Designing Generative AI Search: Search Activity and Demand for Creator Content

Hongxian Huang[1] Jinglong Zhang[3] Jason Chan[2] Anindya Ghose[1]

[1]New York University, Stern School of Business

[2]University of Minnesota, Carlson School of Management

[3]University of Science and Technology of China, School of Management

Abstract

Content creation platforms rely on creators to supply content and on users to provide the attention that sustains content creation. Generative AI search changes this relationship by allowing platforms to answer queries directly alongside helping users discover creator content. We combine a randomized field experiment involving 100,000 users on a leading Chinese question-and-answer platform with a dynamic structural model of search to examine the consequences for search activity and content demand. Continued access to an established AI search feature produces a scale–intensity trade-off: users search more frequently but browse and open less user-generated content (UGC) per session. The expansion in search more than offsets the decline in per-session consumption, increasing aggregate UGC browsing and openings. Broader topical exploration and additional queries pursuing new information tasks or related follow-ups support exploration and continued information seeking as behavioral channels behind increased search activity. Meanwhile, lower opening rates among previewed UGC, together with more pronounced declines for information needs that AI answers can readily satisfy, support information substitution as a channel behind reduced per-session content demand. The model connects these responses through information acquisition and expected search value: AI-provided information reduces the benefit of consuming additional UGC within a session, while higher expected search value encourages further search. Counterfactual simulations evaluate three sets of AI search designs—AI placement and information amount, AI availability across query intents, and information labels—with the goal of improving both search activity and demand for creator content.

Keywords: Generative AI Search; Content Creation Platforms; Search Activity; Demand for Creator Content; User-Generated Content; Platform Design.

1 Introduction

Generative AI Search integrates AI-generated responses into search results, alongside links to underlying content. Users can obtain information from the AI response or consult its sources. Figure 1 illustrates this combination in Google's AI Overviews.

Content creation platforms such as Reddit, Rednote, and Stack Exchange have introduced Generative AI Search (Figure 2). These platforms rely on creators to provide user-generated content (UGC) that attracts users. Users, in turn, provide the audience and attention that support creators' monetary rewards, recognition, and community engagement. Search connects these two sides. By adding its own AI answers, the platform does more than help users discover creator content: it also supplies information that may compete with that content for attention. Generative AI Search can therefore change both how much users search and how much of that activity reaches creators.

Different outcomes are possible. Search activity and demand for creator content may both grow if AI encourages enough additional search. Alternatively, users may search more but open less creator content overall because AI answers meet more of their needs directly. If search activity does not grow, or even declines, substitution toward AI may further reduce demand for creator content. These possibilities depend on two responses: whether AI encourages users to search, and whether it complements or replaces UGC within each search. Understanding both is essential for platforms seeking to improve search while sustaining demand for creator content.

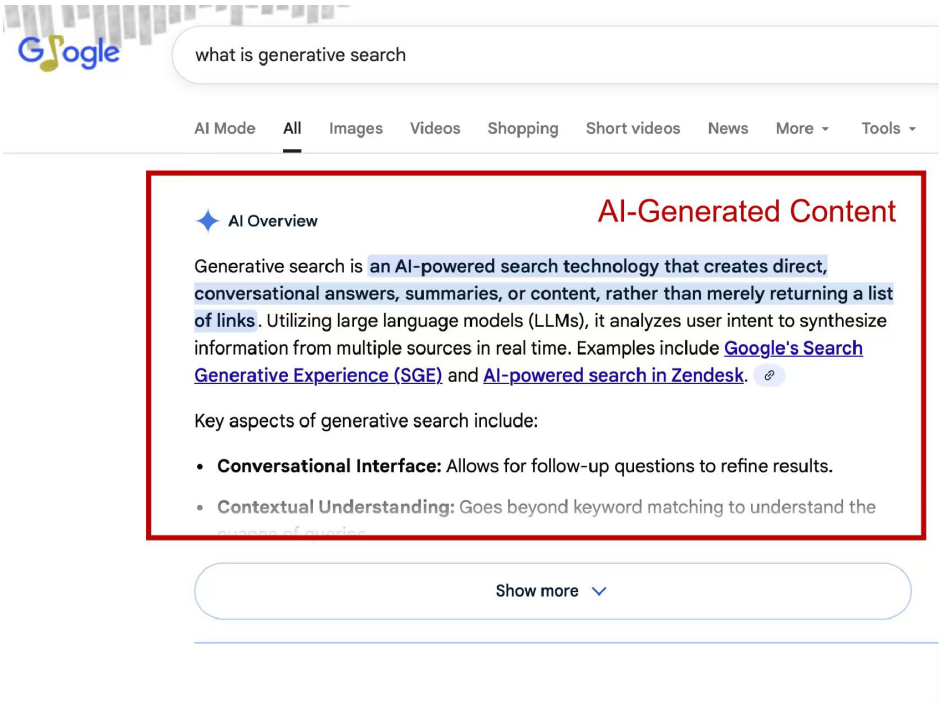


Figure 1: An Example of Google AI Overviews. The red box highlights the AI-generated response on the search-results page.



(a) Reddit

(b) Rednote

(c) Zhihu

Figure 2: Examples of Generative AI Search on Content Creation Platforms. Red annotations identify AI-answer previews and provide translations of the Chinese search queries.

Existing field experiments find fewer clicks to external websites under Google AI Overviews (Agarwal and Sen, 2026; Wang et al., 2026), and greater search participation without a detectable increase in aggregate content consumption on a short-video platform (Bai et al., 2026). We study a question-and-answer platform, where an AI answer and creator-supplied content can supply the information users seek. Beyond evaluating access, we examine how the platform's control over AI placement, availability, and disclosure changes search activity and demand for creator content. We organize the analysis around three research questions:

RQ1. How does Generative AI Search affect users' search activity, and what behavioral channels help explain these effects?

RQ2. How does Generative AI Search affect demand for creator content, both per search and in aggregate, and what behavioral channels help explain these effects?

RQ3. How do the platform's AI search design choices affect search activity and demand for creator content?

We combine a randomized field experiment with a structural model of search. The experiment involved 100,000 Zhihu users from March 30 to April 24, 2025. All had access to the established feature before randomization; the platform retained access for treatment users and removed it for control users. Treatment users could encounter an AI-answer preview for eligible queries and open the full answer. Linked records capture query submissions, preview views, content openings, and navigation to other creator content. We estimate the effect of assigned access by comparing user-level outcomes between the randomized groups.

For RQ1, access to Generative AI Search increases how often and how broadly users search. Users also search over a longer span of the experiment and are more likely to search in both its first and second halves. Most additional same-day queries are classified as related follow-ups or distinct information tasks rather than query reformulations. Together with the wider topical coverage, this query composition supports continued and exploratory information seeking.

For RQ2, treatment users view fewer UGC previews and open fewer UGC pages per observed search session. They also open a smaller fraction of the UGC pages they preview, measured by the click-through rate (CTR). For the average user, a larger share of searches includes UGC previews but no UGC opening. These responses vary with the information users seek. Clicking and CTR decline significantly for factual, procedural, and evaluation/recommendation queries, whose information needs AI answers can more readily satisfy. In contrast, the effects are not statistically significant for experience/opinion queries, where users may place greater value on personal accounts and perspectives. Together, the preview-to-opening responses and intent differences support information substitution: AI answers can provide information that users would otherwise seek by opening UGC.

Lower consumption per search does not translate into lower total demand for creator content. Both UGC previews and openings increase over the experiment because search expansion more than offsets the per-session declines. The CTR decline is larger among the first five original results, which the platform prioritizes using relevance and quality, but aggregate opening gains have similar point estimates in absolute size for the first five and later results. Access also increases the number of sessions leading from an initial UGC page to additional creator content. The overall pattern is a scale–intensity trade-off: users consume less creator content in each search, yet their additional searches increase total consumption.

To address RQ3, the structural model connects these responses through information acquisition and prospective search value. Opening AI or UGC provides information at a reading cost; accumulated information reduces the benefit of further openings within the session. The expected value of search affects both daily participation and subsequent-query initiation. We estimate these choices jointly using linked records of users' daily search activity, query submissions, result-preview views, and content openings, together with text-based measures of AI information and expected reading length. The model reproduces the qualitative scale–intensity pattern.

Counterfactual simulations compare alternative designs with the platform's existing AI-enabled search design. We examine where an AI answer appears and how much information it provides, which query intents receive AI answers, and labels that indicate an answer's information amount before opening. Moving a more informative AI answer lower in the results can increase UGC openings per session but reduce search activity. Reallocating the same AI-availability budget across intents or adding truthful information labels can increase total UGC openings from search results. These designs offer aggregate gains without an established joint improvement in search activity and UGC demand per session.

This paper makes three contributions. First, we extend evidence on Generative AI Search beyond query volume and search result clicks to topical coverage, successive-query relationships, and the relationship between recorded previews, content openings, and subsequent navigation. Comparisons across original result ranks and search intents trace differences in browsing, clicking, and CTR. These patterns provide behavioral evidence consistent with information substitution. Second, we connect sequential search to search initiation

in a setting where information is itself consumed. This links interface design to both demand for creator content per session and the frequency of search. Third, we examine how a platform-owned information source redirects attention through placement, allocation, and disclosure, distinguishing engagement with AI from demand for creator content. These behavioral channels may also operate on other content creation platforms and for comparable informational queries on general-purpose search engines, where AI-generated responses and underlying sources can address the same information need.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 introduces the platform, field experiment, data, and measures. Section 4 presents the experimental evidence on search activity and demand for creator content, including behavioral patterns that help explain these responses. Section 5 develops the structural model. Section 6 explains identification and estimation, and Section 7 presents the parameter estimates and model-implied predictions. Section 8 evaluates alternative AI search designs through counterfactual simulations. Section 9 discusses business implications, limitations, and directions for future research. The appendices provide additional measurement details, robustness analyses, and supporting model and simulation results.

2 Related Literature

2.1 Consumer Search and Interface Design

Search models explain how information-acquisition costs shape consideration and choice. Kim et al. (2010) combine sequential search and product choice using aggregate Amazon data. Randomized hotel rankings allow Ursu (2018) to separate position effects on search from effects on purchase conditional on search. Chen and Yao (2017) and De los Santos and Koulayev (2017) account for consumers' sorting and filtering when evaluating search interfaces and rankings.

In these product-search settings, information helps users choose a purchase. On a question-and-answer platform, acquiring information is itself consumption: users may benefit from several UGC pages and an AI answer in one session. We model cumulative information separately from reading and previewing effort, then link expected session value to daily participation and further queries. The design of Generative AI Search therefore shapes both what users consume within a search session and how often they search.

2.2 Generative AI Search, Search Activity, and Demand for Creator Content

Evidence on Generative AI Search varies across settings. On Meituan, Zheng et al. (2025) find improved query formulation and more focused evaluation of relevant merchants and categories. Here, AI supports a purchase decision rather than supplying the product consumed. Using randomized retention or removal of Google AI Overviews, Agarwal and Sen (2026) find fewer unpaid external-result clicks and more searches without outbound clicks, but no detectable change in overall search frequency or destination engagement

conditional on clicking. Wang et al. (2026) compare Google's existing interface with AI removal and an AI Mode-only interface, providing further evidence on publisher referrals.

On a Brazilian short-video platform, Bai et al. (2026) study query assistance, AI summaries, and conversational follow-up. They find greater search participation, no detectable change in weekly query counts, and fewer videos shown per query without detectable changes in post-query viewing time or aggregate consumption. These findings highlight that participation, frequency, and consumption can respond differently. This distinction relates to earlier research showing that information aggregation can support visits to underlying sources, as documented for Google News in Spain and Germany (Calzada and Gil, 2020; Athey et al., 2021).

A contemporaneous study by Lee et al. (2026) examines Generative AI Search at The Washington Post. Considering AI answers and opened articles together, they find less concentrated topic consumption and greater overlap across readers. They also document increased search frequency alongside lower article consumption per search, with a modest increase in article consumption per reader.

Whereas that study examines search within a publisher's own news archive, we study a Q&A platform that connects users with content contributed by individual creators. In both settings, AI answers and underlying content can address the same informational need. Our question concerns how the platform's own information provision affects demand for the creator content it intermediates. We trace recorded UGC previews, page openings, and further navigation, and distinguish per-session demand from aggregate demand. The click-through rate measures opening among recorded previews, while the preview–opening funnel distinguishes searches with no preview, previews but no opening, or an opening. Comparisons across original ranks and search intents characterize differences in browsing, clicking, and CTR. This evidence is consistent with information substitution.

Our structural model links information acquisition within sessions to expected search value, which affects subsequent queries and daily search participation. This connection allows us to evaluate how AI placement, informativeness, allocation, and disclosure reshape search activity and demand for existing creator content.

2.3 Platform-Owned Content and the Allocation of Attention

Platforms connect creators and users through cross-side participation incentives (Rochet and Tirole, 2003; Parker and Van Alstyne, 2005; Armstrong, 2006). Their own offerings can change both demand and contributions. Amazon's entry is associated with higher product demand but discourages affected sellers from pursuing growth (Zhu and Liu, 2018); Google Photos' entry increases innovation among photography apps, consistent with greater attention to related apps (Foerderer et al., 2018). Amaldoss and Shin (2026) show theoretically that platform-owned content and UGC can be strategic complements: platform content may attract users and increase UGC provision even when both sources serve similar functions.

We examine demand for existing creator content when the platform supplies AI information and controls its placement and allocation. A fixed-budget comparison examines whether reallocating existing availability

across search needs can improve both search activity and demand for UGC per session. We also draw on information design (Kamenica and Gentzkow, 2011) to study truthful labels about an answer's information amount. Holding content fixed distinguishes the value of helping users assess an answer from the value of changing the answer itself.

3 Setting, Field Experiment, and Measurement

3.1 Search on a Content Creation Platform

Our setting is Zhihu, a leading Chinese content creation platform organized around questions and answers. Users can post questions on the platform, and a single question can receive multiple answers. Search connects users with this accumulated content. A submitted query returns a ranked list of user-generated content (UGC) previews. According to the platform, this ranking reflects its assessments of relevance to the query and content quality. Each preview displays part of a question and answer, along with selected engagement indicators. Clicking a preview opens the full content page. Users can then visit other content by the same creator, or content by another creator. Earlier studies examine participation and incentives on this platform (Wang et al., 2022; Liu and Cong, 2023; Cao et al., 2024).

Zhihu introduced the Generative AI Search feature on January 21, 2025. For suitable queries, it adds an AI-generated answer preview among the ranked UGC previews. When available, the AI preview appears first. Clicking the AI-answer preview opens the full AI answer. Both AI answers and UGC therefore first appear as previews. Viewing a preview and opening the full content are separate actions. Figure 3 shows the two interfaces used in the experiment.

The platform uses an eligibility classifier to decide which queries receive an AI answer. A query qualifies when it seeks either objective information or experience-based information for which a sufficiently standardized response can be generated. According to the platform, the classifier is trained on manually labeled queries and operates in real time.

This setting makes both search activity and demand for creator content economically relevant. Users search on the platform to obtain information that creators have provided. Their attention supports platform revenues and continued creator participation. Our analysis follows the demand side of this relationship, from search activity to interactions with creator content.

3.2 Randomized Retention and Removal

The experiment ran from March 30 through April 24, 2025. The platform selected 100,000 accounts registered by August 1, 2024, without conditioning on prior activity or search behavior. All selected accounts had access to the feature before randomization. The platform randomly assigned 50,000 accounts to retain access and 50,000 to have access removed. Assignment remained fixed at the account level throughout the experiment.



Figure 3: The Experimental Search Interfaces

Notes: The left panel shows an eligible treatment-group query. An AI-generated answer preview appears above the conventional UGC previews. The right panel shows the conventional interface, which is displayed to control-group users and to treatment-group users whose queries are ineligible. Clicking either type of preview opens the corresponding full content.

Treatment-group users retained access subject to the platform's query-eligibility rule. Control-group users received only conventional UGC results. The search logs record AI-preview exposure in 17.92% of treatment-group search sessions. Actual exposure and AI-answer openings depend on the queries users submit and their subsequent choices. The analysis compares assigned access to Generative AI Search at the user level in the full randomized sample.

The feature had been available for more than two months before randomization. This removal design evaluates access to an established feature separately from its initial launch (Lei et al., 2025; Chan et al., 2024). We examine possible adjustment to removal by comparing the two halves of the experiment (Appendix D).

3.3 Linked Data and the Units of Observation

A *search session* starts with one submitted query. It includes actions on the search results page and subsequent navigation through content pages. It ends when the user stops that search sequence or submits another query. A new query always starts a new session.

We link random assignment records with search logs and content-navigation records. The assignment records identify each user's group and define the full randomized sample. Search logs identify the user, date,

Table 1: Data Sources, Observation Levels, and Uses

Data and period	Observation level	Use in the analysis
Random assignment records	User account	Identify treatment assignment and the full randomized sample, including those who do not search during the experiment.
Pre-treatment activity: March 15–29	User account	Measure search activity, UGC preview views, and page openings before randomization to assess baseline balance between the two groups.
Experimental activity: March 30–April 24 (26 days)	Search session, preview view, and content opening	Measure search activity, views of UGC previews, UGC page openings, and navigation to additional creator content. These records are aggregated to construct user-level outcomes.
Inputs for structural estimation: March 30–April 14 (16 days)	User-day, search session, result preview, and content item	Link daily search participation and query sequences to users' preview and opening choices and the characteristics of the corresponding content.

Notes: All dates are in 2025. A user refers to a randomized account. The experimental analysis covers the full 26-day period. Structural estimation uses linked behavioral and content records from the first 16 days of the same experiment.

query text, and session. They also record which previews are viewed and which full content pages are opened. Content-navigation records link an initial UGC opening from search to subsequent UGC openings within the same session. Together, these records trace query submission, preview viewing, content opening, and continued navigation.

The experimental analyses include all 100,000 randomized accounts: 50,000 treatment and 50,000 control.

Pre-treatment account records provide age group, gender, identity verification, and creator status. Identity verification refers to verification of the user's real-world identity. Pre-experiment creator status indicates whether the user had published at least one content item before randomization. Account age is the number of days since registration, measured on March 29, 2025. Missing demographic values are included as separate categories when these characteristics are used as controls. Search, UGC preview-view, and page-opening histories from March 15–29 provide the pre-treatment behavioral measures.

Table 1 summarizes the observation levels and analysis windows. The experimental analysis uses all 26 days. Structural estimation uses the first 16 days, linking search paths to content characteristics and result positions. These inputs include AI-answer text and references, UGC length, and result positions. The linked records serve different purposes: assignment defines the experimental comparison, session and navigation logs measure behavior, and content records characterize the choices in the structural model. The shorter structural window is part of the same experiment, not a separate treatment. Section 6.1 describes the structural sample and the corresponding 16-day comparisons.

3.4 Measures of Search Activity and Demand for Creator Content

We examine demand for creator content using logged UGC preview views and page openings, reported separately. Per-session demand summarizes these behaviors within observed search sessions; aggregate

demand summarizes their totals over the observation period. Our measures cover interactions on the search-results page and subsequent navigation within the same search session. Table 2 defines the outcomes. Counts and totals include all randomized users, with zeros for users who do not search. Per-session outcomes are defined only for users with at least one search session.

We distinguish direct and downstream interactions by how users reach UGC. Direct interactions include views of UGC previews on the search-results page and openings of the full pages from those previews. Downstream UGC openings occur when users navigate from an initially opened UGC page to additional creator content within the same search session. Users may reach this additional content through creator profile pages or other UGC pages.

We measure demand for this downstream UGC in two ways: the number of sessions with at least one downstream UGC opening and the number of downstream UGC openings. A session with several such openings contributes once to the first measure and contributes its recorded downstream opening count to the second. Both measures exclude the initial direct UGC opening.

Table 2 also distinguishes distinct pages from opening events. A UGC preview or page contributes once to a session's distinct count, even if viewed or opened repeatedly; it can contribute again in another session. The separate opening-event measure retains repeat openings. For per-session outcomes, we first divide each searching user's count by that user's sessions and then average across users. These outcomes therefore differ from a ratio of pooled totals and from unconditional counts and totals for all randomized users.

The ranking-specific measures use the same UGC definition as the main interaction measures and the preceding-opening sequence decomposition. Top5 contains original result positions 0–4, restored before AI insertion, and Non-top5 contains later positions. Overall UGC outcomes also retain records whose position is unavailable. For CTR, we first compute the fraction of distinct previewed UGC pages opened within a search, then average eligible searches within each user and weight users equally. Eligibility requires a recorded preview in the relevant range. The preview–opening funnel instead includes all observed searches. Appendix E details the construction, coverage, and inference.

We measure topical coverage using the platform's 34 broad query categories and a common set of 300 semantic topics. The semantic topics group queries by similarities in meaning and capture finer differences in subject matter. Coverage measures how many subjects a user searches; submitting more queries can itself increase this count. We also measure how evenly a user's queries are distributed across the semantic topics represented in those queries. Normalized topic entropy equals zero when only one topic is represented and is undefined when there is no valid topic assignment, including when the user does not search. Appendix A provides the construction details.

Table 2: Observed Behaviors and Outcome Construction

Outcome	Construction and interpretation
Search sessions	Number of queries submitted by a user during the observation period. Each query starts a new search session.
Distinct queries	Number of distinct query texts submitted by a user over the observation period.
Active search days	Number of days on which a user submits at least one query.
Topical coverage	Numbers of distinct platform categories and semantic topics represented in a user's queries.
UGC previews viewed	Number of distinct UGC previews viewed on the results page within a session, whether or not the full content is opened.
Distinct direct UGC pages opened	Number of distinct UGC pages opened directly from result previews within a session.
Direct UGC page-opening events	Number of direct UGC openings from the results page, including repeated openings within a session.
Sessions with downstream UGC openings	Number of search sessions containing at least one downstream UGC opening after the initial direct opening.
Downstream UGC openings	Recorded UGC openings reached through subsequent navigation within the same search session, excluding the initial direct opening.

Notes: Total direct UGC measures sum the corresponding session-level counts over the observation period. For UGC previews viewed and distinct direct UGC pages opened, distinctness is defined within each session. The same preview or page can contribute again in a later session. Distinct queries are instead counted over the user's full observation period. User-level per-session outcomes divide each user's total by that user's session count. All UGC measures exclude AI answers. Downstream measures include only activity within the originating search session.

