite. Cognitive style matching (Hauser et al. 2009) likewise tends to increase engagement rather than reduce it. Supply constraints are also implausible: Meituan's 2024 report puts the platform at roughly 1.45 million active merchants, and search results outside exact-name queries can return thousands of merchants.

In sum, the empirical patterns are most consistent with an information-provision mechanism: AI-generated answers help consumers better understand their underlying needs and broaden awareness of relevant attributes, which, together with a complementary reallocation of attention toward category-level information, reshapes consumer search behavior. One limitation is that we do not observe the content of AI-generated answers, due to production-cost considerations, which limits our ability to disentangle these channels more directly. More broadly, our findings point to a key gap in existing consumer search models: how to formally characterize the process by which consumers translate underlying needs into expressed queries, whether categories, brands, or product attributes²².

²²Consistent with this view, documenting how users who initially do not articulate their needs through search engines gradually shift toward more intention-expressing queries would provide complementary evidence.

5 Conclusion

In this paper, we provide empirical evidence in literature on whether generative search can increase consumer purchases and provide empirical documentation of how generative search changes consumer search behavior using a large-scale field experiment conducted on a leading Chinese platform Meituan. Our field experiment shows that generative search can significantly boost consumer purchases for consumers who had submitted intention-expressing search queries. Further analysis reveals that generative search facilitates more effective and diverse keyword queries, reduces exploratory browsing and clicking while concentrating evaluation within relevant categories and merchants. Together, these findings can inform firms' investment decisions regarding generative search and contribute empirical insights for future theoretical work on consumer search models that allow search to begin at the intention-expression stage.

References

- Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434), 444–455.
- Bagwell, K. (2007). The economic analysis of advertising. *Handbook of Industrial Organization*, 1701–1844.
- Bhui, R. and P. Jiao (2023). Attention constraints and learning in categories. *Management Science* 69(9), 5394–5404.
- Bordalo, P., N. Gennaioli, and A. Shleifer (2022). Salience. *Annual Review of Economics* 14, 521–544.
- Branco, F., M. Sun, and J. M. Villas-Boas (2012). Optimal search for product information. *Management Science* 58(11), 2037–2056.
- Caplin, A. (2016). Measuring and modeling attention. *Annual Review of Economics* 8, 379–403.
- Chen, J., Q. Dong, H. Li, X. He, Y. Gao, S. Cao, Y. Wu, P. Yang, C. Xu, Y. Hu, Q. Ai, and Y. Liu (2025). Qilin: A multimodal information retrieval dataset with app-level user sessions. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '25*, pp. 3670–3680.
- Dietvorst, B. J., J. P. Simmons, and C. Massey (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144(1), 114–126.
- Dzyabura, D. and J. R. Hauser (2019). Recommending products when consumers learn their preference weights. *Marketing Science* 38(3), 417–441.
- Gardete, P. M. and M. Hunter (2024). Multiattribute search: Empirical evidence and information design. *Marketing Science* 43(5), 1052–1080.
- Gordon, B. R., F. Zettelmeyer, N. Bhargava, and D. Chapsky (2019). A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science* 38(2), 193–225.
- Griffin, A. and J. R. Hauser (1993). The voice of the customer. *Marketing Science* 12(1), 1–124.

- Guo, L. and J. Zhang (2012). Consumer deliberation and product line design. *Marketing Science* 31(6), 995–1007.
- Hauser, J. R., G. L. Urban, G. Liberali, and M. Braun (2009). Website morphing. *Marketing Science* 28(2), 202–223.
- Honka, E., A. Hortacsu, and M. A. Vitorino (2017). Advertising, consumer awareness, and choice: Evidence from the u.s. banking industry. *Rand Journal of Economics* 48(3), 611–646.
- Honka, E., A. Hortacsu, and M. Wildenbeest (2019). Empirical search and consideration sets. *Handbook of the Economics of Marketing*, 193–258.
- Honka, E., S. Seiler, and R. Ursu (2024). Consumer search: What can we learn from pre-purchase data? *Journal of Retailing* 100(1), 114–129.
- Jeziorski, P. and S. Moorthy (2018). Advertiser prominence effects in search advertising. *Management Science* 64(3), 1365–1383.
- Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review* 101(6), 2590–2615.
- Ke, T. T., Z.-J. M. Shen, and J. M. Villas-Boas (2016). Search for information on multiple products. *Management Science* 62(12), 3576–3603.
- Lei, X., Y. Chen, and A. Sen (2025). Trade-offs in leveraging external data capabilities: Evidence from a field experiment in an online search market. *Management Science*.
- Liu, M., X. Tang, S. Xia, S. Zhang, Y. Zhu, and Q. Meng (2026). Algorithm aversion: Evidence from ridesharing drivers. *Management Science* 72(1), 193–203.
- Lu, H., L. Zhang, and Y. Zhu (2025). The power of linear programming in sponsored listings ranking: Evidence from a large-scale field experiment. Working paper.
- Morozov, I. (2023). Measuring benefits from new products in markets with information frictions. *Management Science* 69(11), 6988–7008.
- Narayanan, S. and K. Kalyanam (2015). Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science* 34(3), 388–407.
- Rothschild, M. (1974). Searching for the lowest price when the distribution of prices is unknown. *Journal of Political Economy* 82(4), 689–711.
- Sun, C. (2025). How does prepopulating search bars with keywords affect online consumer behavior? a field experiment. *Marketing Science* 44(6), 1217–1231.
- Timoshenko, A. and J. R. Hauser (2019). Identifying customer needs from user-generated content. *Marketing Science* 38(1), 1–20.
- Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science* 37(4), 530–552.
- Zheng, S., S. J. Tong, H. E. Kwon, G. Burtch, and X. Li (2025). Frontiers: Recommending what to search: Sales volume and consumption diversity effects of a query recommender system. *Marketing Science* 44(3), 516–524.