3.5 Query Intent and Independent Validation

The platform's eligibility classifier decides whether to provide an AI answer. Our separate research classifier identifies what a query asks for. It uses query text to assign one of six focal intents or a seventh, ambiguous/other category. Queries with missing or unassigned labels are included in this seventh category. The same seven categories are used in the structural model and its intent-specific policy comparisons.

The focal categories distinguish facts and explanations; procedures and how-to requests; evaluations and recommendations; experiences and opinions; open-ended exploration; and current, local, or navigational information. For example, "What is marginal cost?" seeks a factual explanation, whereas "Which phone should I buy?" seeks a recommendation. The ambiguous category covers queries that do not support a reliable primary classification. Appendix B provides the full taxonomy and rules for resolving overlaps between categories. The classifier uses only query text, without information on treatment assignment, AI eligibility, user characteristics, or observed behavior.

Three human annotators, all Zhihu users, independently classified a random sample of 1,000 distinct queries. They did not see the research labels or one another's responses. At least two annotators agreed on 967 queries; the remaining 33 received three different labels. Table 3 summarizes the validation results. Human-majority labels agree with the research labels in 77.6% of the 967 cases. Agreement rises to 86.8% among the 658 queries on which all three annotators agree. Agreement is lowest for exploratory queries. Only eight of the 967 queries carry this label under the research classification. In the summary table, Cohen's

Table 3: Validation of the Research Intent Labels

Comparison	Estimate	95% bootstrap interval
Agreement among all three human annotators	65.8%	–
Human-majority agreement with research labels	77.6%	[74.9%, 80.4%]
Cohen's kappa: human majority versus research labels	.710	[.675, .746]
Macro F1 across the seven categories	.684	[.650, .720]
Agreement with research labels in the human-unanimous subsample	86.8%	–

Notes: All-three agreement uses the full sample of 1,000 independently annotated queries. Human-majority agreement, Cohen's kappa, and macro F1 use the 967 queries with a majority label. The unanimous-subsample comparison uses 658 queries. Intervals use 1,000 bootstrap resamples; dashes indicate unreported intervals. Kappa adjusts agreement for chance. Macro F1 averages the category-specific F1 scores equally across the seven categories; F1 is the harmonic mean of precision and recall. The research labels serve as the reference for computing classification metrics. The exercise assesses agreement between independent labelings. Human annotators observe query text but no treatment assignment, AI eligibility, or behavioral records.

kappa adjusts agreement for chance, while macro F1 gives each intent equal weight rather than allowing common intents to dominate the metric. These statistics describe agreement with the research labels, not accuracy against an established ground truth. The bootstrap intervals quantify uncertainty in this validation sample. Full class-specific statistics appear in Appendix Table A4.

3.6 Sample Characteristics and Pre-Treatment Balance

The average randomized user submits 4.88 queries and searches on 1.89 days during the 26-day experiment. Among the 52,711 users who search, the average user views 9.72 UGC previews and opens 1.56 distinct UGC pages directly from search per session. Across all randomized users, total UGC preview views average 48.92 and total distinct direct UGC openings average 8.18 per user during the experimental period. Table 4 reports the full summary statistics. Search activity is unevenly distributed: the median user submits one query, and the standard deviation is 15.09.

The means for distinct direct UGC openings and direct page-opening events are close, indicating that repeated openings add little to the total direct-opening count. Downstream UGC openings vary substantially across users: their standard deviation is 129.57, compared with a mean of 22.60. We therefore report both the number of sessions with downstream UGC openings and the number of downstream UGC openings.

Table 5 presents the complete pre-treatment balance checks. These comparisons cover account characteristics, missing demographic information, search activity, and direct and downstream UGC interactions. The treatment and control groups have similar pre-treatment characteristics and behavior. None of the reported differences is statistically significant at the 5% level. Standardized mean differences (SMDs) also indicate small imbalances: the largest absolute SMD among the reported measures is 0.0162. The behavioral comparisons use the pre-experiment period, not outcomes potentially affected by treatment. Counts retain users with no search, whereas the per-session comparisons in Panel D require a positive pre-treatment session count. The difference column is treatment minus control; the reported tests assess baseline comparability rather than establishing that the two populations are identical.

Table 4: Summary Statistics for the Randomized Sample

Outcome	Mean	SD	Median
A. Pre-treatment characteristics			
Age group (0–5; observed accounts)	2.1883	1.4617	2
Account age in days	1702.7374	1110.9950	1548
Male (observed accounts)	.3984	.4896	0
Identity verified	.0030	.0543	0
Pre-experiment creator	.3823	.4859	0
B. Search scale and coverage over 26 days			
Search sessions	4.88	15.09	1
Distinct queries	4.12	12.28	1
Active search days	1.89	3.39	1
Unique platform categories	1.55	2.61	1
Unique semantic topics	3.04	6.87	1
Normalized topic entropy	.6742	.4337	.9409
C. Aggregate UGC demand over 26 days			
UGC previews viewed	48.9175	177.5008	4
Distinct direct UGC pages opened	8.1809	31.7909	0
Direct UGC page-opening events	8.2670	32.0459	0
Sessions with downstream UGC openings	1.9650	7.4072	0
Downstream UGC openings	22.5998	129.5716	0
D. Direct UGC demand per session, among searchers			
UGC previews viewed	9.7160	8.3428	7.2000
Distinct direct UGC pages opened	1.5583	1.5991	1.0000
Direct UGC page-opening events	1.5768	1.6164	1.0000

Notes: Behavioral counts and totals use all 100,000 randomized accounts, including zeros. Account characteristics describe the same randomized cohort. Entropy and Panel D use the 52,711 users who search. Age and gender are observed for 89,929 and 82,325 accounts, respectively. SD is the cross-user standard deviation. Panel D averages each user's own per-session ratio. Behavioral measures cover March 30–April 24, 2025; account age is measured on March 29. Distinct direct UGC pages are counted within each session before summation over the experiment.

Table 5: Pre-Treatment Balance

Measure	Treatment	Control	Difference	p-value	SMD
A. Account characteristics					
Age group	2.1934	2.1833	.0101	.3001	.0069
Age group missing	.1023	.0992	.0031	.1034	.0103
Account age in days	1700.1138	1705.3611	-5.2473	.4552	-.0047
Male	.3991	.3978	.0013	.7068	.0026
Gender missing	.1790	.1745	.0045	.0645	.0117
Identity verified	.0031	.0028	.0002	.4849	.0044
Pre-experiment creator	.3820	.3825	-.0005	.8759	-.0010
B. Search activity					
Any pre-treatment search	.3444	.3454	-.0010	.7344	-.0021
Search sessions	2.8591	2.8125	.0466	.4714	.0046
Distinct queries	2.4210	2.3913	.0297	.5794	.0035
Active search days	1.0626	1.0577	.0049	.7228	.0022
Sessions per active search day	2.2530	2.2500	.0030	.8810	.0016
C. Aggregate UGC demand					
Total UGC previews viewed	28.0582	27.9474	.1108	.8910	.0009
Total distinct direct UGC pages opened	4.7191	4.7125	.0066	.9609	.0003
Total direct UGC page-opening events	4.7555	4.7507	.0048	.9717	.0002
Sessions with downstream UGC openings	1.1283	1.1284	-.0001	.9965	.0000
Downstream UGC openings	12.8471	12.9551	-.1080	.8483	-.0012
D. Direct UGC demand per session, among pre-treatment searchers					
UGC previews viewed per session	9.3065	9.3078	-.0013	.9877	-.0002
Distinct direct UGC pages opened per session	1.5384	1.5634	-.0250	.1328	-.0162

Notes: Behavioral measures cover March 15–29, 2025. Each randomized group contains 50,000 accounts. Counts, totals, and activity indicators include all randomized accounts, with zeros for no activity. Per-active-day and per-session measures use users with positive denominators and average each user's own ratio. Age and gender comparisons use observed values; their missing-value indicators are compared separately. Account age is measured on March 29, 2025. Differences are treatment minus control and are calculated from unrounded means, so they may differ slightly from differences between the displayed values. Distinct direct UGC pages are counted within each session before summation. SMD is the treatment–control difference divided by $\sqrt{(s_T^2 + s_C^2)/2}$, where s_g^2 is the within-group sample variance; for binary measures it is $p_g(1 - p_g)$.

4 Experimental Evidence

We address the experimental research questions in sequence. For RQ1, we examine how often and how broadly users search, then use query sequences to characterize the additional activity. For RQ2, we examine UGC consumption within searches, connect preview-to-opening outcomes and intent differences to information substitution, and assess aggregate demand over the experiment. This order distinguishes changes within a search from the consequences of users searching more often.

4.1 Empirical Specification

We compare user-level outcomes between the randomly assigned groups. Each outcome covers the full experiment, and each account receives equal weight. We estimate ordinary least squares (OLS) regressions:

$$Y_i = \alpha + \beta \,\mathrm{Treat}_i + X_i'\gamma + \varepsilon_i. \tag{1}$$

Here Y_i is user i's outcome and Treat_i indicates assigned access. The primary specification includes assignment and an intercept; the adjusted specification adds pre-treatment age group, gender, identity verification, and creator status in X_i. Controls are categorical, with separate missing-value categories. The coefficient β is the treatment–control difference, adjusted when controls are included. We report heteroskedasticity-robust standard errors.

For counts and totals, all 100,000 randomized accounts enter the regression, and β is the user-level intent-to-treat effect of assignment to continued feature access. Per-session outcomes are defined for users who search. Normalized topic entropy additionally requires a valid topic distribution. The coefficients for these outcomes compare observed searchers across groups. CTR additionally conditions on recorded previews in the relevant result range; it is not an unconditional assignment effect for all randomized users.

Estimation in original units preserves zeros and gives a direct per-user interpretation (Chen and Roth, 2024). Appendix D reports Poisson pseudo-maximum-likelihood (PPML) estimates for counts and totals, as well as period-specific OLS estimates. These checks compare results across specifications and periods of the experiment. Tables report coefficients and standard errors for both the unadjusted and adjusted specifications. Figures show 95% confidence intervals, calculated as the estimate plus or minus 1.96 standard errors. The supplementary sequence, ranking, and intent comparisons use the unadjusted assignment contrast and two-sided, unadjusted single-test p-values. Cross-range and cross-intent contrasts retain covariance among outcomes for the same user; their confidence intervals are pointwise.

Table 6: Effects on Search Scale and Topical Coverage

Outcome	No controls	With controls	Users
Search sessions	0.3666***	0.3779***	100,000
	(0.0955)	(0.0939)	
Distinct queries	0.2744***	0.2842***	100,000
	(0.0778)	(0.0763)	
Active search days	0.0701***	0.0753***	100,000
	(0.0215)	(0.0204)	
Unique platform categories	0.0491***	0.0530***	100,000
	(0.0165)	(0.0156)	
Unique semantic topics	0.1363***	0.1443***	100,000
	(0.0435)	(0.0421)	
Normalized topic entropy	−0.0007	−0.0009	52,711
	(0.0038)	(0.0037)	

Notes: User-level OLS estimates for March 30–April 24, 2025. Counts include randomized users with zero activity. Entropy is conditional on a valid topic distribution. Controls and inference follow Equation (1). Parentheses report heteroskedasticity-robust standard errors. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

4.2 RQ1: Search Activity

We first ask whether access expands search activity and the range of information users seek. We then use query text and timing to classify additional searches as reformulations, related follow-ups, or distinct information tasks.

4.2.1 Search Expands in Scale and Coverage

Table 6 shows that assigned access to Generative AI Search increases search sessions, distinct queries, and active search days. The primary estimates are 0.3666 additional sessions, 0.2744 additional distinct queries, and 0.0701 additional active search days per user. Each is significant at the 1% level. The adjusted estimates are similar.

Treatment also increases the number of represented platform categories and semantic topics. The estimated increases are 0.0491 categories and 0.1363 topics per user. Thus, the additional activity reaches a wider set of subjects. The estimated entropy difference is close to zero and not statistically significant, indicating no detectable change in topic evenness among observed searchers.

These measures establish broader search activity, not greater learning or more intensive information acquisition within each query.

Continued search use. We next ask whether this greater use extends across the experimental period. Following the distinction between active-day frequency and the range of search dates in Song et al. (2013), our primary measure is *search-use span*: the number of days between a user's first and last search during March 30–April 24. Users with no searches or searches on only one day receive zero. Table 7 shows that access

Table 7: Effects on Continued Search Use

Outcome	No controls	With controls	Users
Search-use span (days)	0.1404***	0.1534***	100,000
	(0.0489)	(0.0463)	
Search in both halves (pp)	0.5282**	0.5969**	100,000
	(0.2648)	(0.2521)	

Notes: Each user receives equal weight. These analyses include the full sample of 100,000 users: 50,000 treatment users and 50,000 control users. Search-use span is the last search date minus the first search date, with zero for no searches or a single active day. Search in both halves equals one for at least one search in March 30–April 11 and at least one in April 12–24, 2025, and zero otherwise; its coefficients and standard errors are expressed in percentage points (pp). Users with zero activity remain in both regressions. Controls are age group, gender, identity-verification status, and pre-experiment creator status, with separate missing-value categories. Parentheses report heteroskedasticity-robust standard errors. Significance uses unadjusted two-sided tests: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

increases this span by 0.1404 days, or 3.26% relative to the control-group mean of 4.3114 days ($p = 0.0041$). The estimate with baseline controls is 0.1534 days ($p < 0.001$).

As a complementary measure, we record whether a user searches in both 13-day halves of the experiment. This share is 22.8060% in treatment and 22.2778% in control, a difference of 0.5282 percentage points ($p = 0.0460$); the adjusted difference is 0.5969 percentage points ($p = 0.0179$). Both measures include users with no search activity rather than conditioning on first-half participation. Together, they show that greater search volume is accompanied by a longer span of use and more cross-period repeat use within the experiment.

The expansion is not confined to a single kind of query. Figure 4 reports positive count effects in all six focal intents. These estimates capture the effects of assigned access to Generative AI Search on query counts in each intent category. Recorded AI-preview exposure rates vary by intent: 34.27% for procedural queries, 30.46% for factual queries, and 12.73% for exploratory queries. These are descriptive rates among treatment-group queries. Appendix B.3 reports the full estimates and AI-preview exposure rates. Intent categories differ in baseline volume, so larger count effects need not imply larger proportional responses.

4.2.2 Where the Additional Sessions Occur

We divide each user's search sessions into three categories: the first session during the experimental period, the first session on each additional active day, and later sessions on the same day. Let S_i be user i's total sessions and D_i the number of active search days. Let $A_i = \mathbf{1}\{S_i > 0\}$ equal one if the user searches at all and zero otherwise. Then

$$S_i = A_i + (D_i - A_i) + (S_i - D_i). \tag{2}$$

The terms count sessions in these categories, respectively.

Because Equation (2) holds for every account, OLS coefficients for the components estimated on the same sample with the same specification add to the total-session coefficient. Later same-day sessions account for

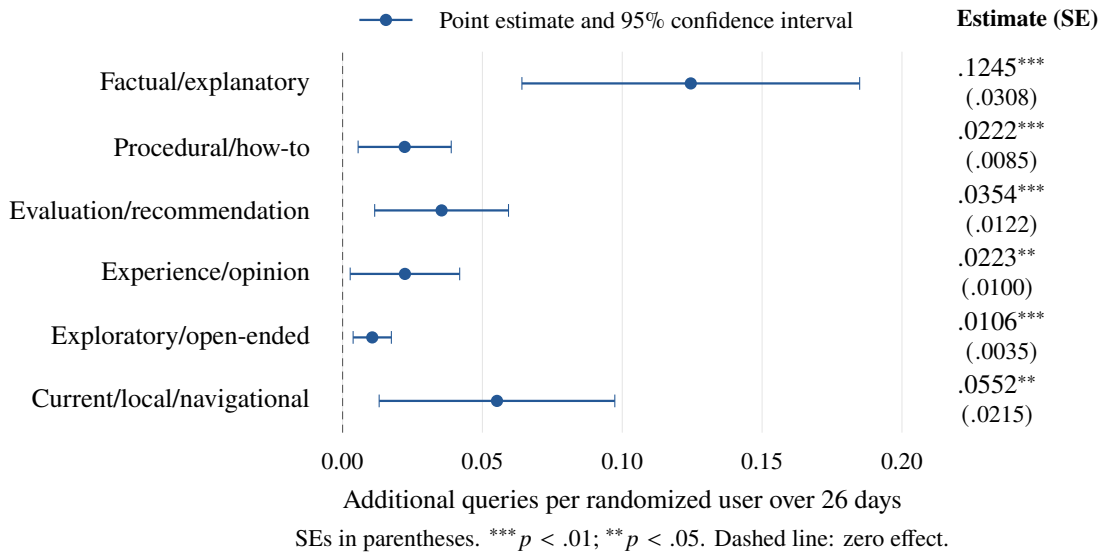


Figure 4: Search Growth Across Query Intents

Notes: Dots show the no-controls user-level OLS estimates; capped lines show 95% normal-approximation confidence intervals based on heteroskedasticity-robust standard errors, also printed in parentheses below each estimate. All 100,000 randomized users enter each regression. The classifier uses query text only and is independent of the platform's AI-eligibility classifier. Exploratory/open-ended labels have low agreement in the human-validation sample, so that estimate warrants caution.

80.9% of the estimated increase. First sessions on additional active days account for 16.9%. The remaining 2.2% comes from the difference in whether users search at all.

We next compare each later query with the immediately preceding query from the same user on the same day. Closely similar queries within 30 minutes are classified as *reformulation/repair*. Related queries within 30 minutes that do not meet that near-duplicate rule are *related follow-ups*. Queries separated by more than 30 minutes or by low semantic similarity are classified as *distinct information tasks*, including similar text after a longer gap. These labels describe how adjacent queries relate in text and timing; they do not measure task completion. Appendix C states the exact rules.

Figure 5 reports the estimated increases and their confidence intervals. Reformulation/repair accounts for 18.0% of the increase in later same-day queries, related follow-ups for 14.8%, and distinct information tasks for 67.3%. Most additional same-day queries are therefore classified as follow-ups or distinct tasks rather than reformulations.

Together, these results answer RQ1: access expands the frequency and breadth of search, with most additional activity occurring later on the same day. Most additional same-day queries are classified as related follow-ups or distinct information tasks. Together with the broader topical coverage, this pattern supports continued and exploratory information seeking.

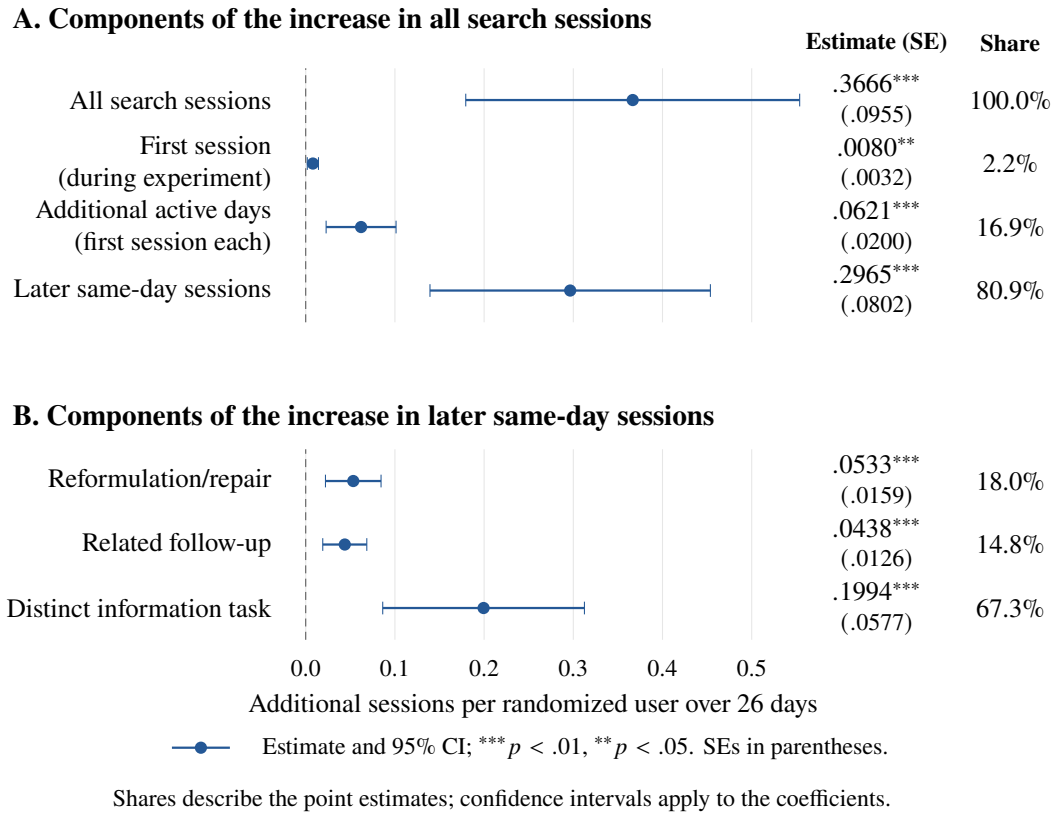


Figure 5: Most Additional Sessions Occur Later on the Same Day

Notes: Coefficients and capped 95% confidence intervals use the no-controls OLS estimates and heteroskedasticity-robust standard errors, printed in parentheses below the estimates. The dashed line is zero. The first-session estimate in Panel A is significant at 5%; the other displayed coefficients are significant at 1%. Shares divide component estimates by 0.3666 in Panel A and 0.2965 in Panel B. They describe the estimated increase, rather than the distribution of all observed queries. All rows retain the same 100,000 accounts. Rounding affects the displayed sums. Table A7 also reports adjusted estimates.

4.3 RQ2: Demand for Creator Content

We next ask how access to Generative AI Search affects demand for creator content. We begin with consumption per search, then examine preview-to-opening behavior and differences across result ranks and information needs. Finally, we combine the within-search responses with search expansion to assess total demand over the experiment.

4.3.1 Demand for Direct UGC per Search Session

Table 8 reports demand for creator content per observed search session. Among users who search, the treatment group views 0.1487 fewer UGC previews and opens 0.0447 fewer distinct direct UGC pages per session than the control group.

From UGC previews to page openings. Fewer openings per search could reflect fewer recorded previews, a lower fraction of previewed results being opened, or both. To distinguish these patterns, we use the same

Table 8: Demand for UGC per Search Session

Outcome	No controls	With controls	Users
UGC previews viewed	−0.1487**	−0.1514**	52,711
	(0.0727)	(0.0723)	
Distinct direct UGC pages opened	−0.0447***	−0.0458***	52,711
	(0.0139)	(0.0138)	

Notes: Each outcome divides a user's total by that user's number of search sessions and weights searching users equally. These comparisons are conditional on search participation. Each UGC preview and distinct direct UGC page is counted once per session. Both outcomes use the same UGC definition as the main analysis. Controls and inference follow Equation (1). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 9: Preview-to-Opening Outcomes Overall

Outcome	Control (%)	Treatment (%)	Effect (pp)
CTR	17.5287	17.0469	−0.4818***
			(0.0976)
No preview	0.0000	0.4077	0.4077***
			(0.0237)
Preview but no opening	29.6362	31.1372	1.5009***
			(0.2691)
At least one opening	70.3638	68.4551	−1.9086***
			(0.2697)

Notes: All measures use the same UGC definition as the main analysis. CTR is averaged over searches with at least one preview within each eligible user; 26,155 Control and 26,539 Treatment users enter. The three search states use all 487,940 searches by 52,711 searching users and sum to one within each search. Users receive equal weight. The states describe logged exposure and openings. Overall includes usable results whose original rank cannot be assigned. All effects are Treatment minus Control, in percentage points. Heteroskedasticity-robust standard errors are in parentheses. Significance uses unadjusted two-sided p values: $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$. Appendix E details the construction and inference.

UGC definition as the preceding count outcomes. For each search, the click-through rate (CTR) is the number of distinct opened results divided by the number of distinct previewed results, provided that at least one preview is recorded. We average this ratio within each eligible user and then weight users equally.

Table 9 shows that overall CTR falls by 0.482 percentage points. A complementary decomposition separates three mutually exclusive states: no preview, a preview but no opening, and at least one opening. Unlike CTR, these states use all observed searches as the denominator before averaging within users. Treatment increases the share with a preview but no opening by 1.501 percentage points and the share with no preview by 0.408 percentage points. The share with at least one opening falls by 1.909 percentage points.

Users are less likely to open UGC even after seeing its preview. For the average user, a larger share of searches also includes previews but no opening. Lower UGC consumption therefore involves changes after recorded preview exposure, not only fewer previews. We next examine where these responses occur and how they vary with the information users seek.

Table 10: Browsing, Clicking, and Click-Through by Original UGC Rank

Outcome	Top 5	Non-top 5	Top 5 − Non-top 5
Browsing: previews per search	-0.0349^{***}	-0.1176	0.0827
	(0.0064)	(0.0715)	(0.0706)
Clicking: openings per search	-0.0338^{***}	-0.0110	-0.0228^{**}
	(0.0058)	(0.0107)	(0.0102)
CTR (percentage points)	-0.5798^{***}	-0.0445	-0.5352^{***}
	(0.1307)	(0.1032)	(0.1434)

Notes: The sample contains 487,940 searches by 52,711 searching users. All measures use the same UGC definition as the main analysis. Browsing and clicking include zero outcomes in every user's observed searches. CTR is averaged only over searches with a preview in the relevant range; eligible Control/Treatment users number 26,084/26,467 (Top 5) and 20,051/20,128 (Non-top 5). Users receive equal weight within each outcome. The last column directly compares the two treatment effects, preserving within-user cross-range dependence; eligible users can differ across CTR cells. All effects are Treatment minus Control. Standard errors are in parentheses. Significance uses unadjusted two-sided p values: $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$.

4.3.2 Differences across Original Result Ranks

Browsing, clicking, and click-through by original rank. We next examine whether the change in demand differs across result ranks. Top 5 refers to original positions 0–4, before AI insertion; Non-top 5 includes all original positions indexed 5 or higher. The ranking reflects the platform's assessments of query relevance and content quality. The comparison therefore distinguishes the results the platform prioritizes from later-ranked alternatives, rather than simply dividing the page into two arbitrary regions. The same UGC definition and user weighting apply to both ranges. Browsing counts distinct recorded previews, clicking counts distinct opened results, and CTR is the opened-to-previewed fraction when at least one result in the range is previewed. Appendix E describes rank reconstruction and records with unavailable positions.

Table 10 shows significant reductions in Top 5 previews and openings per search. The Non-top 5 estimates for these counts are not statistically resolved from zero. The distinction is especially clear for CTR: it falls by 0.580 percentage points in the Top 5, whereas the Non-top 5 estimate is −0.045 percentage points ($p = .666$). A direct comparison confirms that the CTR decline is larger in the Top 5: the Top 5-minus-Non-top 5 treatment-effect difference is −0.535 percentage points (95% CI $[-0.816, -0.254]$, $p < .001$). The opening-count decline also differs significantly between the two ranges ($p = .025$), but the preview-count difference does not ($p = .241$). These comparisons use the covariance between user-level outcomes. The sharper Top 5 CTR decline shows that reduced opening after a preview also affects the results prioritized for relevance and quality. This pattern is consistent with AI substituting for information in highly ranked UGC. Rank captures the platform's combined prioritization through relevance, quality, and placement.

The preview–opening funnel by rank. Table 11 uses the three mutually exclusive search states to locate these changes within each rank range. Treatment increases the share of searches with a Top 5 preview but no Top 5 opening by 1.297 percentage points, while reducing the share with a Top 5 opening by 1.844 percentage

Table 11: Search States by Original UGC Rank

State (percentage points)	Top 5	Non-top 5	Top 5 − Non-top 5
No preview	0.5474^{***}	1.0176^{***}	-0.4701
	(0.0756)	(0.3175)	(0.3296)
Preview but no opening	1.2969^{***}	-0.3970	1.6939^{***}
	(0.2733)	(0.2586)	(0.3693)
At least one opening	-1.8444^{***}	-0.6206^{**}	-1.2238^{***}
	(0.2775)	(0.2519)	(0.3390)

Notes: For each search and rank range, the states are $\mathbf{1}\{P = 0\}$, $\mathbf{1}\{P > 0, C = 0\}$, and $\mathbf{1}\{C > 0\}$, where P and C count distinct previewed and opened UGC results. They sum to one. Every observed search enters each state's denominator before averaging within each of the 52,711 searching users. The last column directly compares the treatment effects using their within-user covariance. All effects are Treatment minus Control. Standard errors are in parentheses. Significance uses unadjusted two-sided p values: $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$.

points. For Non-top 5 results, the no-preview share increases and the any-opening share decreases, but the change in preview-without-opening is not significant.

The difference across ranges is significant for preview-without-opening (1.694 percentage points, $p < .001$) and any opening (-1.224 percentage points, $p < .001$), but not for no preview ($p = .154$). The evidence for cross-range differences is therefore clearest in opening outcomes and in the share with recorded previews but no opening. This pattern further supports the information-substitution interpretation.

4.3.3 Differences across Search Intents

Information substitution predicts that the response in UGC opening depends on the information a query seeks. The intent taxonomy distinguishes requests for explanations, instructions, and decision support from requests for personal experiences or open-ended exploration (Rose and Levinson, 2004; Morris et al., 2010; Marchionini, 2006). An AI answer can directly supply an explanation, a set of steps, or a comparison, reducing the need to open UGC for that information. Users seeking first-hand experiences or a range of perspectives may instead retain a reason to consult individual creators. This distinction motivates examining how opening responses vary with the content of users' information needs.

Table 12 compares browsing, clicking, and CTR across the six search-intent categories. All three measures decline significantly for factual/explanatory, procedural/how-to, and current/local/navigational queries. Evaluation/recommendation queries show a distinct pattern: browsing has no statistically significant change, while both clicking and CTR decline. Thus, this category shows a change in opening behavior after recorded preview exposure alongside no significant decline in preview counts. The experience/opinion and exploratory/open-ended estimates are not statistically significant for any of the three measures.

Formal comparisons establish that the opening responses differ across intents. Joint tests reject equality of the six treatment effects for clicking ($p = .033$) and CTR ($p < .001$), but not for browsing ($p = .176$). The factual/explanatory treatment effect is more negative than the experience/opinion effect by 0.1121 openings

Table 12: Browsing, Clicking, and Click-Through by Search Intent

Intent	Browsing	Clicking	CTR (pp)
Factual/explanatory	−0.3819***	−0.1029***	−0.889***
	(0.1231)	(0.0242)	(0.168)
Procedural/how-to	−0.4585**	−0.1399***	−1.105***
	(0.2034)	(0.0394)	(0.252)
Evaluation/recommendation	0.0178	−0.1148***	−1.150***
	(0.1888)	(0.0402)	(0.238)
Experience/opinion	−0.0114	0.0092	−0.123
	(0.1860)	(0.0418)	(0.253)
Exploratory/open-ended	0.1594	0.0541	0.630
	(0.3997)	(0.0783)	(0.485)
Current/local/navigational	−0.2946***	−0.0791***	−0.351**
	(0.1021)	(0.0194)	(0.163)
Equality across intents (p)	.176	.033	< .001

Notes: All outcomes use the same UGC definition as the main analysis. Browsing and clicking count distinct UGC previews viewed and pages opened per search. They average all observed searches within each user and intent, including zero counts. CTR averages session opening-to-preview ratios over searches with at least one recorded UGC preview within each user and intent, and is reported in percentage points. Users receive equal weight within each cell and can enter several categories. The last row tests equality of the six treatment effects for each outcome. Appendix E.1 provides sample counts and inference details. All effects are Treatment minus Control. User-clustered standard errors are in parentheses. Significance uses unadjusted two-sided p values: $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$.

per search ($p = .014$) and 0.767 percentage points of CTR (95% CI $[-1.345, -0.188]$, $p = .009$). The larger decline for factual/explanatory than experience/opinion queries, together with lower clicking and CTR for procedural and evaluative queries, aligns with AI answers meeting information needs that would otherwise lead users to open UGC. Combined with the overall CTR decline and the increase in searches with previews but no opening, these intent differences support information substitution as a behavioral channel. The evidence links lower opening after preview exposure to the type of information users seek. Appendix E.1 reports the direct comparisons and the full rank-by-intent CTR breakdown. The latter also shows a positive Non-top 5 CTR effect for experience/opinion queries ($p = .038$), highlighting variation across both information types and result ranges.

4.3.4 Aggregate Demand for Direct and Downstream UGC

Aggregate demand for UGC moves in the opposite direction from demand per search session. Over the full experiment, access increases the number of UGC previews viewed by 3.3168 and the number of distinct direct UGC pages opened by 0.4212 per randomized user (Table 13).

Access also increases the number of sessions with downstream UGC openings by 0.1224 ($p < 0.01$). Users therefore have more opportunities to navigate from an initial UGC page to additional creator content within the same search session. Total downstream UGC openings increase by 1.4895 per user (standard error: 0.8208; $p < 0.10$), a suggestive rather than a 5%-significant increase.

Table 13: Aggregate Demand for Direct and Downstream UGC

Outcome	No controls	With controls	Users
UGC previews viewed	3.3168***	3.4059***	100,000
	(1.1240)	(1.1073)	
Distinct direct UGC pages opened	0.4212**	0.4346**	100,000
	(0.2014)	(0.1984)	
Direct UGC page-opening events	0.4240**	0.4377**	100,000
	(0.2030)	(0.2000)	
Sessions with downstream UGC openings	0.1224***	0.1249***	100,000
	(0.0469)	(0.0460)	
Downstream UGC openings	1.4895*	1.5196*	100,000
	(0.8208)	(0.8132)	

Notes: Outcomes are totals over the 26-day experiment. The table reports assignment contrasts for all randomized users, including zeros for nonsearchers. Each UGC preview and each distinct direct UGC page is counted once per session before counts are summed across sessions. Direct UGC page-opening events include repeated openings. Downstream UGC openings follow navigation from an initially opened UGC page and remain within the same search session. Controls and inference follow Equation (1). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Aggregate demand by original rank. The aggregate pattern extends beyond the first five results. Table 14 reports increases of 1.4693 and 1.8267 UGC previews per user in the Top 5 and Non-top 5, respectively. Relative to their respective control means, both increases are about 7.0%. The estimated increases in distinct direct UGC openings are nearly identical in absolute terms: 0.2126 and 0.2072 per user. The Top 5 opening effect is significant at the 5% level, whereas the Non-top 5 estimate is marginally significant at the 10% level. Direct comparisons do not detect differences between the two ranges in total previews ($p = .477$) or distinct openings ($p = .941$).

These results distinguish the aggregate consequences for results that the platform prioritizes on relevance and quality from those for later-ranked alternatives. The sharper reduction in Top 5 CTR within searches does not translate into a decline in total openings of Top 5 content. At the same time, the positive, less precisely estimated Non-top 5 opening effect suggests that additional demand is not confined to the highest-ranked results. Similar aggregate opening point estimates therefore coexist with different within-search responses across the ranking.

Table 14: Aggregate UGC Demand by Original Rank

Outcome	Top 5	Non-top 5	Top 5 − Non-top 5
Total UGC previews viewed	1.4693***	1.8267**	−0.3574
	(0.4259)	(0.7561)	(0.5023)
Total distinct direct UGC pages opened	0.2126**	0.2072*	0.0054
	(0.1058)	(0.1081)	(0.0726)

Notes: Outcomes sum each user's session-level counts over the 26-day experiment. Users with no relevant activity receive zero; users are weighted equally. A result is counted once per session before counts are summed across sessions. Top 5 comprises original positions 0–4 before AI insertion, and Non-top 5 comprises all original positions 5 or higher. The UGC definition follows the main analysis. The first two columns report Treatment-minus-Control estimates without additional controls; the last column compares these effects using their within-user covariance. Standard errors are in parentheses. Stars use unadjusted two-sided p values: $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$.

4.3.5 Connecting Search Scale to Aggregate Demand for UGC

Aggregate demand for UGC depends on both the number of search sessions and demand per session. An accounting decomposition shows how the group differences in these two quantities combine. For group $z \in \{C, T\}$, where C denotes control and T treatment, let S_z be the mean number of sessions per randomized user. Let M_z be the group's total for the UGC outcome being decomposed divided by its total search sessions. Thus, M_z weights every session equally, whereas Table 8 weights each searching user equally. The mean total for that outcome per randomized user is $\overline{Y}_z = S_z M_z$. The treatment–control difference can therefore be written as

$$\overline{Y}_T - \overline{Y}_C = \underbrace{(S_T - S_C)\frac{M_T + M_C}{2}}_{\text{search-activity component}} + \underbrace{(M_T - M_C)\frac{S_T + S_C}{2}}_{\text{per-session component}}. \tag{3}$$

Each component values the change in one margin at the two groups' average of the other. Both are in units of the corresponding UGC outcome per randomized user and sum exactly to the total difference. They are accounting components, not separately identified causal effects. Plug-in estimates use observed group means and pooled ratios; delta-method standard errors retain user-level covariance between session counts and the corresponding UGC outcome (Appendix C.4).

Figure 6 illustrates the decomposition for distinct direct UGC pages opened. The positive search-activity component, 0.6148, exceeds the negative per-session component, −0.1936, leaving a total effect of 0.4212. Unlike the user-level comparison in Table 8, this component uses pooled distinct direct UGC openings per session and expresses its contribution in total openings per randomized user. The other direct UGC measures follow the same directional pattern (Appendix C.4). The decomposition shows why lower demand for UGC

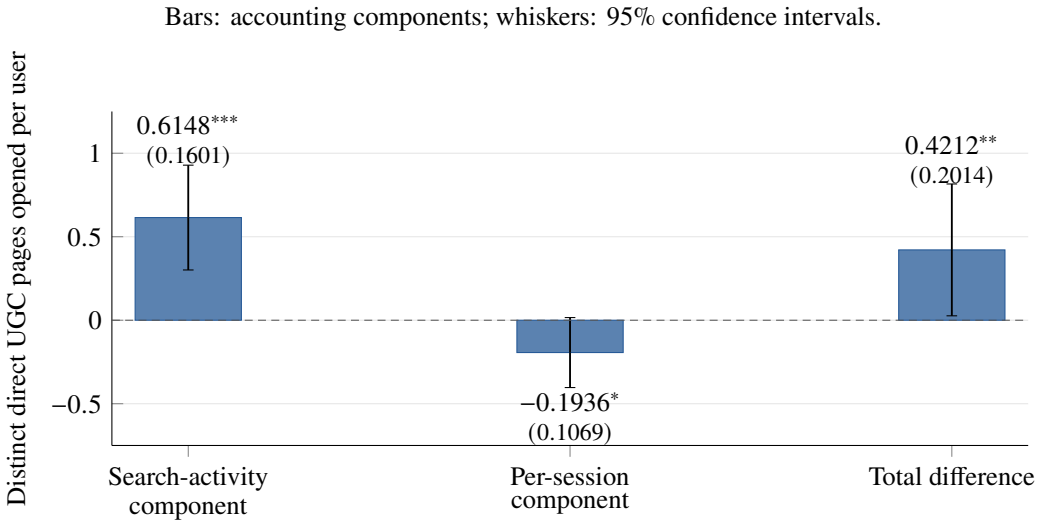


Figure 6: Search Expansion More Than Offsets Lower Demand for UGC per Session

Notes: Exact accounting decomposition of the treatment–control difference in total distinct direct UGC pages opened per randomized user over 26 days. The first two point estimates sum to the third. Whiskers use the user-level delta-method standard errors printed below the estimates. The per-session component uses pooled distinct direct UGC openings per session, rather than the mean of users' individual ratios. The components rearrange observed group means and are not separately identified causal effects. Appendix C.4 reports the remaining outcomes and their standard errors.

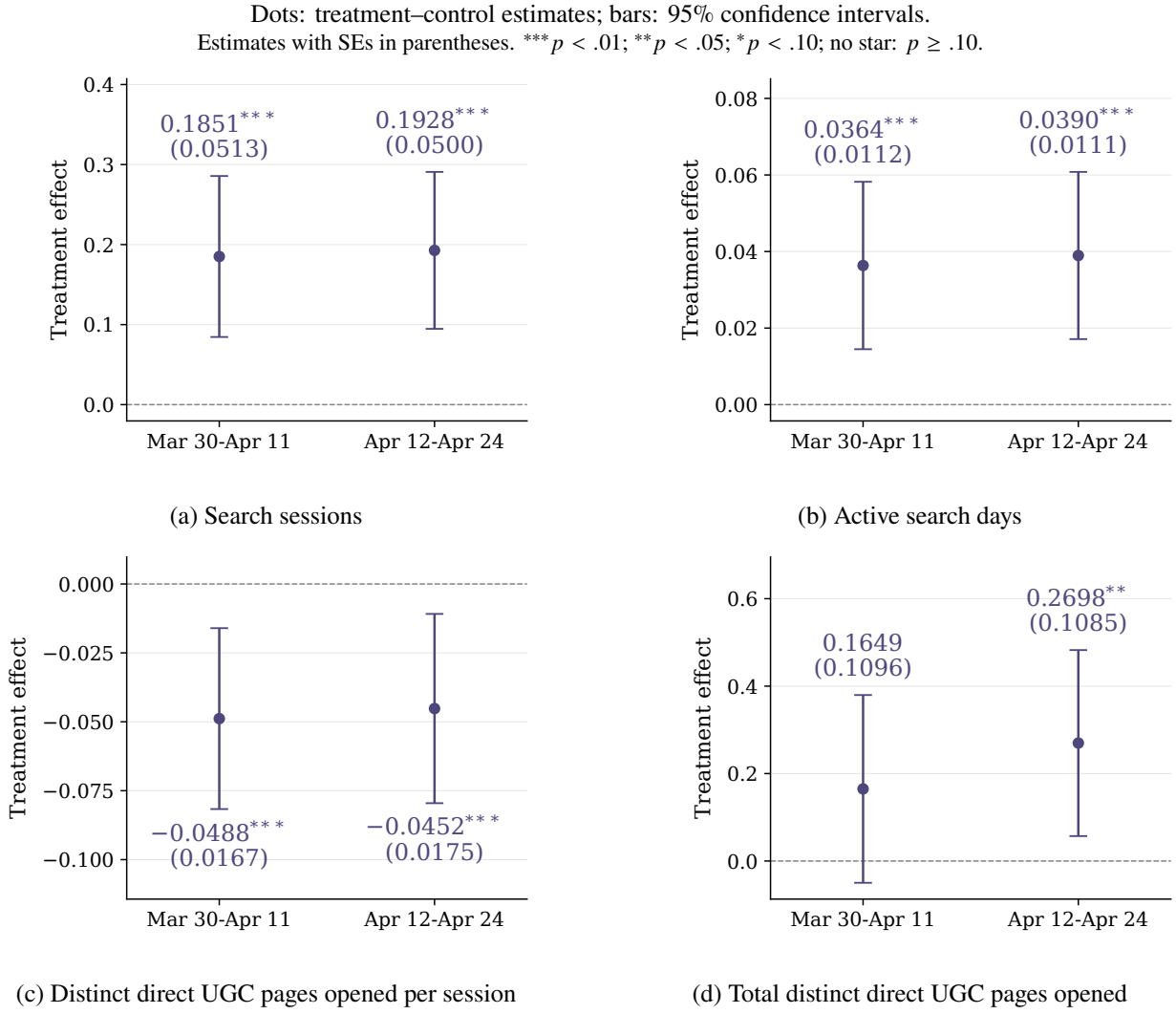
per search session can coexist with higher aggregate demand. The later structural model will examine how users' choices connect these two margins and how alternative Generative AI Search designs change them.

4.4 Robustness of the Experimental Pattern

We examine whether the findings depend on the regression specification, the period of the experiment, or the rules used to classify adjacent queries, and supplement them with a difference-in-differences check.

First, PPML expresses changes in counts and totals in proportional terms while retaining users with zero activity. The unadjusted estimates imply 7.8% more search sessions, 6.9% more distinct queries, and 3.8% more active search days. Each is significant at the 1% level. The total number of UGC previews viewed increases by about 7.0% ($p < 0.01$), and distinct direct UGC pages opened increase by 5.3% ($p < 0.05$). Sessions with downstream UGC openings also increase ($p < 0.01$), as do total downstream UGC openings ($p < 0.10$). These estimates express the same aggregate-demand pattern in proportional terms. Table A11 in Appendix D reports the estimated coefficients and their standard errors.