Online Appendix

A Meituan APP

Figure A1. The First Page of Meituan APP

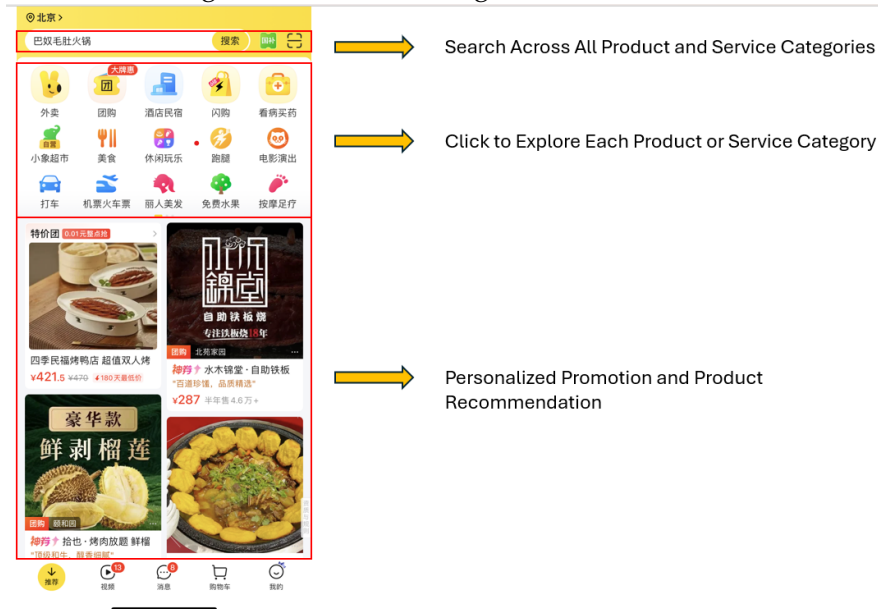


Figure A2. Consumer Shopping Process

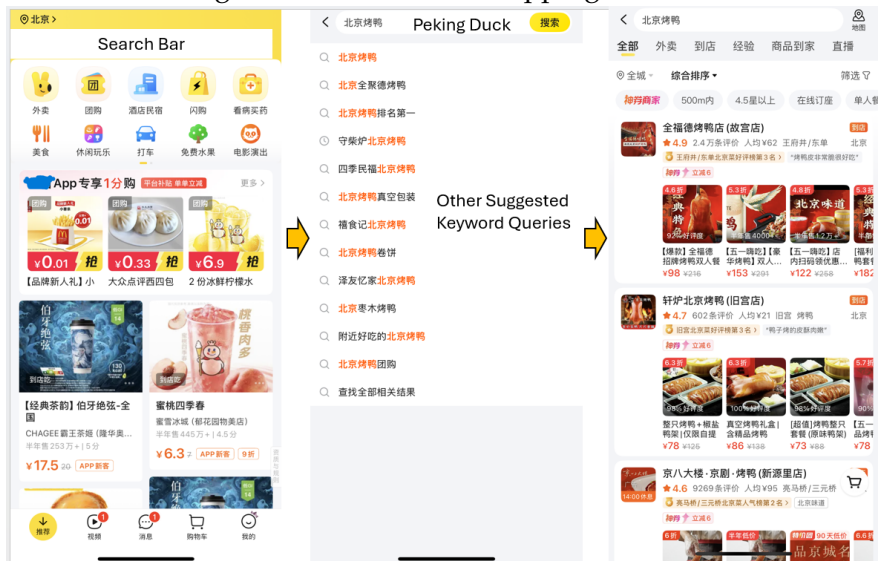
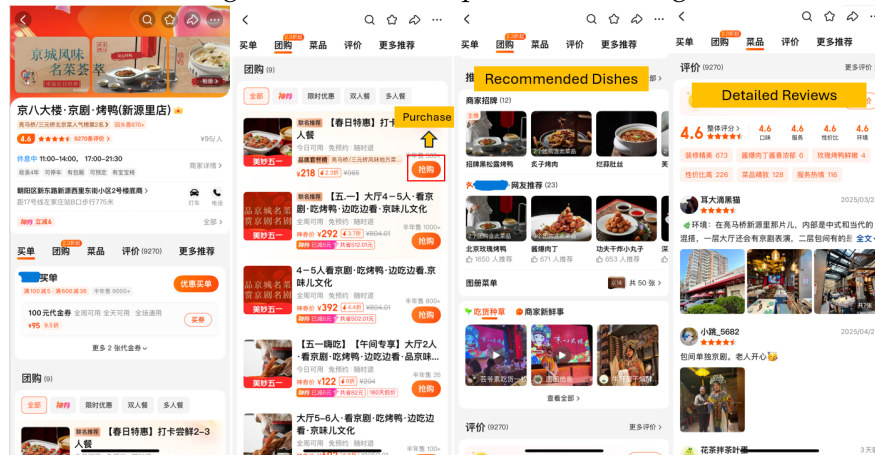