Second, we reconstruct outcomes separately for March 30–April 11 and April 12–24, two 13-day periods. Figure 7 shows the four core search and direct UGC opening outcomes with their 95% confidence intervals. Both halves show the same directional pattern: more search sessions and active search days, fewer distinct direct UGC pages opened per observed session, and more total distinct direct UGC pages opened. The results are broadly consistent across the two periods; see Appendix D for details. We highlight four key outcomes in Figure 7 because they jointly capture the scale–intensity trade-off central to our research questions. Panels A



(a) Search sessions

(b) Active search days

(c) Distinct direct UGC pages opened per session

(d) Total distinct direct UGC pages opened

Figure 7: Core Experimental Outcomes in Each 13-Day Period

Notes: Period-specific OLS includes the four baseline controls in Equation (1), including categories for missing demographics. Intervals use user-clustered standard errors, also printed beneath each estimate. Stars use the original regression p-values. Effects on search counts and total UGC openings are per randomized user over the indicated 13 days and retain nonsearchers as zeros. Panel C uses each searching user's per-session ratio and weights these users equally. The periods are March 30–April 11 and April 12–24, 2025. Outcomes are rebuilt within each period.

and B show search activity, Panel C shows distinct direct UGC pages opened per session, and Panel D shows total distinct direct UGC pages opened. The comparison shows that the overall pattern is not confined to one half of the experiment.

Third, Appendix D.3 reports a difference-in-differences check for eight user-level search and aggregate UGC consumption outcomes.

Finally, we vary the adjacent-query classification across nine combinations of 20-, 30-, and 60-minute windows and lenient, baseline, and strict semantic thresholds. The estimated share of additional same-day queries classified as follow-ups or distinct tasks remains between 80.7% and 84.3%, compared with 82.0% under the baseline rule. The adjusted range is similarly 80.6%–84.2%. Most additional same-day queries

therefore fall into these two categories under all nine rules. Appendix D provides the regression specifications, and Table A8 reports the classification sensitivity.

These findings motivate a model that links choices within search sessions to search initiation and evaluates how AI design changes both search activity and demand for creator content.

5 A Two-Level Model of Search and Demand for Creator Content

The model is motivated by the behavioral patterns in the experiment. Continued AI access increases search activity while reducing demand for UGC per session. Lower opening rates among recorded previews, together with differences in opening responses across information needs, support information substitution within a search. These findings motivate a two-level model that connects the benefits of acquiring information within a session to the incentive to initiate search. The link is prospective search value: the expected net benefit of a session before its results are encountered. This structure allows us to evaluate AI placement, answer informativeness, and the information shown before opening.

Within a session, users choose whether to open each previewed result and whether to view another preview. Openings provide information at a reading cost; additional previews also require effort. Information already acquired can reduce the benefit of further UGC, while the prospect of a more valuable session can encourage another query or search on another day. These are the model's channels for connecting per-session demand for UGC to search scale. A separate conditional relationship links direct UGC openings to downstream exploration. Figure 8 provides a flowchart of the model's decision sequence.

5.1 Timing and Persistent User Heterogeneity

Let i, t, s index users, days, and search sessions, and n index successive previews viewed within a session. Let k denote query intent, using the seven categories introduced in Section 3. The index n counts encounters, not positions in the ranked list. As in the experimental analysis, a preview shows part of a search result and an opening is a logged click into its full content.

At the start of a day, a user chooses whether to search. Submitting a query starts a session with a first preview, incurring the same class-specific previewing cost as each subsequent preview. At each preview, the user first chooses whether to open the content or skip it. An opening provides information and incurs reading cost. The user then chooses whether to view another preview or stop viewing results, taking account of any information just acquired. After the session, the user may submit another query that day. Sections below first describe the within-session problem and then use its expected value to explain these two search-initiation decisions.

There is at most one AI answer per session. We model each result's first preview and whether it is opened, merging repeated records as described in Appendix F.1. Once previewed, a result cannot be encountered

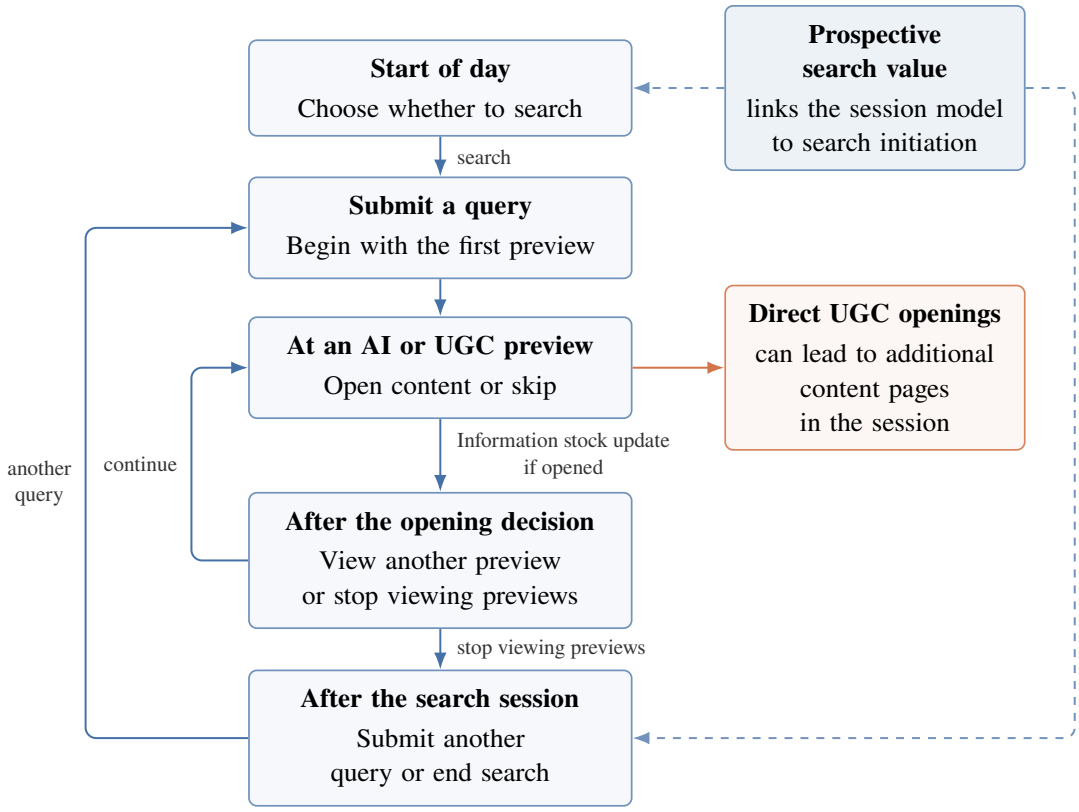


Figure 8: From Choices Within a Search Session to the Scale of Search Activity

Notes: Solid arrows show the decision sequence. Dashed arrows show how the prospective value of a search affects daily search participation and the initiation of another same-day session. Downstream UGC openings are modeled as a separate conditional outcome within the recorded query session; this outcome does not enter the preview-sequence decision problem.

again in the modeled sequence, whether opened or skipped. Users need not view results in rank order: a later preview may be at a lower position or at an earlier position not yet viewed. The recorded session also includes downstream navigation and ends when search ceases or another query is submitted.

For the structural analysis, non-AI result types are pooled under the UGC category. The model retains this pooled definition when predicting direct openings under alternative designs.

Persistent user heterogeneity. Two unobserved latent classes, $h_i \in \{1, 2\}$, capture persistent differences in information values, effort costs, and search propensity. Class remains fixed throughout the experiment. Recorded characteristics Z_i predict Class 2 membership:

$$\pi_{i2} = \mathbb{P}(h_i = 2 \mid Z_i) = \Lambda(Z_i'\delta), \qquad \pi_{i1} = 1 - \pi_{i2}, \tag{4}$$

Here δ contains membership coefficients and $\Lambda(u) = \exp(u)/[1 + \exp(u)]$. The ten elements of Z_i are an intercept, five platform age-group indicators, age missing, female, gender missing, and pre-experiment creator status. Reference categories are age group 0 (age <20), male users, and noncreators.

Assumption 1 (Stable expectations across sessions). Users' expectations about the search environment under a given display rule remain stable across sessions. The feature's availability for more than two months before randomization motivates this restriction, which abstracts from further cross-session learning. Within a session, users still learn the information in an opened answer, keep track of results already previewed, and update whether AI may still appear. An opening does not change the UGC-quality primitives, but its information increment can reveal the layout. Daily shocks are independent conditional on class and environment. Information stocks reset with a new query: the model treats it as a new information task rather than carrying knowledge across related queries.

5.2 The Search Environment and What Users Know

Let $g = 0$ denote conventional search without AI and $g = 1$ the existing selective AI feature. We call these environments NoAI and Current, respectively. Assignment changes available answers and users' expectations, not a separate treatment coefficient in utility.

The differences in clicking and CTR across intents in Section 4.3.3 motivate conditioning the search environment on the information sought. Accordingly, AI availability, answer-information distributions, expected reading lengths, UGC information increments, and encounter probabilities vary with query intent. Information-valuation weights vary across persistent user classes; intent-specific responses can arise through differences in these search environments.

Assumption 2 (Policy-invariant organic UGC supply). Potential UGC results have a common distribution across experimental environments. The platform confirmed that underlying UGC search is non-personalized and that AI availability changes neither retrieval nor organic ranking.

To evaluate continued search, users need expectations about which result they may encounter next. These expectations depend on which results remain available: once a result has been previewed, it is removed from subsequent consideration, whether opened or skipped.

The rank-specific opening and CTR patterns in Section 4.3.2 motivate retaining both where results are encountered and the information provided by the underlying UGC. We distinguish display rank, starting at zero on the combined results page, from UGC-only rank, which excludes the AI answer. We track previously previewed UGC results at UGC-only ranks zero through four individually. For results at later UGC-only ranks, we track how many remain rather than the identity of each result.[1] We refer to this group of later-ranked results as the tail. Within a query, these later results are treated as equivalent in their information increment and expected reading cost. Inserting AI preserves all 222 UGC candidates but shifts the display ranks at and below its position. Positions whose source, information increment, or existence differs across possible layouts remain distinct in the computation.

[1]This group initially contains 217 candidates, covering UGC-only ranks 5–221 in the constructed search environment. This is a modeling choice for the number of available results, not the number a user actually views before stopping.

We calibrate initial display-position weights from the control group's actual first previews, conditional on query intent and expected UGC reading length, and use these weights in both experimental environments. For subsequent previews, we estimate display-position weights from observed browsing prefixes pooled across treatment and control, through the first opening inclusive. At any display position, AI and UGC receive the same encounter weight. The probabilities depend on query intent, the current display position, and the results remaining under each possible layout. They do not have a separate treatment, source-type, or opening-choice coefficient.

The data reveal the next result only when a user continues browsing. We therefore need an assumption to describe what users who stop would have encountered had they continued. Assumption 3 below states that, conditional on the modeled state, observed transitions among continuing users also apply to users who stop.

Assumption 3 (A common conditional distribution of subsequent previews). Given query intent, the current display position, and the results remaining under a possible layout, the next-result distribution observed among users who continue also applies to users who stop. The encounter weights do not depend further on information already acquired, prior opening decisions, expected reading length, or user class. User beliefs over possible layouts can nevertheless change with the observed history.

The assumption separates the next result a user might encounter from the decision to continue. Two users facing the same conditional distribution can make different choices because they have acquired different amounts of information or face different costs. It rules out additional private information about unusually attractive remaining results that both encourages continuation and changes the next-result distribution beyond the conditions specified above.

Assumption 4 (Browsing without replacement). Let $\ell \in \{0, 1\}$ index layouts without and with an AI answer, and let $\mathcal{D}_\ell$ be the display positions in that layout. For a possible next display rank r, define its remaining availability by

$$R_{r\ell}(x) = \mathbf{1}\{r \in \mathcal{D}_\ell \text{ and rank } r \text{ has not been previewed}\}. \tag{5}$$

Here $\mathbf{1}$ is an indicator, and the current preview is already recorded as viewed in x. Let $\omega_{kjr} > 0$ be the per-result encounter weight at display rank r for intent k and current display position j. The weight is the same for AI and UGC at the same display position. Removing viewed results and rescaling within each layout gives

$$P_{k\ell}(r \mid j, x) = \frac{\omega_{kjr} R_{r\ell}(x)}{\sum_{e \in \mathcal{D}_\ell} \omega_{kje} R_{e\ell}(x)}. \tag{6}$$

Users average these layout-specific probabilities using their current beliefs $b_\ell(x)$. Initial previews use the calibrated initial weights in place of ω_{kjr}. Each denominator is computed within its layout before averaging;

AI is a possible result in this same distribution. For computation, B records individually tracked display positions and c counts viewed results in the common exchangeable tail. Positions whose source, UGC information, or existence differs across layouts remain explicit; tail probability is its per-result weight times the number remaining. A viewed result has zero return probability. For example, viewing display rank two removes that result even if it is not opened; it does not remove unviewed ranks zero or one. Further previews are unavailable once all results in the realized layout have been viewed. Exhaustion is recognized before the continuation choice, with no extra preview cost or choice shock. The same encounter law enters users' expected values and simulated paths.

We use browsing without replacement as an approximation to search over distinct results. In the estimation sample, additional recorded openings of the same UGC result within a session account for only 0.96% of all recorded UGC opening events, supporting our abstraction from repeated openings.

Information available to users. Under Current, an available AI answer occupies display rank zero, with intent-specific availability $\rho(k)$ estimated while allowing an available answer to remain unviewed. Availability means that the answer can be encountered; opening it remains the user's choice. Display rank zero need not be the first preview. A UGC preview at display rank zero rules out AI under Current; a UGC preview at another rank need not do so. After AI is previewed, no second AI answer can arrive, whether the user opens or skips it.

When a counterfactual policy places AI later in the search results, users form expectations about whether and when they will encounter an AI answer. These expectations depend on the display rule and the results already viewed, and are updated as browsing proceeds without AI appearing.

AI and UGC are drawn jointly through the display-position law. Under an AI insertion at position p, a UGC result at display rank r has UGC-only rank $r - \mathbf{1}\{r > p\}$ if AI is available, and rank r otherwise. Before the layout is revealed, users integrate over these possibilities using their observable history (see Appendix F.3 for details).

Users observe source and rank at a preview, plus reference count for AI, but not full-answer length or complete AI information. The interface's limited previews motivate expectations about full-answer information and reading length, described below. For UGC, rank proxies information. Equating calibrated distributions with subjective beliefs is a maintained assumption, not direct measurement of expectations.

Opening acquires the measured information; skipping leaves information stock unchanged. Later results become known only when encountered.

5.3 Information Acquired and Expected Reading Cost

We distinguish the search environment from information already acquired. Conditional on user class h and policy g, the state immediately after preview n and before the opening decision is (x_n, m_n), where

$$\begin{aligned} x_n &= \left(k, b^L, j_n, \mathrm{src}_n, \zeta_n, \mathcal{B}_n, \boldsymbol{b}_n, z_n^A\right), \\ m_n &= (m_n^A, m_n^U). \end{aligned} \tag{7}$$

The search-environment state x_n has the following components. Query intent k and expected UGC length group b^L remain fixed within the session; together they determine the expected UGC reading length. The current preview has display rank j_n and source $\mathrm{src}_n \in \{A, U\}$. Under Current, its AI preview profile is $\zeta_n = (b_n^{\mathrm{ref}}, \mu_{L,n}^A)$: the reference-count group and the expected AI reading length formed from the preview's cues. Set $\zeta_n = \varnothing$ at a UGC preview. Neither realized AI information nor actual full-answer length is known from this baseline profile. The disclosure policies considered below augment ζ_n with the displayed information label. The seven intent categories include Ambiguous / Other / Unclassified, which includes missing labels.

The set $\mathcal{B}_n$ contains all display positions previewed so far, including j_n, whether opened or skipped. Together with policy g, it determines which results remain in each possible layout. The belief vector is

$$\boldsymbol{b}_n = (b_{0n}, b_{1n}), \qquad b_{\ell n} = \Pr(\ell \mid \mathcal{H}_n^{\mathrm{obs}}, k, b^L, g), \quad b_{0n} + b_{1n} = 1, \tag{8}$$

where $\mathcal{H}_n^{\mathrm{obs}}$ contains previews and information revealed by earlier openings, not future results. Thus $b_\ell(x_n) = b_{\ell n}$. Finally, z_n^A indicates whether AI has been previewed, including at n; it does not indicate whether AI has been opened. Once $z_n^A = 1$, the single AI opportunity has been consumed, even if skipped.

The stock vector m_n records information obtained from earlier openings; both stocks begin at zero. Previewing alone changes x_n but not m_n. Opening updates stock and may also update the layout belief, whereas skipping leaves both unchanged. Opening history matters for subsequent within-session choices through these stocks and beliefs, not as a separate list of past opening decisions. Appendix F.3 describes the equivalent computational representation of viewed results. Table 15 collects the notation; below we suppress the subscript n on x and m where no ambiguity arises.

AI information amount and density. Let N count distinct query-relevant semantic units in the full AI answer: propositions, instructions, explanations, conditions, comparisons, or concrete recommendations. Repetition counts once, qualifications and attribution are retained, and generic or unrelated text is excluded. This equal-weight content measure follows semantic-unit approaches (Brown et al., 2008; Lex et al., 2012; Nenkova and Passonneau, 2004). It measures query-relevant content expressed in an answer, not independently verified accuracy or how much an individual user understands. Appendix F.2 details extraction and review.

Table 15: Core Model Notation

Symbol	Meaning and units
i, t, s, n	User, day, search session, and preview-view sequence index
h, k, g	User class, query intent, and AI search environment
x_n, m_n	Search-environment state and acquired-information stocks at preview n, before opening
$j_n, \mathrm{src}_n, \zeta_n$	Current display rank, source, and AI preview profile (empty for UGC)
$\mathcal{B}_n, z_n^A$	All display positions already previewed, including the current one; indicator that AI has been previewed
$B, c, R_{r\ell}$	Viewed individually tracked display positions, viewed common-tail count, and remaining availability at display rank r in layout ℓ
$\ell, b_\ell(x)$	AI-absent or AI-present layout and the user's current belief over layouts
N, ν	AI semantic-unit count and its scaled value, $\nu = N/10$
L, D	AI answer length in characters and semantic units per 100 characters
m^A, m^U	AI information stock measured by ν; UGC information stock between zero and one
$r^{\mathrm{org}}, w_k(r)$	UGC-only rank, excluding AI, and the information weight assigned to an opening
S^U, τ_k	Sum of UGC opening weights and the intent-specific stock scale
b^{ref}, b^L	AI reference-count group and expected UGC length group
$G_h(m)$	Gross utility from information already acquired
μ_L^j, c_h^j	Expected normalized length and reading cost for source $j \in \{A, U\}$
a_n, c_n	Indicators for opening the current result and viewing another preview
$C_{h,n}, M_{h,n}$	Value of continuing; value of the choice between continuing and stopping
$Q_{h,n}^0, Q_{h,n}^1, V_{h,n}$	Skip value, opening value, and value before choosing at a preview
$v_h(k, g)$	Expected net value of starting a search session
$\mathcal{V}_h^Q, \mathcal{V}_h^D$	Prospective values used for another same-day query and daily search

Notes: Utilities, costs, and prospective values use the normalized choice-utility scale. Preview sequence n counts encounters; it is not the result's rank. Each symbol is explained with its equation in the main text.

We distinguish information amount from concentration. Let L be the non-whitespace character count of the normalized full-answer text, excluding hidden reasoning. Define

$$\nu = \frac{N}{10}, \qquad D = \frac{100N}{L} \quad (L > 0). \tag{9}$$

Thus ν is information amount in units of ten semantic units, whereas D is concentration per 100 characters. Amount enters stock and expected length enters cost; density does not enter utility separately. Opening sets $m^A = \nu$ for the session.

Before opening, users form expectations about the amount of information in the full AI answer. At an AI preview, this belief is

$$\nu \sim \mathcal{H}^A(\cdot \mid k, b^{\mathrm{ref}}), \tag{10}$$

Here $\mathcal{H}^A$ is the distribution of possible AI information amounts, and b^{ref} denotes the visible reference-count group: 0, 1–3, 4–7, or at least 8. We construct this distribution from valid information measurements within each intent and reference-count group. Reference count is visible before opening (Figure 9); the red box

marks this cue on the AI preview. It informs expected information, while position determines when AI may appear. The count is a signal used to form beliefs, not the number of semantic units in the answer or a separate utility component. The realized information amount is only revealed on opening.



Figure 9: Reference Count Displayed in the AI-Answer Preview. The red box highlights the number of referenced content items shown before opening the full answer.

UGC information. We use UGC rank as a signal of relevance and quality. The index r^{org} records the UGC-only rank, excluding the AI answer, starting at zero. Inserting AI changes display rank without changing this UGC-only rank or removing a UGC result. For intent k, let $F^{\text{org}}_{k,0}(r^-)$ denote the share of viewed UGC results in the control group with UGC-only ranks strictly smaller than r, meaning positions earlier in the list. We assign each distinct UGC opening at UGC-only rank r the information weight

$$w_k(r) = 1 + \left[1 - F^{\text{org}}_{k,0}(r^-)\right]. \tag{11}$$

The weight ranges from one to two and is larger nearer the top. Both arms use the same rule. Write $\widetilde{w}_k(d) = w_k(d)$ for UGC-only rank groups $d = 0, \ldots, 4$. Tail openings instead use $\widetilde{w}_k(T) = \bar{w}_{kT}$, the mean raw rank weight of control tail previews within intent. Each new tail opening earns this increment; each tail preview removes one result from those still available.

Information accumulates across openings, with decreasing additions to the stock. If S^U is the sum of these grouped weights for opened results, the stock and its update after one more opening are

$$m^U = 1 - e^{-S^U/\tau_k}, \qquad m^{U+} = 1 - (1 - m^U)e^{-\widetilde{w}_k(d)/\tau_k}. \tag{12}$$

The superscript + denotes stock after opening. The scale $\tau_k > 0$ converts cumulative opening weights into a stock between zero and one. It is fixed so that the 90th percentile of positive end-of-session *sums of weights*, using individual ranks in the no-AI reference, maps to a stock of 0.9. We retain this scale after grouping later ranks, so the grouped stock's percentile need not equal 0.9 (see Appendix F.2 for details). Stock rises toward one: the same opening weight adds less to the stock when more information has already been acquired, as represented by $1 - m^U$ in the update. Weights and scales are fixed before behavioral estimation.

Gross information utility. The decline in opening among recorded previews, together with larger clicking and CTR declines for factual/explanatory than experience/opinion queries, motivates overlapping information benefits from AI and UGC. We assume that the two information stocks summarize the benefits acquired within a query. Users value information from both sources and obtain smaller additional benefits as their stocks grow. We represent this with

$$G_h(m^A, m^U) = \log(1 + \chi_{Ah} m^A + \chi_{Uh} m^U). \tag{13}$$

The nonnegative χ_{Ah} and χ_{Uh} value source-specific stocks. Logarithmic utility implies diminishing returns and, when both weights are positive, a lower benefit from a given UGC increment after AI information. We maintain this diminishing-returns structure to represent overlapping information needs. We normalize the coefficient outside the logarithm to one to fix the scale of information utility for identification. The AI and UGC stocks use different measurement scales, so the sizes of χ_{Ah} and χ_{Uh} alone do not rank the value of reading an AI answer versus a UGC page. The relevant benefit is the change in G_h produced by that opening.

Expected reading cost. Expected normalized answer length $\mu_L^j(z)$ proxies anticipated effort for source $j \in \{A, U\}$. Here z contains the query and preview information used to predict length: query text and intent, plus visible reference count for AI. It differs from the user characteristics Z_i that predict class membership. Separate predictors for AI and UGC put length on a common normalized log-length scale. Predictions are trained on other queries (see Appendix F.2 for details).

The reading cost of an opening is proportional to this expected length:

$$c_h^j(z) = \gamma_{Lh} \mu_L^j(z). \tag{14}$$

The common within-class coefficient $\gamma_{Lh} \geq 0$ gives the effort cost per unit of expected normalized length; source-specific expectations may differ. This treats normalized AI character counts and UGC word counts as comparable effort proxies. Conditional on modeled information and search state, expected length enters utility only through reading cost. This exclusion separates effort from information benefits. Query-held-out checks assess average length calibration, not the recovery of subjective expectations.

UGC length expectations form at query submission and remain fixed throughout the session, including when AI appears first. Thus all UGC openings in a given query carry the same expected-length cost for a user of class h; the cost can differ across queries. Opening a UGC page does not revise the expected length of the remaining pages. For computation, we group similar expected lengths within intent and let b^L index these groups. This approximation reduces the number of environments to solve without adding a new behavioral parameter.

5.4 Dynamic Choices Within a Search Session

Choice timing and shocks. The user first decides whether to open the current result, then learns any information revealed by opening, and finally chooses whether to view another preview. Let $V_{h,n}(x, m; g)$ be the expected *additional* net utility from the current preview onward, conditional on class h, search state x, acquired stock m, and environment g. It includes the options to open, skip, continue, and stop, but does not count previously acquired utility again. The current result is already recorded as viewed and cannot be encountered again.

Opening and continuation choices each include temporary utility shocks that are independent conditional on class and state. The shocks follow Type-I extreme-value distributions with mean zero and unit scale. This specification yields the logit probabilities below. $V_{h,n}$ averages over the opening-choice shocks before they are realized.

We solve backward through the decision sequence. First consider the choice *after* opening or skipping, when the information stock has already been updated. Viewing another preview offers its expected value less previewing cost. Stopping adds no further payoff from this session: the user retains all information already acquired, but incurs no further reading or previewing cost. The value of continuing, $C_{h,n}$, and the expected value of choosing between continuing and stopping, $M_{h,n}$, are

$$C_{h,n}(x, m; g) = \begin{cases} \mathbb{E}_{\mathcal{G}_g(\cdot|x)}[V_{h,n+1}(x', m; g)] - \kappa_{Bh}, & \text{if another preview remains,} \\ -\infty, & \text{otherwise,} \end{cases} \tag{15}$$

$$M_{h,n}(x, m; g) = \log\{1 + \exp[C_{h,n}(x, m; g)]\}. \tag{16}$$

The next-preview distribution $\mathcal{G}_g(\cdot \mid x)$ uses the joint AI–UGC display law in Equation (6), averaged over the user's current layout beliefs and possible AI preview profiles. At the next preview, x' records the observed display rank and source, removes that result, and updates the user's beliefs. These updates precede the next opening choice. An AI preview removes no UGC result. After exhaustion, $M = 0$ and continuation probability is zero.

Each preview, including the first, costs $\kappa_{Bh} \geq 0$, regardless of rank gap or prior effort. This is the effort of inspecting a preview, not a cost proportional to scrolling distance. Thus C compares that cost with the next

result's expected value. In Equation (16), the term 1 represents stopping with zero additional payoff (i.e., $\exp(0)$), while $\exp(C)$ represents continuing. The log-sum expression gives the expected maximum payoff over the two choices, averaging over the choice shocks.

The earlier opening decision anticipates these subsequent choices. Skipping preserves the information stock. Opening adds information, incurs reading cost, and changes the value of continuing. Let T_U and T_A denote the information-state transition functions following a UGC opening and an AI opening, respectively:

$$T_U(m) = (m^A, m^{U+}), \qquad T_A(m, \nu) = (\nu, m^U).$$

The UGC update follows Equation (12); its intent and UGC-only rank arguments are suppressed. If the layout is unresolved, write $T_U(m; \ell)$ for the corresponding stock update. Opening reveals its information increment, updating the layout belief only when the possible increments differ; write x^1_ℓ for the resulting state. AI sets its stock to realized ν. Let x^0 and x^1 denote search states after skipping and opening. After an AI preview, $z^A = 1$ under either action, even if opening reveals $\nu = 0$; zero AI stock therefore does not imply that AI remains available. The values of skipping (Q^0) and opening (Q^1) are

$$Q^0_{h,n}(x, m; g) = M_{h,n}(x^0, m; g), \tag{17}$$

$$Q^1_{h,n,U}(x, m; g) = \mathbb{E}_{\ell|x}\left[G_h(T_U(m; \ell)) - G_h(m) - c^U_h + M_{h,n}(x^1_\ell, T_U(m; \ell); g)\right], \tag{18}$$

$$Q^1_{h,n,A}(x, m; g) = \mathbb{E}_{\nu\sim\mathcal{H}^A}\left[G_h(T_A(m, \nu)) - G_h(m) - c^A_h + M_{h,n}(x^1, T_A(m, \nu); g)\right]. \tag{19}$$

The three terms in an opening value are the additional information benefit, given by the change in G_h, the reading cost, and the value of deciding what to do afterward. A skip has no information benefit or reading cost, but preserves the option to continue. For UGC, the information increment is specified by UGC-only rank and intent; the expectation integrates unresolved layouts before opening and uses the revealed increment afterward. For AI, the user does not yet know how much information the full answer contains when deciding whether to open it. The expectation therefore averages the information gain and the value of subsequent choices over possible realizations of ν. Once the answer is opened, continuation uses the realized ν. The expectation is therefore taken before the opening choice, not before the later continuation decision. The pre-choice value and probabilities are

$$V_{h,n}(x, m; g) = \log\{\exp(Q^0_{h,n}) + \exp(Q^1_{h,n})\}, \tag{20}$$

$$\mathbb{P}(a_n = 1 \mid x, m, h, g) = \Lambda(Q^1_{h,n} - Q^0_{h,n}), \tag{21}$$

$$\mathbb{P}(c_n = 1 \mid x^{a_n}, m^{a_n}, h, g) = \Lambda\{C_{h,n}(x^{a_n}, m^{a_n}; g)\}. \tag{22}$$

Here Q^1 uses the current source. The indicator $a_n = 1$ denotes opening, and $c_n = 1$ denotes viewing another preview; c_n is distinct from the count c of tail previews already viewed. The states (x^{a_n}, m^{a_n}) reflect the

opening decision, including revealed AI information when applicable. The opening probability rises with the net advantage of opening over skipping, $Q^1 - Q^0$. The continuation probability depends on C at the updated state. Reading cost is already incurred at that point and is not charged a second time. We fix the within-session discount factor at one because the observed choices provide limited information for identifying time discounting separately from the costs and benefits of continued search.

This structure separates the value already obtained from the value of further search. Acquired information is retained on stopping. Each opening earns only its additional information benefit, so gains along a path sum to the final information utility. Let N_s^B be the total number of AI and UGC previews viewed in session s. Excluding choice shocks, the path payoff is

$$G_h(m_{\text{terminal}}) - \sum_{\text{openings}} c_h^j - N_s^B \kappa_{Bh}, \tag{23}$$

Here m_{terminal} is the final acquired stock. Stopping retains this information without a second terminal reward.

5.5 From Session Value to Search Initiation

The experiment shows that most additional sessions occur later on the same day, with a further increase in active search days (Section 4.2.2). This pattern motivates separate equations for another same-day query and daily search participation. We link both decisions to the prospective value produced by the within-session problem. Higher expected search value can encourage more sessions even as acquired information reduces the benefit of additional UGC within a session. The two initiation equations use this expected value without solving a single dynamic optimization problem over all future queries.

Before submitting a query, users consider what the session could deliver. The prospective value averages the initial decision value over possible initial search environments for intent k, then deducts the first-preview cost:

$$v_h(k, g) = \mathbb{E}_{\mathcal{G}_g^0}[V_{h,1}(x_1, (0, 0); g) \mid k] - \kappa_{Bh}. \tag{24}$$

Here $\mathcal{G}_g^0$ includes the query-fixed expected UGC length group, intent-specific AI availability, the first encountered display rank and source, and possible AI reference-count profiles with their expected lengths and information distributions. The state x_1 contains what the user knows on encountering that first preview, not information revealed only by opening. Both stocks begin at zero, but $V_{h,1}$ accounts for all later information, effort, and stopping choices. Search can therefore have positive prospective value before information is acquired.

Assumption 5 (Value-mediated search initiation). Assignment affects initiation only through prospective session value, without a separate treatment term. The value is prospective: it concerns the next search opportunity, not realized satisfaction or the information stock left by the previous session. First-query and

conditional next-query intent distributions are held fixed across environments. We maintain these restrictions to connect activity to modeled information benefits and effort.

After a query with intent k, the user considers searching again that day. Let $\mu_0(k' \mid k)$ be the probability that the next same-day query has intent k', conditional on the current intent k. We estimate this distribution from control-group query sequences. The expected value of the next search opportunity determines the probability of another query:

$$\mathcal{V}_h^Q(k, g) = \sum_{k'} \mu_0(k' \mid k) v_h(k', g), \tag{25}$$

$$\mathbb{P}(H_s = 1 \mid h, k, g) = \Lambda\{\alpha_h^Q + \rho^Q \mathcal{V}_h^Q(k, g)\}. \tag{26}$$

$H_s = 1$ denotes another same-day query. Its class-specific baseline is α_h^Q, and ρ^Q is the common response to prospective value. The next intent is drawn from $\mu_0(\cdot \mid k)$ and stocks reset.

At the start of a day, users have not yet submitted the first query. Let $\mu_0(k)$ be the control-group distribution of first-query intents on active search days. We use this distribution to calculate the expected value of starting search. This value governs whether any search occurs that day:

$$\mathcal{V}_h^D(g) = \sum_{k} \mu_0(k) v_h(k, g), \tag{27}$$

$$\mathbb{P}(A_{it} = 1 \mid h, g) = \Lambda\{\alpha_h^D + \rho^D \mathcal{V}_h^D(g)\}. \tag{28}$$

$A_{it} = 1$ denotes at least one search that day. The class-specific intercept α_h^D captures baseline participation, and ρ^D its common sensitivity to expected value. We also refer to this daily participation decision as daily activation. The two coefficients ρ^Q and ρ^D allow same-day querying and daily participation to respond differently to prospective value on the same session-utility scale.

The two prospective values use different intent weights. At day start, the user averages over possible first-query intents. After a session of intent k, the user averages over the intents that tend to follow k. The underlying session values $v_h(k', g)$ are the same in both cases. Thus a design change first alters the value of search for each intent, then changes the incentives to initiate sessions through these averages.

The maintained common intent laws are motivated by experimental-group comparisons that do not detect differences in first-query intent or next-query intent conditional on current intent ($p \approx 0.762$ and $p \approx 0.765$, respectively; see Appendix F.3 for details). These checks concern observed searches, not latent information needs, and do not establish exact equality or invariance under every new policy. Holding the intent laws fixed isolates the response in search initiation. Policies can still change the mix of realized sessions because continuation to another query depends on current intent. The experimental evidence on topical breadth is separate from this model of search volume. Appendix F.5 shows how daily activation and repeated same-day queries determine expected session counts.

5.6 Downstream UGC Openings

Direct UGC openings can lead users to additional creator content within the same session. We model whether downstream UGC navigation occurs and how many UGC pages are opened when it does. Let n_s^U be the number of distinct direct UGC openings in session s. Let Y_s count UGC content-page openings reached through subsequent navigation within that session. It excludes direct search-result openings and creator-profile pages.

Conditional downstream response. Conditional on user class and the number of direct UGC openings, downstream outcomes follow the same relationship across search environments. They do not enter the within-session opening and continuation values. This auxiliary relationship connects direct UGC openings to further exploration without modeling each subsequent navigation choice. Its invariance is maintained for policy prediction, not implied by randomized assignment. Sessions with no direct UGC opening have $Y_s = 0$ in the model. For sessions with $n_s^U > 0$, the two relationships are

$$\mathbb{P}(Y_s > 0 \mid h, n_s^U) = \Lambda\{\beta_{0h}^H + \beta_{1h}^H \log(1 + n_s^U)\}, \tag{29}$$

$$Y_s - 1 \mid Y_s > 0, h, n_s^U \sim \text{NB2}(\mu_{sh}, \phi), \qquad \log \mu_{sh} = \beta_{0h}^N + \beta_{1h}^N \log(1 + n_s^U). \tag{30}$$

The first equation describes whether any downstream opening occurs. The second describes the number of openings beyond the first, conditional on downstream navigation. The negative-binomial distribution (NB2) allows this excess count to vary more than a Poisson count with the same mean. The intercepts $\beta_{0h}^H, \beta_{0h}^N$ and slopes $\beta_{1h}^H, \beta_{1h}^N$ relate direct openings to downstream incidence and the excess count conditional on navigation. This excess can equal zero when only one downstream page is opened. Conditional on $Y_s > 0$, expected total openings are $1 + \mu_{sh}$. NB2 variance is $\mu_{sh} + \mu_{sh}^2/\phi$, with common dispersion parameter $\phi > 0$. Policy predictions retain these conditional relationships while direct openings change; downstream outcomes do not feed back into preview values.

6 Structural Estimation and Identification

6.1 Estimation Sample and Empirical Inputs

The structural sample covers March 30–April 14, 2025. We use this 16-day window because the available AI-answer text records provide consistent daily coverage only through April 14. These texts are needed to measure query-relevant information and answer length, which inform the model's information benefits and expected reading costs. We link them to users' search histories over the same period. The experimental analysis uses the full 26-day experiment because its search, preview-view, and page-opening outcomes can be measured from behavioral logs without these text inputs.

Drawing on the experimental cohort of 100,000 accounts, we construct a balanced daily panel of linked accounts, including users who never search. Appendix F.1 details the linked panel and measurement. The linked data contain 304,306 search sessions and 3,070,457 distinct result-slot previews, including 28,232 AI previews. Nonsearchers contribute the likelihood of daily inactivity and help identify persistent differences in activity.

Session-level records identify whether users open an AI answer. Among AI previews viewed, 79.2115% lead to an opening. Appendix F.1 describes outcome measurement and the construction of within-session choice sequences.

Information measures, expected lengths, rank weights, stock scales, search environments, and intent distributions are constructed before behavioral estimation and held fixed as we estimate preferences and costs. These empirical inputs describe the information and search opportunities users face; they are not additional behavioral parameters. Appendix F.2 reports measurement checks and query-held-out length predictions. Estimates and standard errors condition on these inputs.

6.2 Identification

Identification uses variation in opening, continuation, and search-initiation choices across user histories and search environments. The parameters are estimated jointly: the same information values and effort costs must explain both current choices and subsequent behavior. Random assignment provides exogenous variation in access to Generative AI Search. The information measures, expectations, and timing restrictions described above determine how other observed differences enter the model.

UGC information weight, χ_{Uh}. Variation in accumulated UGC stock and the information increment from the current result helps identify χ_{Uh}. Control-group histories provide these comparisons without an AI answer in the search environment. Conditional on expected reading cost and the remaining search opportunities, opening and continuation choices reveal how users value additional UGC information at different stock levels. The fixed rank-weight formula and intent-specific stock scale anchor the units in which this information value is measured.

AI information weight, χ_{Ah}. Variation in the expected amount of AI information across query intents and reference-count groups helps identify χ_{Ah} from AI-opening choices. After an answer is opened, variation in the information it contains provides a second source of identification through subsequent UGC openings and continuation decisions. The AI weight must explain both the willingness to acquire information and the response after acquiring it, conditional on reading cost and the existing UGC stock. The opening decision and later choices are estimated jointly, with selection into opening governed by the modeled state and user

class. These comparisons use the fixed $N/10$ information scale and the diminishing returns and cross-source interaction specified in Equation (13).

Reading-cost coefficient, γ_{Lh}. Variation in expected answer length across queries within intent and between AI and UGC helps identify γ_{Lh}. Conditional on expected information returns, longer expected answers increase the effort associated with an opening. The identifying exclusion is that expected length affects utility through reading effort, given the modeled information and search state.

Opening and subsequent continuation choices help distinguish this cost from the information weights. Reading cost is incurred when a result is opened. Acquired information remains in the stock and changes the benefits of later openings. The model therefore uses both the response to expected length before opening and the response to acquired information afterward.

Preview cost, κ_{Bh}. Variation in continue-versus-stop choices as the expected value of the next preview changes helps identify κ_{Bh}. These choices are observed after both openings and skips. Conditional on the updated information stock and the next-preview distribution, they reveal the effort users are willing to incur to inspect another result.

The data contain no sessions without a viewed preview. We charge the same κ_{Bh} for the first and each subsequent preview, with no separate first-preview parameter or continuation intercept. Relative to a free-first-preview normalization with all other primitives fixed, this subtracts κ_{Bh} from session value and is absorbed by the class-specific search-initiation intercepts. The comparison of continuation choices uses the next-preview distribution and its maintained assumptions described in Section 5.2.

Search-initiation intercepts and responses to value. Variation in prospective search value helps identify the initiation coefficients, also called value loadings. For another same-day query, ρ^Q uses value differences across current-query intents and randomized search environments. For daily search, ρ^D uses the experiment-induced difference in day-start value. Conditional on these values, the levels of search initiation help identify the class-specific intercepts, α_h^Q and α_h^D.

The identifying exclusion is that assignment affects initiation through the modeled prospective value, with no separate treatment term in either initiation equation. Stable expectations and common intent distributions provide the basis for comparing these values across environments.

For the daily equation, the identifying contrast is

$$\operatorname{logit} p_h^D(1) - \operatorname{logit} p_h^D(0) = \rho^D \left\{ \mathcal{V}_h^D(1) - \mathcal{V}_h^D(0) \right\}, \tag{31}$$

where $p_h^D(g)$ is the daily search probability for class h in environment g, and $\operatorname{logit}(p) = \log[p/(1-p)]$. Differencing removes the class-specific intercept. The session model supplies the change in prospective value.

The corresponding change in daily search log odds helps identify how strongly initiation responds to that value through ρ^D. The same-day equation uses an analogous comparison conditional on current intent. Because class membership is unobserved, these contrasts are recovered jointly from users' behavioral histories.

Downstream coefficients and dispersion. Among sessions with at least one direct UGC opening, variation in the number of direct openings helps identify the downstream slopes. Its relationship with the probability of any downstream UGC navigation informs β_{1h}^H. Its relationship with the number of downstream openings beyond the first, conditional on downstream UGC navigation, informs β_{1h}^N. The corresponding outcome levels help identify the intercepts β_{0h}^H and β_{0h}^N.

Variation around the conditional mean of the excess count, $Y_s - 1$, helps identify the negative-binomial dispersion parameter ϕ. These coefficients describe conditional relationships with direct UGC openings, as specified in the downstream model.

Latent classes and class-membership coefficients, δ. Persistent differences in behavior across a user's repeated choices help distinguish the two latent classes. Opening, continuation, search initiation, and downstream UGC openings jointly inform class membership. Conditional independence of the temporary choice shocks allows the model to separate persistent user differences from occasion-specific variation.

Associations between recorded characteristics and these behavioral patterns help identify δ in the class-membership probability $\Lambda(Z_i'\delta)$. Class-membership coefficients and class-specific behavioral parameters are estimated jointly. Ordering the daily initiation intercepts gives the classes consistent labels.

We normalize the opening-choice shock scale to one to set utility units. Imposing the same scale for continuation shocks is an additional restriction on their relative variability. The fixed information-stock measures set information units, while the unit coefficient outside the information-utility logarithm fixes its curvature relative to the choice shocks. A within-session discount factor of one rules out discounting across previews. These maintained restrictions, together with the empirical inputs, underpin the parameter interpretations above. Appendix F.6 reports local numerical diagnostics for the joint estimates.

6.3 Joint Likelihood and Inference

We estimate the model by finite-mixture maximum likelihood (McLachlan and Peel, 2000), choosing parameters that make the observed user histories most likely. For each class, the history combines choices within sessions, subsequent queries, daily search, and downstream UGC openings. Let ℓ_{ih}^W, ℓ_{ih}^Q, ℓ_{ih}^D, and ℓ_{ih}^Y

be the corresponding log likelihoods for user i conditional on class h. We combine them before averaging across the two possible classes:

$$L_{ih} = \exp(\ell_{ih}^{W} + \ell_{ih}^{Q} + \ell_{ih}^{D} + \ell_{ih}^{Y}), \tag{32}$$

$$\mathcal{L}(\theta) = \sum_{i} \log \left[\sum_{h=1}^{2} \pi_{ih} L_{ih} \right]. \tag{33}$$

L_{ih} is the probability of user i's full history conditional on class h. The prior weight π_{ih} comes from observed characteristics. The inner sum allows either class to generate the history; the outer sum adds user log likelihoods. The vector θ contains all parameters estimated jointly.

Following the nested estimation logic of Rust (1987), for each trial parameter vector we solve both classes' search problems by working backward from the point at which no results remain. The resulting values determine opening, continuation, and search-initiation probabilities. Those probabilities give the likelihood of observed histories, which the optimizer uses to update all parameters together. Prospective values are recomputed as parameters change. Here the finite set of remaining results permits backward solution, rather than an infinite-horizon fixed-point calculation.