Figure A3. An Example Merchant Page



B GenAI-generated Answers Examples

- **Example 1:**

Intention-expressing search query: Restaurants suitable for birthday celebrations in Beijing

GenAI-generated Answers: In Beijing, there are numerous restaurants suitable for birthday celebrations, combining aesthetically pleasing environments, delicious cuisine, and attentive service. Rouwenwu: A Japanese-style barbecue restaurant located in Xidan, with an average consumption of 200 RMB. Featuring a strong Japanese ambiance and elegant decor, the restaurant offers complimentary longevity noodles and desserts for birthday celebrations, accompanied by thoughtful service. K Thai • Starry Tent Restaurant • Hong Kong Hot Pot: Situated in Sanlitun, this internet-famous restaurant is renowned for its romantic ambiance and signature dishes such as abalone, sea cucumber, and fish maw chicken hot pot. Ideal for photography and gatherings, it is particularly suitable for birthday parties among friends. Sashibi Western Cuisine: Specializing in Mexican cuisine, this restaurant is located at No. 74 Wudaoying Hutong, Andingmen, Dongcheng District, with an average consumption of 93 RMB. The elegant atmosphere, combined with in-house facilities like TVs, gaming consoles, and board games, makes it perfect for young adults' gatherings.

- **Example 2:**

Intention-expressing search query: Which beverages from Heytea are considered to be the most palatable?

GenAI-generated Answers: The following beverages are repeatedly highlighted across various reviews and user feedback as popular choices at Heytea: The "Fleshy" Fruit Series: Beverages such as Fleshy Grape and Fleshy Mango Pomelo Sago are favored by many customers for their generous fruit content and refreshing taste. The Cheese Foam (Zhizhi) Series: These fruit teas topped with a layer of cheese foam combine the light sweetness of fruit with the savory creaminess of cheese, offering a balanced and multi-layered flavor experience. Special Recommendation: For those who enjoy distinctive flavors, the Hot Taro Bobo Milk is worth trying. This drink stands out for its rich and creamy taro taste, offering a warm and indulgent sensory experience.

- **Example 3:**

Intention-expressing search query: Places of interest in Xi'an

GenAI-generated Answers: Appreciate chime bell performances at the Bell Tower and savor delicacies in the Muslim Quarter. Watch owl doll shows and taste biangbiang noodles in Datang Ever-bright City. Overlook the night city from the Big Wild Goose Pagoda and admire the musical fountain. Explore Xiaozhai Seg Shopping Center and walk to the Big Wild Goose Pagoda. Xi'an, equipped with transportation hubs like Xi'an North Railway Station and Xi'an Xianyang International Airport, ensures convenient travel for visits to this ancient capital.

C Additional Analysis of the Subsample

In this section, we present additional analyses conducted using the subsample. In Table A1, we provide the randomization check results for the sub-sample ($N = 90,000$), along with the group means for gender, bachelor's degree, university-student status, and discount-promotion sensitivity, as well as the distribution across levels for total earnings, recent purchase frequency, and consumption level, together with p-values from randomization checks.

Table A1. Randomization Check ($N = 90,000$).