Each observed choice contributes its model-implied probability. The AI-opening value averages over anticipated information before entering the logit choice probability. Continuation after opening conditions on the updated information stock. When AI availability is unresolved, the within-session likelihood sums over the complete AI-present and AI-absent paths compatible with the observed history. Each path includes its layout probability, encounter probabilities, and behavioral choices; we do not mix the two layouts separately at each choice. Choice probabilities use only the information available to the user at that point. The researcher's later inference about the layout does not enter the user's earlier beliefs. Appendix F.6 details the likelihood calculation.

Observed behavior also updates our assessment of class membership. The posterior probability weights a class's prior probability by how well it explains the user's full history:

$$\pi_{ih}^{\text{post}} = \frac{\pi_{ih} L_{ih}}{\sum_{h'=1}^{2} \pi_{ih'} L_{ih'}}. \tag{34}$$

The denominator normalizes these weights to sum to one across classes; h' indexes the alternatives in that denominator. The posterior summarizes evidence about the user's persistent class.

The model has 33 free parameters: ten class-membership coefficients, three common parameters ($\rho^{Q}, \rho^{D}, \log \phi$), and ten behavioral coefficients per class. The NB2 size ϕ is parameterized through its logarithm.

We maximize the observed likelihood with L-BFGS-B (Byrd et al., 1995), using multiple starting points and parameter scaling. Standard errors cluster by user and allow covariance across all behavioral blocks and classes. Appendix F.6 documents the backward solution, optimizer, and covariance calculation.

7 Structural Estimates and Model-Implied Predictions

7.1 Information Value, Effort, and Persistent Heterogeneity

Table 16 reports the 33 estimated parameters and their user-clustered standard errors. Panels A–C contain the within-session, search-initiation, and downstream parameters. Panel D reports the ten class-membership coefficients. Shared parameters appear once across the two class columns.

Both classes have positive estimated information weights and positive reading and preview-inspection costs. The average posterior probability of Class 1 is 82.4913%. This class has a lower baseline propensity to search than Class 2. At the same expected answer length, Class 1 faces a higher reading cost but pays less to inspect a preview.

The information weights operate through logarithmic utility. An AI opening contributes query-relevant information measured by $N/10$. UGC openings accumulate rank-weighted information on a bounded stock scale. The weights are interpreted on these source-specific scales; their magnitudes alone do not measure a preference for AI over UGC. The utility gain from either source depends on its information increment and the stocks already acquired.

Reading and previewing involve distinct costs. An increase of one tenth in expected normalized length adds 0.2299 units of reading cost in Class 1 and 0.1578 in Class 2. The preview cost is paid whenever a user views a preview, including the first, whether or not the result is opened.

Panel D relates recorded characteristics to class membership. Pre-experiment creators have 2.4200 times the prior odds of belonging to Class 2, holding the other characteristics fixed. Age group 2 has higher prior odds than the reference group; groups 4 and 5 have lower odds. These associations help describe persistent differences in search behavior.

7.2 Higher Session Value Can Generate More Search

The subsequent-query loading is $\widehat{\rho}^{Q} = 0.5267$ (SE 0.0919), and the daily-activation loading is $\widehat{\rho}^{D} = 0.3350$ (SE 0.1036). A one-unit increase in prospective value multiplies the odds of another same-day query by 1.6934 and the odds of daily activation by 1.3979. These transformations use normalized utility units; a policy's activity effect depends on the value change it produces and baseline participation.

Table 17 reports estimated gains in prospective search value under the existing selective AI feature, with user-clustered delta-method standard errors. Figure 10 displays the same contrasts with 95% confidence intervals. Each intent-specific entry subtracts the NoAI value from the Current value for the same class

Table 16: Structural Parameter Estimates

Parameter	Class 1: estimate (SE)	Class 2: estimate (SE)
Panel A. Information value and search costs		
Expected-length reading cost, γ_{Lh}	2.2992 (0.0351)	1.5784 (0.0126)
AI-information weight ($N/10$), χ_{Ah}	30.7495 (2.5284)	34.4251 (2.0969)
UGC-information weight, χ_{Uh}	1.2527 (0.1785)	0.9508 (0.0765)
Preview cost, κ_{Bh}	0.2040 (0.0036)	0.3212 (0.0021)
Panel B. Subsequent-query initiation and daily activation		
Subsequent-query intercept, α_h^Q	-1.1977 (0.2395)	-0.6885 (0.2178)
Daily-activation intercept, α_h^D	-4.4416 (0.2654)	-1.6772 (0.2471)
Subsequent-query value loading, ρ^Q	Shared: 0.5267 (0.0919)	
Daily-activation value loading, ρ^D	Shared: 0.3350 (0.1036)	
Panel C. Downstream UGC openings		
Initiation intercept, β_{0h}^H	-1.6674 (0.0561)	-1.4402 (0.0392)
Initiation slope, β_{1h}^H	1.6740 (0.0502)	1.7657 (0.0364)
Excess-count intercept, β_{0h}^N	0.7584 (0.0486)	1.8085 (0.0299)
Excess-count slope, β_{1h}^N	0.7989 (0.0383)	0.4723 (0.0170)
Count-dispersion parameter, ϕ	Shared: 0.4562 (0.0039)	

Panel D. Prior membership in Class 2: coefficients in $Z_i'\delta$	
Parameter	Estimate (SE)
Intercept	-1.8109 (0.0604)
Age group 1	-0.0073 (0.0445)
Age group 2	0.2403 (0.0459)
Age group 3	0.0534 (0.0479)
Age group 4	-0.5137 (0.0687)
Age group 5	-1.7034 (0.1592)
Age missing	0.2965 (0.0513)
Female	-0.0155 (0.0229)
Gender missing	-2.5807 (0.0787)
Pre-experiment creator	0.8838 (0.0208)

Notes: Parentheses contain user-clustered robust standard errors conditional on fixed first-stage inputs. Posterior class shares are 82.4913% for Class 1 and 17.5087% for Class 2. The reference categories in Panel D are platform-coded age group 0, male users, and noncreators. Both downstream slopes multiply $\log(1 + n_s^U)$; the excess-count equation models openings beyond the first, $Y_s - 1$, conditional on $Y_s > 0$. On its natural scale, $\widehat{\phi} = 0.4562$ (SE 0.0039).

Table 17: Prospective-Value Gains Under Existing AI Access

Value contrast	Class 1 (SE)	Class 2 (SE)
Factual / Explanatory	0.2594 (0.0067)	0.3667 (0.0071)
Procedural / How-to	0.3409 (0.0083)	0.4804 (0.0086)
Evaluation / Recommendation	0.1573 (0.0040)	0.2191 (0.0041)
Experience / Opinion	0.1874 (0.0048)	0.2645 (0.0051)
Exploratory / Open-ended	0.0985 (0.0024)	0.1375 (0.0025)
Current / Local / Navigational	0.1187 (0.0031)	0.1670 (0.0033)
Ambiguous / Other / Unclassified	0.0605 (0.0016)	0.0845 (0.0016)
Day start, first-intent weighted	0.1436 (0.0037)	0.2021 (0.0039)

Notes: Differences compare Current selective AI access, with an available AI answer at display rank zero, with no AI. Values are in normalized utility units, not percentages, and use the same UGC environment. Parentheses report delta-method standard errors of the differences, using the user-clustered structural-parameter covariance matrix and retaining covariance between Current and NoAI values. Day-start gains use fixed control-group first-query intent weights. Inference conditions on the estimated environment, information beliefs, expected lengths, and intent weights; it does not include uncertainty in these inputs.

and intent, holding the UGC environment fixed. Query-start value rises for every intent and both classes. Procedural and factual queries gain most; Ambiguous / Other / Unclassified gains least. These contrasts combine selective availability with the value of a search environment in which an available AI answer occupies display rank zero, not necessarily the first preview. They are changes in expected normalized utility, not opening rates or monetary welfare. The last row averages these query-start value gains using the control group's first-query intent distribution, connecting within-session choices to the incentive to begin searching on a day.

Class 2 obtains larger value gains and has a larger AI-information weight. Reading effort, the value of UGC, and later choices also affect session value. First-intent weighting gives day-start gains of 0.1436 and 0.2021 in Classes 1 and 2, respectively. Thus the model links the prospect of acquiring useful AI information to more frequent search, even when that information reduces further UGC openings within a session.

7.3 Downstream UGC Openings and Conditional Predictions

Panel C of Table 16 links direct UGC openings to further content exploration. Relative to one direct opening, three direct openings multiply the odds of downstream UGC navigation by 3.1910 in Class 1 and 3.4003 in Class 2. Conditional on downstream navigation, expected openings beyond the first increase by factors of 1.7398 and 1.3873. The common dispersion estimate, $\widehat{\phi} = 0.4562$, captures variation around these conditional means.

Table 18 compares observed outcomes with fitted conditional means for selected choice margins. Conditional predictions are evaluated on the observed choice risk sets. They average choice probabilities using each user's full-history posterior class weights and the researcher's full-session layout weights. These in-sample conditional predictions average probabilities at observed histories. The counterfactuals generate new histories under the estimated model and maintained environmental restrictions.

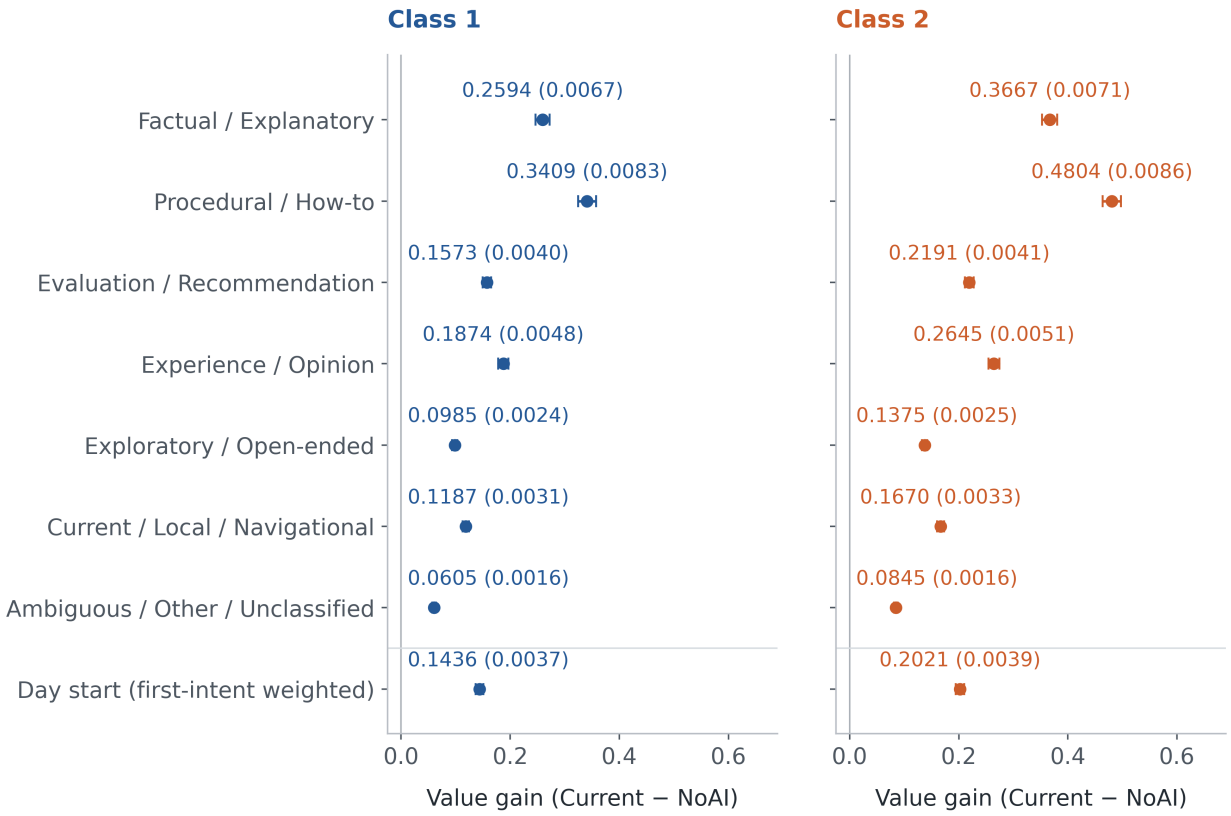


Figure 10: Prospective Search Value Rises Across Query Intents

Notes: Points reproduce the Current-minus-NoAI value gains in Table 17; annotations give estimates with standard errors in parentheses. Bars are pointwise 95% delta-method confidence intervals (estimate ±1.96 standard errors), conditional on the same fixed environmental inputs. The horizontal scale is normalized utility, not percentage growth or monetary welfare.

Table 18: Model Fit: Selected Outcomes

	Control		Treatment	
Outcome	Observed	Fitted	Observed	Fitted
UGC opening given preview (%)	16.81	16.60	16.51	16.58
AI opening given preview (%)	—	—	79.21	84.65
UGC opening per preview after AI opening (%)	—	—	16.05	15.43
Continue previewing after opening UGC (%)	96.18	89.16	96.51	89.63
Continue previewing after skipping UGC (%)	88.81	90.03	89.70	90.31
Another same-day query (%)	60.67	60.13	62.12	62.61
Active user-day (%)	7.24	7.24	7.48	7.49
Any downstream UGC opening (%)	56.80	56.97	56.95	56.83
Downstream UGC openings (per session)	6.38	6.39	6.45	6.39
Downstream UGC openings, if positive	11.23	11.20	11.32	11.23

Notes: Selected in-sample conditional means over the March 30–April 14, 2025 structural window. Fitted values use full-history posterior class weights and the researcher's full-session layout weights at observed histories. Dashes indicate empty control-group AI risk sets. The third row conditions on an earlier AI opening in the same session. Continuation rows include terminal previews. All three downstream measures condition on at least one direct UGC opening in the session. The penultimate row includes zero downstream counts; the last row additionally conditions on positive downstream openings and reports openings per such session.

8 Designing Generative AI Search

To address RQ3, we examine how the platform's design choices shape search activity and demand for creator content. We consider placement and information amount, allocation across query intents, and information disclosure before opening. The first changes the answer's timing and content; the second reallocates existing availability while holding answer characteristics fixed; the third changes what users know about an unchanged answer. Each comparison evaluates both search activity and demand for creator content.

We use two design objectives throughout. First, we ask whether a design improves both search activity and direct UGC openings per session relative to the existing design, which we call Current. This is a joint, or Pareto, improvement in these two behavioral outcomes, not a welfare comparison. Second, when joint improvement is not established, we examine which designs increase aggregate demand for creator content, measured by total direct UGC openings per user. We also report this aggregate outcome when both margins improve. Downstream openings are reported separately and are not part of this design objective. Failure to establish a joint improvement does not imply that Current is optimal. Designs can trade off the two margins and still perform better on aggregate direct UGC demand. An effect unresolved from zero, however, does not by itself establish such a trade-off.

8.1 Policy Comparisons and Simulation

Under each design, users reconsider opening and continuation choices over the finite set of remaining search results. The resulting session value affects daily search participation and subsequent-query initiation. Direct UGC openings also generate downstream UGC openings through the estimated conditional relationship, which remains outside the session decision problem.

NoAI removes AI while preserving the common UGC supply distribution. Current retains selective, intent-specific availability and always places an available AI answer at display rank zero, without forcing its preview first. Across policies we hold preferences, costs, observed user characteristics, the UGC environment, expected reading lengths, and first- and next-query intent laws fixed. The distribution of realized sessions can change because initiation responds to prospective value.

An AI *opportunity* means an answer is available at a designated position. *Exposure* occurs when the user reaches its preview. An *opening* is the subsequent choice to read the answer. Availability, exposure, and openings need not change together. An availability budget therefore differs from a target for AI use.

The main policy evaluation contains 100 paired replications of 1,000 users over 16 days. Policies share user and event draws within each replication. Per-user outcomes retain inactive users. Every query initiates one session, and the per-session outcome divides direct UGC openings by session count within a replication before averaging across replications. Appendix G.6 describes aggregation and pairing.

The main tables and figures report Monte Carlo point estimates and error from simulating a finite number of user histories at the fitted parameters. We describe a simulated increase or decrease as resolved only when its paired 95% Monte Carlo interval excludes zero. Joint improvement requires this criterion to hold for both margins. These intervals measure simulation precision at fixed parameters, not statistical inference for policy effects estimated from data.

The reference trade-off. In the common-population simulation, Current generates 3.2034 sessions per user, compared with 2.8644 under NoAI. Direct UGC openings per session fall from 1.6656 to 1.6402, while total direct openings rise from 4.7714 to 5.2545 per user. More frequent search outweighs lower demand for UGC per session. Appendix G.2 reports the remaining benchmark outcomes.

8.2 Placement and AI Information Amount

Moving AI answer down the results page gives users earlier opportunities to open UGC, but can reduce the value of initiating search. We ask whether a more informative answer at the same reading length can offset that loss.

We compare display positions zero through three and information multipliers 0.8, 0.9, 1.0, 1.1, and 1.2. Position zero places AI at the top of the displayed list, but does not force its preview first. AI and UGC at the same display position receive the same encounter weight. Previewing a lower-ranked result leaves unseen higher-ranked results available. Moving AI changes UGC display ranks but not UGC-only ranks, information quality, or the supply of 222 UGC results. The availability probabilities remain at their Current levels.

The multiplier scales the semantic-unit count N and its stock $N/10$, holding expected answer length and reading cost fixed. A multiplier of 1.2 represents 20% more query-relevant information at the same length. Users anticipate the new information distribution and acquire the information after opening. Position zero with multiplier one reproduces Current, giving 20 displayed combinations including that benchmark.

Joint improvement in search activity and per-session demand. Figure 11 highlights a scale–intensity trade-off. Across the information levels considered, delaying the AI answer increases direct UGC openings per session but reduces search activity. Providing more information at the same expected reading length can increase search activity and partly offset the loss from delayed placement. However, the designs that raise search activity relative to Current have per-session effects that are not resolved from zero by the paired Monte Carlo intervals. The evaluated designs therefore do not establish joint improvement in search activity and direct UGC demand per session.

Aggregate demand when joint improvement is not established. With the same 20% information increase, position two reduces sessions by 1.2552% while raising direct UGC openings per session by 3.0761%. Total direct UGC openings rise by 1.7850%. The design-selected total-opening rule, marked by stars in Figure 11,

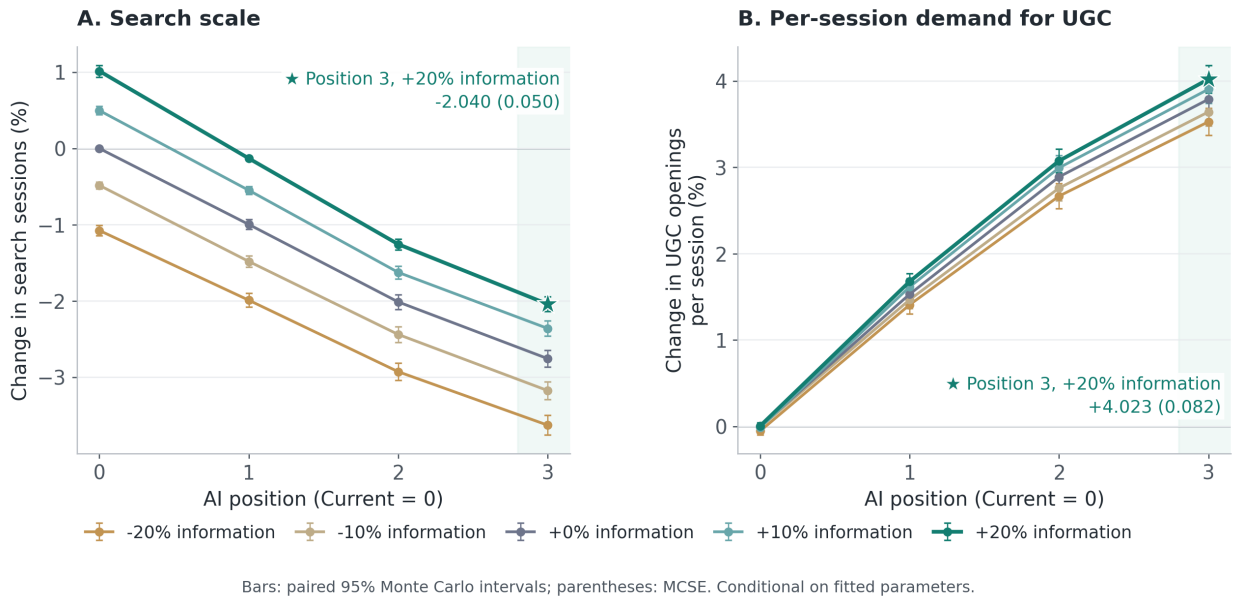


Figure 11: Placement and Information Amount Change Search Scale and UGC Intensity

Notes: Effects are relative to Current. Positions are zero-based display ranks and multipliers change semantic information at fixed expected reading length. Availability and the common UGC environment are fixed. Error bars are 95% paired Monte Carlo intervals conditional on the estimated parameters; annotated effects have Monte Carlo standard errors in parentheses. Stars mark the design selected for total direct UGC openings per user, which also has the largest point estimate in the evaluated grid. They are not significance markers or a claim of optimality outside that grid.

is position three with 20% more information. Sessions fall by 2.0397%, while direct UGC openings per session rise by 4.0227% and total direct openings rise by 1.9018%. Within the evaluated grid, this design has the largest point estimate of total direct UGC demand and serves the aggregate-demand objective while retaining a scale–intensity trade-off.

8.3 Allocating AI Opportunities Across Query Intents

The platform can also change which queries receive AI while preserving the current answer and its display position zero. We evaluate reallocations of intent-specific availability under the model's full Current budget. Within each intent, answer information, reference-count profiles, expected reading length, and UGC supply remain fixed.