Consumer Traits	Treatment	Control	P-value
Gender (Male = 1)	0.414	0.412	0.750
Bachelor's Degree or Higher (Has = 1)	0.670	0.668	0.617
University Student (Yes = 1)	0.042	0.040	0.197
Sensitivity to Discount Promotions	0.527	0.529	0.384
Total Earnings (Level = 0)	0.406	0.403	0.538
Total Earnings (Level = 1)	0.562	0.565	
Total Earnings (Level = 2)	0.024	0.025	
Total Earnings (Level = 3)	0.007	0.007	
Recent Purchase Frequency (Level = 0)	0.049	0.048	0.173
Recent Purchase Frequency (Level = 1)	0.183	0.185	
Recent Purchase Frequency (Level = 2)	0.117	0.117	
Recent Purchase Frequency (Level = 3)	0.408	0.406	
Recent Purchase Frequency (Level = 4)	0.139	0.136	
Recent Purchase Frequency (Level = 5)	0.104	0.109	0.327
Consumption Level (Level = 0)	0.063	0.065	
Consumption Level (Level = 1)	0.178	0.175	
Consumption Level (Level = 2)	0.118	0.121	
Consumption Level (Level = 3)	0.212	0.208	
Consumption Level (Level = 4)	0.319	0.320	
Consumption Level (Level = 5)	0.111	0.112	

Notes. The table reports, for the sub-sample ($N = 90,000$), group means for gender, bachelor's degree, university-student status, and discount-promotion sensitivity, as well as the distribution across levels for total earnings, recent purchase frequency, and consumption level, together with p-values from randomization checks. Differences in the first set of variables are assessed using Welch's t-tests, while the ordinal variables are evaluated using Pearson's chi-squared tests for equality of distributions.

Table A2. Generative Search Effects on Purchases.

	(1)	(2)	(3)	(4)
Generative Search	0.0224***	0.0021***	0.1767***	0.0165***
	(0.0061)	(0.0004)	(0.0478)	(0.0035)
Observations	90,000	90,000	90,000	90,000

Notes. The table reports estimates of the effects of generative search on purchases using the 90,000-user subsample. The dependent variables are (1) number of orders (ITT), (2) order rate (ITT), (3) number of orders (ATT), (4) order rate (ATT). Robust standard errors are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Table A2, we report the effects of generative search on consumer purchases using the

90,000-user subsample. Column (1) presents the ITT on the number of orders (log-transformed), while Column (2) reports the ITT on the order rate. Column (3) presents the ATT on the number of orders (log-transformed), while Column (4) reports the ATT on the order rate. The ATT estimates are obtained by instrumenting actual exposure with treatment assignment (Gordon et al. 2019). Consistent with the full-sample results, generative search significantly increases consumer purchases in this subsample under both ITT and ATT. As expected, the ATT estimates exceed the ITT estimates.

Table A3. Effect of Generative Search on Consumer Search Behavior (ATT).

Outcome Variables	Effect	Standard Error
Number of Sessions	0.0708	(0.0523)
Total Stay Time	-0.2933**	(0.1156)
Number of Search Queries	0.1046*	(0.0610)
Number of Distinct Search Queries	0.1064*	(0.0597)
Number of Intention-Expressing Search Queries	-0.0663***	(0.0238)
Number of Keyword-Based Search Queries	0.1037*	(0.0614)
Number of Sessions with Intention-Expressing Search Queries	-0.0540***	(0.0186)
Number of Browsed Merchants	-0.1493*	(0.0823)
Number of Browsed Merchant Categories	-0.1259*	(0.0714)
Number of Browsed Merchants Per Category	-0.1467	(0.1045)
Number of Total Clicks	-0.0587	(0.0768)
Number of Unique Clicked Merchants	-0.1805**	(0.0740)
Number of Clicked Merchant Categories	-0.1565**	(0.0682)
Number of Clicked Merchants Per Category	0.4666***	(0.0968)
Number of Unique Clicked Merchants Per Category	-0.0422	(0.0317)
CTR	0.0486***	(0.0122)
Unique CTR	0.0036	(0.0048)

Notes. This table reports the ATT of generative search on various consumer search behavior outcomes ($N = 90,000$). The reported effects correspond to the second-stage coefficients from a two-stage least squares (2SLS) specification that instruments actual exposure with treatment assignment. Robust standard errors are also reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3 reports the ATT of generative search on various consumer search behavior outcomes, corresponding to Table 3 in the main manuscript. The reported effects are second-stage coefficients from a two-stage least squares (2SLS) specification that instruments actual exposure with treatment assignment. As expected, the ATT estimates exceed their ITT counterparts, and all coefficient signs and levels of statistical significance are consistent with those in Table 3.