Let w_k be intent k's share in a fixed reference distribution of Current queries and let λ_k^0 be its Current availability probability. The feasible policies satisfy

$$0 \le \lambda_k \le 1, \qquad \sum_{k=1}^{7} w_k \lambda_k = \sum_{k=1}^{7} w_k \lambda_k^0. \tag{35}$$

The right-hand side is the model's Current availability averaged across intents using the fixed reference query mix. It is a model-based availability budget, not an observed AI-preview or opening rate. Every rule retains

Table 19: Balanced Intent Allocation Under the Current AI Budget

Outcome	Balance	Current	Difference (MC SE)
Search sessions/user	3.2034	3.2034	0.0000 (0.0000)
Active search days/user	1.2001	1.2001	0.0000 (0.0000)
Direct UGC openings/session	1.6402	1.6402	0.0000 (0.0000)
Total direct UGC openings/user	5.2545	5.2545	0.0000 (0.0000)
AI openings/user	0.3006	0.3006	0.0000 (0.0000)
Sessions with downstream openings/user	1.3172	1.3172	0.0000 (0.0000)
Downstream UGC openings/user	14.7836	14.7836	0.0000 (0.0000)

Notes: Balance denotes the balanced intent allocation. Difference is Balance minus Current, in the row's units, not a percentage or percentage-point change; contrasts are calculated before rounding. Per-user outcomes cover 16 days and include inactive users. Per-session intensity averages replication-specific ratios. Parentheses report paired Monte Carlo standard errors from 100 replications of 1,000 users at fixed parameters. Policies retain Current placement, answer characteristics, and its full weighted availability budget.

this full budget; none supplies AI to every query. Fixed weights require increases in one intent to be offset elsewhere. Realized AI exposures and openings remain behavioral outcomes, and the realized query mix can change as users respond to the policy.

Joint improvement in search activity and per-session demand. We select a *balanced allocation* that maximizes the smaller proportional gain in sessions and direct UGC openings per session in the searched policy class. Design-sample selection precedes the independent main policy evaluation; no separate holdout validation is used. Table 19 compares this rule with Current. Table 20 provides the exact availability probabilities. Appendix G.4 describes the selection procedure.

The balanced search and both frontier objectives select Current itself, producing exactly zero differences under the paired simulation. The aggregate-demand objective selects a different allocation, as discussed below.

Aggregate demand when joint improvement is not established. The separately selected total-direct-UGC allocation raises sessions by 0.6908% and total direct UGC openings by 0.5925%, while direct UGC openings per session change by -0.0967%. The paired Monte Carlo interval for the latter is $[-0.2253\%, 0.0318\%]$, so the simulation does not resolve either an increase or a decrease in per-session demand. The AI-opening-maximizing allocation raises AI openings by 5.2737%, sessions by 0.6275%, and total direct UGC openings by 0.4176%, while reducing direct UGC openings per session by 0.2099%. Maximizing AI engagement and maximizing aggregate demand for creator content remain distinct objectives. All rules are fixed from design calculations before the main simulation.

Table 20: Availability Probabilities Under the Current Budget

Intent	Weight	Current	Balanced	AI-opening maximum	Total UGC maximum
Factual / Explanatory	21.5541	20.2604	20.2604	0.0000	6.8485
Procedural / How-to	6.3026	23.5802	23.5802	20.4685	35.5743
Evaluation / Recommendation	7.4185	11.3177	11.3177	64.4597	49.5446
Experience / Opinion	6.3022	13.9976	13.9976	66.2361	45.1865
Exploratory / Open-ended	1.3078	6.8151	6.8151	4.8363	89.4450
Current / Local / Navigational	21.7052	9.6539	9.6539	5.0753	0.0000
Ambiguous / Other / Unclassified	35.4096	4.6647	4.6647	0.0000	0.0000

Notes: Entries are percentages. Weights describe the frozen Current query distribution. Each policy retains the model's Current weighted availability budget before rounding, display position zero, and the same conditional answer distribution. Balanced selects Current itself. These probabilities are fixed policy choices, not estimated treatment effects.

8.4 Disclosing AI Information Amount Before Opening

Placement changes when AI is encountered, while allocation changes which queries receive it. Information disclosure instead helps users judge whether an available answer is worth reading. We ask whether this improves search value without changing the answer itself. Relative to Current, we hold actual information, expected reading length, availability, display position zero, and UGC supply fixed.

A common three-label design classifies answers as low, middle, or high in query-relevant semantic units. A second design chooses thresholds by observable intent from a finite candidate family. Full disclosure reveals the exact information amount as an ideal benchmark. Users know the disclosure rule and observe the label at the AI preview. Labels restrict the user's information belief to amounts consistent with the disclosed category. They supply no answer content and incur no additional preview cost. Expected reading length remains unchanged: this comparison isolates learning about information amount, rather than jointly revising beliefs about all answer characteristics.

Joint improvement in search activity and per-session demand. Common labels raise sessions by 0.1626%, while intent-specific labels and full disclosure raise sessions by 0.1839% and 0.2544%, respectively (Figure 12). Direct UGC openings per session change by 0.0285% under common labels, 0.0238% under intent-specific labels, and 0.0303% under full disclosure. All three paired Monte Carlo intervals for that margin include zero. None therefore establishes an increase in per-session demand or a joint improvement in both margins relative to Current.

Aggregate demand for creator content. Common labels raise total direct UGC openings by 0.1903%, while reducing AI openings by 1.6470%. Intent-specific labels raise total direct UGC openings by 0.2073%, and full disclosure yields the largest aggregate gain among the three fixed disclosure designs, 0.2838%. Full disclosure provides an ideal benchmark. Label thresholds are selected before the main simulation and evaluated against both design objectives. Users can value search more while opening AI less often: better

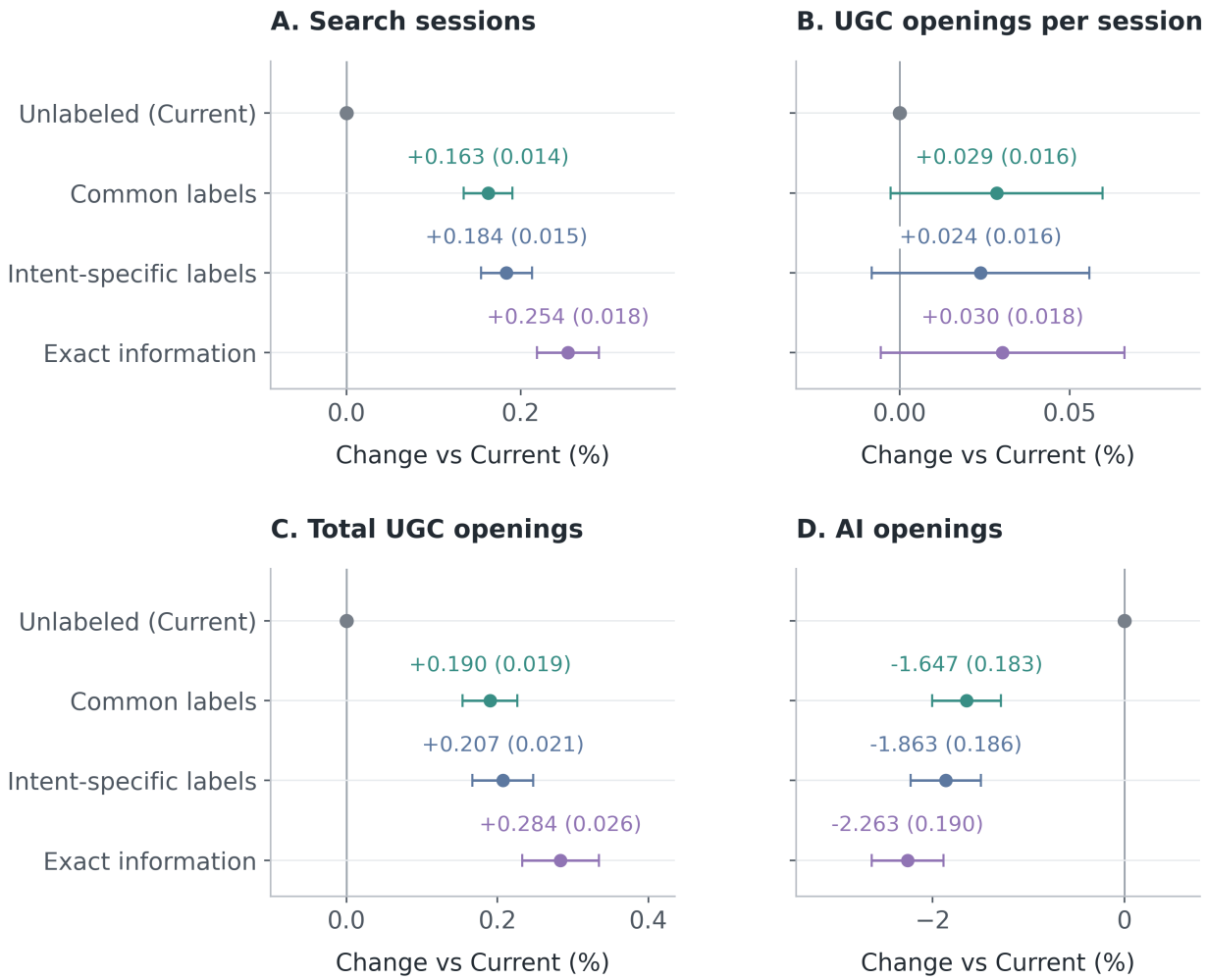


Figure 12: Information Labels Increase Search With Fewer AI Openings

Notes: Points are percentage changes relative to Current without disclosure; parentheses give paired Monte Carlo standard errors of those changes. Bars are 95% Monte Carlo intervals (estimate ±1.96 MC SE) from 100 paired replications at fitted parameters. They measure simulation precision, not statistical inference for estimated policy effects. Panel B measures direct UGC openings per session, averaging each replication's ratio before comparing policy means. Panels A, C, and D compare per-user outcomes over 16 days. The answer, expected reading length, availability, and display position zero are unchanged. Information enters the acquired stock only after opening.

selection into reading can expand simulated search and aggregate demand for UGC without a resolved change in demand per session.

8.5 Interpretation and Scope

The two design objectives separate joint improvement from aggregate demand gains. Delayed higher-information answers increase simulated aggregate demand while retaining a scale–intensity trade-off. The total-direct-UGC allocation and information labels increase simulated search activity and aggregate demand, but their per-session effects are not resolved from zero by the Monte Carlo intervals. The balanced allocation retains Current. These results do not establish joint improvement in both margins, but neither do they establish

that Current is optimal. The aggregate-demand objective can favor a different design despite a trade-off between search activity and demand per session.

These are short-run predictions under fixed creator-content supply and stable expectations within each policy. Additional AI opportunities inherit the calibrated answer distribution of their intent. Increasing information at fixed length is a hypothetical improvement in the answer; disclosure changes what users know about an unchanged answer. Creator entry, content production, and platform costs are outside the model.

9 Discussion and Conclusion

9.1 Main Findings and Business Implications

Our experiment answers RQ1 by showing that access to Generative AI Search increases how often and how broadly users search. Users also search over a longer span and more often use search in both halves of the experiment. Most additional same-day queries are classified as related follow-ups or distinct information tasks. Together with the wider topical coverage, this composition supports continued and exploratory information seeking.

For RQ2, users view fewer UGC previews and open fewer pages per search. They also open a smaller fraction of previewed pages, and a larger share of the average user's searches contains previews but no opening. Clicking and CTR decline significantly for factual, procedural, and evaluation/recommendation queries, whose information needs AI answers can more readily satisfy, but not for experience/opinion queries. Together, the preview-to-opening outcomes and intent differences support information substitution: AI answers can provide information that users would otherwise seek by opening creator content.

Total demand nevertheless rises because users search more. Both UGC previews and openings increase over the experiment, and more sessions lead to additional creator content beyond the initial search result. The CTR decline is larger among the first five results, which the platform prioritizes using relevance and quality. Yet aggregate opening point estimates are positive and similar in absolute size for the Top 5 and Non-top 5, with the latter significant only at the 10% level. Aggregate gains therefore extend beyond the first five results; the decline in per-session consumption is statistically clear only for the first five.

For RQ3, the model links these two responses and evaluates alternative designs. AI information reduces the benefit of further UGC within a session, while higher prospective search value encourages daily participation and additional queries. This connection allows design choices to affect both consumption within a search and how often users return to search.

These findings have three implications for platforms. First, the scale–intensity trade-off calls for tracking search sessions per user, UGC demand per session, and aggregate UGC demand per user. Fewer openings per session can coexist with greater total demand when users search more. AI engagement should be tracked separately: more AI openings need not indicate better support for creator content.

Second, ranking heterogeneity matters because the ranking reflects the platform's assessments of relevance and content quality. Platforms should track demand for the UGC they prioritize as well as discovery beyond the first five results. The sharper Top 5 CTR decline indicates that reduced opening after a preview is not confined to later-ranked content. Yet aggregate opening gains are similar in absolute size across the two ranges. Within-search responses and aggregate demand therefore need to be assessed separately for each range.

Third, the three counterfactual design families illustrate ways to reshape the scale–intensity trade-off. The two design objectives distinguish improving both search activity and demand per session from increasing aggregate direct UGC demand when joint improvement is not established. Moving a more informative AI answer lower in the results raises simulated UGC demand per session but reduces search activity; aggregate direct openings can nevertheless rise. Reallocating AI availability across query intents can produce a modest aggregate gain under the same AI-availability budget, without changing placement or answer characteristics. Truthful information labels can also increase search activity and aggregate direct UGC openings even as AI openings fall. For these allocation and disclosure designs, however, the effects on per-session demand are not resolved from zero by the Monte Carlo intervals. Unlike the placement result, these aggregate gains establish neither a scale–intensity trade-off nor a joint improvement. The absence of an established joint improvement does not show that the existing design is optimal. Together, the comparisons favor evaluating design choices by their effects on search and demand for creator content, not only engagement with the platform's own answer.

9.2 Generalizability.

The behavioral channels examined in our study are most directly relevant to other content creation platforms where AI-generated responses and creator-supplied content address overlapping information needs. Reddit and Stack Overflow provide comparable settings, combining AI-assisted information access with community-created content.[2] In these settings, AI may reduce the need to open additional UGC within a session while making search more valuable and encouraging more frequent use. Our experiment covers a range of search needs: query counts increase across all six focal intent categories, including categories with relatively low AI-preview exposure. This breadth suggests that search expansion is not confined to a single type of information request.

For general-purpose search engines, the comparison is most relevant to queries seeking explanations, how-to guidance, evaluations, recommendations, or experience-based advice. For these searches, an AI response and the underlying sources can address the same information need. The connection is less direct for searches whose primary purpose is to reach a particular website or complete a transaction. The scale–intensity

[2]See `https://redditinc.com/news/introducing-reddit-answers` and `https://stackoverflow.co/company/press/archive/stack-overflow-ai-assist`.

framework therefore provides a useful basis for evaluating AI search in comparable settings. Whether greater search activity offsets lower demand for underlying content per session depends on the mix of search needs and the interface, including how much AI information is visible before clicking and how easily users can access the original sources.

9.3 Limitations and Future Research

First, the counterfactual analysis does not model how users adapt to new interface designs. Our removal experiment examines continued access to an established feature. The counterfactuals, however, consider changes to placement, availability, and information disclosure under stable expectations. Information disclosure also assumes that users understand and trust the labels without incurring additional previewing effort. Future experiments could examine how responses to redesigned interfaces evolve as users learn their rules, interpret their labels, and revise their expectations.

Second, the analysis captures demand responses rather than the full response of the creator ecosystem. We do not estimate creator entry, content production, earnings, or the distribution of additional attention across creators. Holding UGC supply fixed allows us to isolate how AI search changes demand for existing creator content, but excludes feedback from demand to future content supply. Linking these demand responses to creator behavior, monetization, and AI-production costs would clarify the longer-run implications for creator incentives and platform returns. Evidence from other content formats and platforms would also help establish how these responses depend on the setting.

Third, the reduced-form preview-to-opening comparisons reflect responses to AI information together with changes in position, attention, and the composition of observed searches. They provide evidence on behavioral channels rather than separately identified mediation effects. The observed behaviors do not directly measure whether users satisfy their information needs. The logs record search and content openings, not task completion, comprehension, or satisfaction. Opening a page does not establish full reading. Similarly, semantic units measure query-relevant content expressed in an answer, not its independently verified accuracy or the information a user understands. More search therefore need not imply new information needs or higher welfare. Future research could combine behavioral records with task performance, independent accuracy assessments, and user-reported experiences to distinguish greater engagement from better outcomes for users.

References

Agarwal, S. and Sen, A. (2026). The impact of Google AI Overviews on publisher traffic and user experience: Evidence from a field experiment. *SSRN Working Paper 6513059*.

Amaldoss, W. and Shin, W. (2026). Content provision on UGC platforms. *Management Science*, 72(2):1489–1508.

Armstrong, M. (2006). Competition in two-sided markets. *The RAND journal of economics*, 37(3):668–691.

Athey, S., Mobius, M., and Pal, J. (2021). The impact of aggregators on Internet news consumption. NBER Working Paper 28746, National Bureau of Economic Research.

Athukorala, K., Głowacka, D., Jacucci, G., Oulasvirta, A., and Vreeken, J. (2016). Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, 67(11):2635–2651.

Bai, B., Zeng, Z., Zhang, D., and Xu, Z. (2026). The impact of generative AI search on content-sharing platforms. Available at SSRN 6171026.

Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum*, 36(2):3–10.

Brown, C., Snodgrass, T., Kemper, S. J., Herman, R., and Covington, M. A. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, 40(2):540–545.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Calzada, J. and Gil, R. (2020). What do news aggregators do? evidence from Google news in Spain and Germany. *Marketing Science*, 39(1):134–167.

Cao, Z., Zhu, Y., Li, G., and Qiu, L. (2024). Consequences of information feed integration on user engagement and contribution: A natural experiment in an online knowledge-sharing community. *Information Systems Research*, 35(3):1114–1136.

Chan, T. Y., Liao, L., Martin, X., and Wang, Z. (2024). Avoiding peer information and its effects on charity crowdfunding: a field experiment. *Management science*, 70(4):2272–2293.

Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, 139(2):891–936.

Chen, Y. and Yao, S. (2017). Sequential search with refinement: Model and application with click-stream data. *Management Science*, 63(12):4345–4365.

De los Santos, B. and Koulayev, S. (2017). Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science*, 36(4):542–564.

Foerderer, J., Kude, T., Mithas, S., and Heinzl, A. (2018). Does platform owner's entry crowd out innovation? evidence from Google Photos. *Information Systems Research*, 29(2):444–460.

Gouriéroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to Poisson models. *Econometrica*, 52(3):701–720.

Jansen, B. J., Booth, D. L., and Spink, A. (2008). Determining the informational, navigational, and transactional intent of web queries. *Information Processing & Management*, 44(3):1251–1266.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101(6):2590–2615.

Kim, J. B., Albuquerque, P., and Bronnenberg, B. J. (2010). Online demand under limited consumer search. *Marketing Science*, 29(6):1001–1023.

Lee, H. A., Lee, D., Lee, G., and Lee, D. (2026). Breaking news out of the filter bubble: Generative AI Search diversifies collective attention and raises shared information consumption. Working paper, arXiv:2609.38946.

Lei, X., Chen, Y., and Sen, A. (2025). Trade-offs in leveraging external data capabilities: Evidence from a field experiment in an online search market. *Management Science*, 72(4):2998–3015.

Lex, E., Voelske, M., Errecalde, M., Ferretti, E., Cagnina, L., Horn, C., Stein, B., and Granitzer, M. (2012). Measuring the quality of web content using factual information. In *Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality*, pages 7–10. Association for Computing Machinery.

Liu, J. and Cong, Z. (2023). The daily me versus the daily others: how do recommendation algorithms change user interests? evidence from a knowledge-sharing platform. *Journal of Marketing Research*, 60(4):767–791.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons.

Morris, M. R., Teevan, J., and Panovich, K. (2010). What do people ask their social networks, and why? a survey study of status message q&a behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1739–1748. Association for Computing Machinery.

Nenkova, A. and Passonneau, R. (2004). Evaluating content selection in summarization: The pyramid method. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, pages 145–152. Association for Computational Linguistics.

Parker, G. G. and Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management Science*, 51(10):1494–1504.

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of theoretical biology*, 13:131–144.

Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, 1(4):990–1029.

Rose, D. E. and Levinson, D. (2004). Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, pages 13–19.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 55(5):999–1033.

Santos Silva, J. M. C. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Santos Silva, J. M. C. and Tenreyro, S. (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters*, 112(2):220–222.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Song, Y., Shi, X., and Fu, X. (2013). Evaluating and predicting user engagement change with degraded search relevance. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1213–1223. Association for Computing Machinery.

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.

Wang, J., Li, G., and Hui, K.-L. (2022). Monetary incentives and knowledge spillover: Evidence from a natural experiment. *Management Science*, 68(5):3549–3572.

Wang, S. T., Gleason, J., Bart, Y., Wilson, C., and Metaxa, D. (2026). AI in search reduces publisher referrals without improving user experience: Experimental evidence. arXiv:2608.18352, version 1, August 18.

Zheng, S., Zhu, Y., Ye, X., and Shen, L. (2025). Generative search: Evidence from a large-scale field experiment. *Available at SSRN 5297194*.

Zhu, F. and Liu, Q. (2018). Competing with complementors: An empirical look at amazon.com. *Strategic Management Journal*, 39(10):2618–2642.

A Experimental Measurement and Sample Description

A.1 Text Processing and Semantic Topics

We construct a common topic system for treatment and control. Query text first receives Unicode NFKC normalization, which standardizes equivalent character forms. We convert text to lowercase, remove leading and trailing spaces, and collapse repeated spaces. We then pool distinct normalized queries across users, dates, and experimental groups.

Each distinct query receives an embedding, a numerical representation of its meaning, from `paraphrase-multilingual-MiniLM-L12-v2`. We scale these representations to unit length and group them into 300 topics using spherical k-means. This method groups texts by the direction of their embeddings, so queries with similar meanings tend to share a topic. Each distinct text has equal weight when forming the groups. We then assign every occurrence of that text to its topic. These assignments remain fixed when calculating topic counts and entropy for the full experiment or either 13-day period.

Topic counts measure the range of subjects searched. Entropy measures how evenly a user's queries are spread across those subjects. Let K_i be the number of semantic topics represented in user i's queries, and let p_{ik} be the fraction of that user's query occurrences in topic k. For $K_i \geq 2$, normalized topic entropy is

$$H_i = -\frac{\sum_{k=1}^{K_i} p_{ik} \log p_{ik}}{\log K_i}. \tag{36}$$

Here, H_i ranges from zero to one. A higher value indicates a more even distribution across the topics that user searches; it equals one when all represented topics receive the same share of queries. We set $H_i = 0$ when $K_i = 1$ and leave it undefined when $K_i = 0$. The measure adapts Shannon entropy and Pielou's evenness index (Shannon, 1948; Pielou, 1966).

A.2 Additional Descriptive Statistics

Table 4 in the main text reports the core user-level summary statistics and their denominators. Table A1 adds the descriptive breakdown for the six substantive research intents. The seventh category combines ambiguous, other, and unclassified query texts. Total search counts and the structural intent system retain that category alongside the six focal intents.

Table A1: Additional Descriptive Statistics by Focal Search Intent

Outcome	Mean	SD	Median
Factual/explanatory	1.0735	4.8649	0
Procedural/how-to	.3034	1.3355	0
Evaluation/recommendation	.3727	1.9272	0
Experience/opinion	.3123	1.5717	0
Exploratory/open-ended	.0673	.5506	0
Current/local/navigational	1.0565	3.3965	0

Notes: Each row is the number of queries of the named intent per randomized user over March 30–April 24, 2025. All 100,000 randomized accounts enter, with zero for no queries of that type. SD is the cross-user standard deviation. These statistics cover the 26-day experimental window.

Table A2: Research Search-Intent Taxonomy

Category	Primary request and examples
Factual/explanatory	Facts, definitions, concepts, mechanisms, causes, and objective explanations; e.g., "What is marginal cost?"
Procedural/how-to	Steps, implementation, configuration, or troubleshooting; e.g., "How do I convert a PDF to Word?"
Evaluation/recommendation	Comparison or advice to support a choice; e.g., "Which phone should I buy?"
Experience/opinion	Personal experience, reviews, reactions, or subjective views; e.g., "What is it like to work at this company?"
Exploratory/open-ended	Broad understanding, possibilities, or multiple perspectives without a specific factual, procedural, or decision request; e.g., "How will AI change society?"
Current/local/navigational	Timely or location-specific information, a particular person/page/item, or content retrieval; e.g., "Today's gold price" or a named Zhihu page.
Ambiguous/other	Text insufficient to identify a reliable primary request, including isolated names and incomplete fragments.

B Search Intent Classification and Validation

B.1 Taxonomy and Implementation

The taxonomy draws on research on web-search goals, procedural search, decision support, social question asking, and exploration (Broder, 2002; Rose and Levinson, 2004; Jansen et al., 2008; Marchionini, 2006; Athukorala et al., 2016; Morris et al., 2010). Table A2 states the primary distinctions. The shortened category names in the main text retain these definitions. In particular, current/local/navigational includes requests for specific content and entertainment-oriented searches.

We classify each distinct normalized query once using GPT-4.1-mini with deterministic decoding. Blank entries and standard missing-value markers are treated as missing text. The classifier receives the query text, prespecified category definitions, and rules for overlapping requests. Its task is to label the request, using only that text. It receives no treatment assignment, AI eligibility, display information, user characteristics, or behavior, and uses no external search.

The prompt prioritizes explicit experience/opinion language over recommendations, explicit decisions over general exploration, and explicit steps or troubleshooting over factual explanation. Current/local information and specific-target retrieval are assigned to that category. Broad requests not covered by these rules are exploratory. Every valid query receives one primary label. Secondary labels, a mixed-intent indicator, confidence scores, and short explanations support classification checks. Only the primary label enters the analysis. We verify that each classification matches its input query, uses a valid category, and is internally consistent. Unmatched or invalid labels enter ambiguous/other, preserving the seven-category system.

B.2 Independent Human Validation

Three human annotators, all Zhihu users, independently classify a random sample of 1,000 distinct queries with valid analysis labels. They receive the same taxonomy and literal-text rule. They do not see the analysis label, one another's answers, treatment status, AI eligibility, or any behavior. A majority label exists if at least two agree. There are 967 such queries; 33 receive three different labels.

Table A3: Independent Annotation Reliability

Measure	Estimate	95% bootstrap interval
Unanimous agreement	.658	–
All-different tie rate	.033	–
Fleiss' kappa	.691	–
Krippendorff's alpha	.691	–
Pair 1–2: agreement / kappa	.774 / .712	–
Pair 1–3: agreement / kappa	.763 / .693	–
Pair 2–3: agreement / kappa	.746 / .673	–
Human-majority label versus analysis label		
Exact agreement	.776	[.749, .804]
Cohen's kappa	.710	[.675, .746]
Balanced accuracy / macro recall	.677	[.636, .729]
Macro precision	.701	–
Macro F1	.684	[.650, .720]
Weighted F1	.777	–
Unanimous-only agreement / kappa	.868 / .830	–

Notes: Majority comparisons use 967 queries; the unanimous comparison uses 658. Confidence intervals use 1,000 bootstrap resamples. Reported classification metrics treat the analysis label as the reference, not as established ground truth. Kappa and alpha adjust agreement for chance. Precision and recall compare the human-majority and analysis labels; F1 is their harmonic mean. Macro metrics weight categories equally, and weighted F1 weights them by their reference counts. Validation assesses the reproducibility of the labels.

Table A4: Class-Specific Agreement with the Analysis Labels

Intent	Count	Precision	Recall	F1
Factual/explanatory	251	.853	.765	.807
Procedural/how-to	76	.827	.882	.854
Evaluation/recommendation	77	.903	.727	.806
Experience/opinion	60	.764	.700	.730
Exploratory/open-ended	8	.067	.125	.087
Current/local/navigational	199	.759	.663	.708
Ambiguous/other	296	.732	.878	.799

Agreement varies across categories. Table A4 shows particularly low agreement for exploratory/open-ended queries, with only eight reference observations. We therefore interpret that category's experimental result cautiously. The more common categories have higher agreement and provide a firmer basis for describing query patterns.

B.3 Intent-Specific Experimental Results

Table A5 reports OLS estimates underlying Figure 4. Table A6 describes recorded AI-preview exposure in treatment sessions. The experimental estimates use all randomized accounts. Exposure rates describe treatment-group queries and combine eligibility with users' reach of the AI preview. Research intent and platform eligibility are separate classifications.

Table A5: Effects on Query Counts by Intent

Intent	No controls	With controls
Factual/explanatory	.1245***	0.1251***
	(.0308)	(0.0304)
Procedural/how-to	.0222***	0.0225***
	(.0085)	(0.0084)
Evaluation/recommendation	.0354***	0.0356***
	(.0122)	(0.0121)
Experience/opinion	.0223**	0.0232**
	(.0100)	(0.0099)
Exploratory/open-ended	.0106***	0.0107***
	(.0035)	(0.0035)
Current/local/navigational	.0552**	0.0579***
	(.0215)	(0.0212)

Notes: Each regression includes 100,000 randomized users. Controls and heteroskedasticity-robust standard errors follow Equation (1). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A6: Recorded AI-Preview Exposure in Treatment Sessions

Intent	Sessions	Sessions with an observed AI preview	AI-preview exposure rate
Factual/explanatory	56,788	17,299	30.46%
Procedural/how-to	15,722	5,388	34.27%
Evaluation/recommendation	19,519	3,533	18.10%
Experience/opinion	16,172	3,728	23.05%
Exploratory/open-ended	3,628	462	12.73%
Current/local/navigational	54,202	8,194	15.12%
Ambiguous/other/unclassified	87,104	6,756	7.76%
All treatment sessions	253,135	45,360	17.92%

Notes: AI-preview exposure rate is sessions with an observed AI preview divided by treatment sessions within the row. These are observed rates over the 26-day experiment. The structural analysis calibrates AI availability from observed preview occurrence and models users' continuation choices over its 16-day window.

C Search Sequences and Accounting Details

C.1 Adjacent-Query Classification

We compare consecutive queries from the same user on the same day. Sessions are ordered by their earliest valid search-result display time, with the original record order and session identifier breaking ties. Each session after the day's first is paired with the immediately preceding session. The time between their first result displays is Δt, measured in minutes.

We compare query text in two ways: overlap in its written characters and similarity in meaning. Text receives the normalization described in Appendix A. A compact version also removes spaces and punctuation while retaining letters, digits, and Chinese characters. For query q, let $G(q)$ be the set of adjacent two-character combinations in this compact text. For two queries, q and q', their character overlap is measured by the Jaccard similarity

$$J(q, q') = \frac{|G(q) \cap G(q')|}{|G(q) \cup G(q')|}.$$

Table A7: Components of the Estimated Increase in Search Sessions

Outcome	No controls	With controls	Share
A. All sessions			
Total	.3666***	0.3779***	100.0%
	(.0955)	(0.0939)	
First session during experiment	.0080**	0.0059**	2.2%
	(.0032)	(0.0028)	
First sessions on additional days	.0621***	0.0694***	16.9%
	(.0200)	(0.0193)	
Later same-day sessions	.2965***	0.3025***	80.9%
	(.0802)	(0.0794)	
B. Later same-day sessions			
Total	.2965***	0.3025***	100.0%
	(.0802)	(0.0794)	
Reformulation/repair	.0533***	0.0548***	18.0%
	(.0159)	(0.0157)	
Related follow-up	.0438***	0.0448***	14.8%
	(.0126)	(0.0125)	
Distinct information task	.1994***	0.2030***	67.3%
	(.0577)	(0.0572)	

Notes: All rows use the same 100,000 randomized accounts. Shares divide the no-controls estimate by the panel's total, describing the composition of estimated growth. The unclassified component remains in the accounting but rounds to zero at the reported precision. Rounding can prevent displayed sums from matching exactly. Controls and standard errors follow Equation (1). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Here, the numerator counts character pairs shared by the two queries, and the denominator counts pairs appearing in either query. The score ranges from zero for no overlap to one for identical sets. A one-character query is treated as a single token. Semantic similarity, ρ, is the cosine similarity between the unit-normalized multilingual query embeddings. A higher ρ indicates greater similarity in meaning. A pair with missing text or required inputs is unclassified.

We apply the following rules in order. *Reformulation/repair* requires $\Delta t \leq 30$ plus at least one of the following: an exact normalized-text match, a compact-text match, $\rho \geq .92$, or both $\rho \geq .85$ and $J \geq .30$. *Related follow-up* requires $\Delta t \leq 30$ and $\rho \geq .70$ without meeting the repair rule. *Distinct information task* requires $\Delta t > 30$ or $\rho < .70$, with valid inputs. The same thresholds apply to both groups. Similar queries more than 30 minutes apart are therefore classified as distinct. These are classifications of text and timing, not observations of task success or private information needs.

C.2 Sensitivity to Adjacent-Query Classification Rules

Table A8 crosses 20-, 30-, and 60-minute windows with lenient, baseline, and strict semantic thresholds. The estimated non-repair share of additional same-day sessions ranges from 80.7% to 84.3%, compared with 82.0% under the baseline rule. Controlling for observed user characteristics yields a similar range, 80.6%–84.2%. Repair, follow-up, and distinct-task coefficients remain positive in every specification. Thus, the finding that most additional same-day queries fall outside repair is stable across these thresholds.

Table A8: Sensitivity of the Adjacent-Query Decomposition

Window	Semantic rule	Repair	Follow-up	Distinct task	Non-repair share
20 min	Lenient	0.0570	0.0519	0.1876	80.8%
		(0.0168)	(0.0138)	(0.0557)	
20 min	Baseline	0.0528	0.0425	0.2011	82.2%
		(0.0155)	(0.0122)	(0.0583)	
20 min	Strict	0.0465	0.0364	0.2136	84.3%
		(0.0142)	(0.0106)	(0.0611)	
30 min	Lenient	0.0572	0.0539	0.1854	80.7%
		(0.0171)	(0.0142)	(0.0550)	
30 min	Baseline	0.0533	0.0438	0.1994	82.0%
		(0.0159)	(0.0126)	(0.0577)	
30 min	Strict	0.0468	0.0378	0.2119	84.2%
		(0.0146)	(0.0108)	(0.0606)	
60 min	Lenient	0.0570	0.0562	0.1833	80.8%
		(0.0177)	(0.0147)	(0.0540)	
60 min	Baseline	0.0530	0.0451	0.1983	82.1%
		(0.0165)	(0.0129)	(0.0569)	
60 min	Strict	0.0467	0.0382	0.2115	84.2%
		(0.0151)	(0.0111)	(0.0599)	

Notes: Entries are treatment–control differences in later same-day sessions per randomized user over the experiment. Parentheses contain user-level heteroskedasticity-robust standard errors. All specifications include users with no searches and use the same adjacent-query similarities. The lenient, baseline, and strict rules use cosine cutoffs of (.65, .82, .90), (.70, .85, .92), and (.75, .88, .94) for relatedness, repair supported by character overlap, and near duplication, respectively. The character-overlap cutoff remains .30. Exact or compact-text matches remain repair within the applicable time window. The non-repair share is one minus the repair coefficient divided by the coefficient for all later same-day sessions. The single unclassified transition remains in the accounting; its treatment coefficient is .00002. The total later-same-day coefficient is .2965 (.0802) in every specification.

C.3 UGC Interactions Before a Later Query

Supplementary outcomes partition later searches by the preceding session's direct UGC openings. Table A9 uses five-minute, 30-minute, and same-day windows. These are windows between sessions, not alternative definitions of the sessions themselves. Counts with zero, one, and at least two preceding UGC openings add to the corresponding total. The table also partitions later queries by preceding UGC preview views. Most estimated growth occurs after sessions with zero or one direct opening. This partition describes treatment effects on search paths. Because preceding UGC interactions can also respond to treatment, these partitions do not identify their causal effects on further search.

C.4 Accounting for Aggregate Demand for UGC

Group means and plug-in estimates. For each group $z \in \{C, T\}$, let n_z denote the number of users, q_i user i's total search sessions, and y_i that user's total count for the UGC outcome being decomposed. Both counts cover the 26-day experiment, and users with no searches contribute zeros. For distinct direct UGC

Table A9: Subsequent Queries by Prior-Session UGC Interactions

Transition / preceding UGC interactions	No controls	With controls	Share
Within 5 minutes: total	.1578*** (.0435)	0.1619*** (0.0431)	100.0%
0 direct UGC openings	.0953*** (.0198)	0.0976*** (0.0196)	60.4%
1 direct UGC opening	.0449** (.0185)	0.0459** (0.0184)	28.5%
2+ direct UGC openings	.0175 (.0130)	0.0183 (0.0129)	11.1%
Within 30 minutes: total	.2079*** (.0575)	0.2127*** (0.0570)	100.0%
0 direct UGC openings	.1094*** (.0220)	0.1119*** (0.0218)	52.6%
1 direct UGC opening	.0635*** (.0227)	0.0648*** (0.0226)	30.5%
2+ direct UGC openings	.0350 (.0216)	0.0360* (0.0214)	16.8%
Same day: total	.2965*** (.0802)	0.3025*** (0.0794)	100.0%
0 direct UGC openings	.1342*** (.0260)	0.1372*** (0.0257)	45.3%
1 direct UGC opening	.0973*** (.0330)	0.0993*** (0.0328)	32.8%
2+ direct UGC openings	.0649** (.0325)	0.0660** (0.0322)	21.9%
Within 30 minutes, by preview views			
0 UGC previews	.0056*** (.0004)	0.0056*** (0.0004)	2.7%
1 UGC preview	.0024 (.0037)	0.0027 (0.0037)	1.2%
2+ UGC previews	.1998*** (.0553)	0.2044*** (0.0548)	96.1%

Notes: User-level OLS, $N = 100,000$. Parentheses are heteroskedasticity-robust standard errors. Shares use no-controls estimates and sum subject to rounding. Each partition by preceding UGC interactions is mutually exclusive within its window. These are treatment effects on path counts, not causal effects of prior UGC interactions. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

openings, we count each opened UGC item once within a session and then sum across sessions. The inputs to the decomposition are

$$S_z = \frac{1}{n_z}\sum_{i\in z} q_i, \qquad \overline{Y}_z = \frac{1}{n_z}\sum_{i\in z} y_i, \qquad M_z = \frac{\sum_{i\in z} y_i}{\sum_{i\in z} q_i} = \frac{\overline{Y}_z}{S_z}.$$

Thus, M_z is the pooled count of that UGC outcome per session. The per-session regressions instead use each searching user's own outcome-to-session ratio. The two calculations weight users differently.

Substituting these sample quantities into Equation (3) gives the plug-in estimates

$$\widehat{A} = (S_T - S_C)\frac{M_T + M_C}{2}, \qquad \widehat{I} = (M_T - M_C)\frac{S_T + S_C}{2}, \qquad \widehat{\Delta} = \overline{Y}_T - \overline{Y}_C = \widehat{A} + \widehat{I}.$$

Here $\widehat{A}$ is the search-activity component, $\widehat{I}$ the per-session component, and $\widehat{\Delta}$ the total difference. All three are in units of the corresponding UGC outcome per randomized user. Although their computed values are fixed in the observed sample, they would vary across samples. Their standard errors measure this sampling uncertainty.

User-level variation and standard errors. We first calculate the sample covariance matrix of session counts and the corresponding UGC outcome counts within each group:

$$\widehat{\Omega}_z = \frac{1}{n_z - 1}\sum_{i\in z}\begin{pmatrix} q_i - S_z \\ y_i - \overline{Y}_z \end{pmatrix}\begin{pmatrix} q_i - S_z & y_i - \overline{Y}_z \end{pmatrix}.$$

Table A10: Accounting Decomposition of Aggregate UGC Outcomes

Outcome	Total difference	Search-activity component	Per-session component
UGC previews viewed	3.3168***	3.6748***	−.3580
	(1.1240)	(.9578)	(.5049)
Distinct direct UGC openings	.4212**	.6148***	−.1936*
	(.2014)	(.1601)	(.1069)
Direct UGC opening events	.4240**	.6213***	−.1972*
	(.2030)	(.1618)	(.1076)
Sessions with downstream UGC openings	.1224***	.1476***	−.0252
	(.0469)	(.0385)	(.0216)
Downstream UGC openings	1.4895*	1.6978***	−.2083
	(.8208)	(.4433)	(.6290)

Notes: Sessions per user are 4.6961 in control and 5.0627 in treatment in this accounting sample; their difference is .3666 (SE .0955). Outcomes are in original units per randomized user. Components add exactly before rounding. They do not estimate separately randomized mechanisms. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Its diagonal entries are the sample variances of the two user-level counts; its off-diagonal entries are their sample covariance. Dividing this matrix by n_z gives the estimated covariance matrix of the two group means. Stacking the control and treatment means gives

$$\widehat{\theta} = (S_C, \overline{Y}_C, S_T, \overline{Y}_T)', \qquad \widehat{V}_\theta = \begin{pmatrix} \widehat{\Omega}_C/n_C & 0_{2\times 2} \\ 0_{2\times 2} & \widehat{\Omega}_T/n_T \end{pmatrix}.$$

The zero blocks treat the two groups as independent samples. Aggregating at the user level retains the dependence among a user's search sessions and between that user's search activity and UGC outcome counts.

Let $F(\theta)$ be the function that calculates the search-activity component, per-session component, and total difference from these four means, using $M_z = \overline{Y}_z/S_z$, so that $\widehat{F} = F(\widehat{\theta}) = (\widehat{A}, \widehat{I}, \widehat{\Delta})'$. Let J be its derivative matrix, evaluated at $\widehat{\theta}$. Each row records how one estimate changes with the four input means. The delta method gives

$$\widehat{V}_F = J\widehat{V}_\theta J', \qquad \widehat{\mathrm{SE}}(\widehat{F}_j) = \sqrt{[\widehat{V}_F]_{jj}}, \quad j \in \{A, I, \Delta\}.$$

We evaluate the derivatives numerically. This calculation propagates uncertainty in both the group means and the pooled ratios, including their covariance, into each estimate. It also retains the covariance between the two components, which share the same inputs.

For distinct direct UGC openings, the estimates and standard errors are 0.6148 (0.1601) for the search-activity component, −0.1936 (0.1069) for the per-session component, and 0.4212 (0.2014) for the total difference. The 95% confidence intervals in Figure 6 equal each estimate plus or minus 1.96 times its standard error. Table A10 reports the same calculations for all UGC outcomes, with standard errors in parentheses.

D Experimental Robustness

D.1 Proportional Changes in Search and UGC Demand

The primary regression estimates changes in the number of searches or content interactions. As a robustness check, we use Poisson pseudo-maximum likelihood to estimate proportional changes in those counts and totals (Gouriéroux et al., 1984; Santos Silva and Tenreyro, 2006, 2011):

$$\mathrm{E}[Y_i \mid \mathrm{Treat}_i, X_i] = \exp(\alpha + \beta\,\mathrm{Treat}_i + X_i'\gamma). \tag{37}$$

The variables follow Equation (1): Y_i is the user's outcome, Treat_i indicates retained access, and X_i contains pre-treatment characteristics. The equation gives the expected outcome conditional on assignment and those characteristics. The exponential keeps this mean positive. The intercept α sets the baseline, γ contains the control coefficients, and β gives the proportional treatment change through $\exp(\beta) - 1$.

This specification retains users with zero outcomes. PPML estimates the conditional mean without requiring a Poisson outcome distribution. We apply it to counts and totals, using the same randomized-user sample as the primary regressions.

The unadjusted estimates imply approximately 7.8% more sessions, 6.9% more distinct queries, and 3.8% more active search days. Total UGC-preview views rise by about 7.0% and distinct direct UGC openings by 5.3%. The positive count effects appear in all six focal intents. The increase in total downstream openings is less precise and significant at 10%. The central findings are similar under the additive and proportional specifications.

D.2 The Two Halves of the Experiment

To examine adjustment after removal, we split the experiment into March 30–April 11 and April 12–24, two 13-day periods. Each account contributes one observation per period for counts and totals. Per-session outcomes are defined only in periods in which that user searches. We allow the treatment–control difference to vary between the two halves:

$$Y_{ip} = \alpha + \lambda\,\mathbf{1}\{p = 2\} + \beta_1 \mathrm{Treat}_i \mathbf{1}\{p = 1\} + \beta_2 \mathrm{Treat}_i \mathbf{1}\{p = 2\} + X_i'\gamma + \varepsilon_{ip}, \tag{38}$$

Here, $p \in \{1, 2\}$ denotes the period, and Y_{ip} is user i's outcome in that period. The indicator $\mathbf{1}\{p = j\}$ equals one in period j. The intercept α is the first-period baseline, and λ allows the baseline to change in the second period. The coefficients β_1 and β_2 are the period-specific treatment–control differences. Positive values indicate higher outcomes under retained access. Pre-treatment controls X_i, their coefficients γ, and the residual ε_{ip} follow the main specification. Standard errors are clustered by user. Counts, totals, and ratios are reconstructed within each period.

Nineteen of the 20 outcomes retain their sign across periods. Search scale, topical coverage, all six focal intent counts, and total direct UGC measures have positive estimates in both halves; per-session UGC estimates are negative. Entropy is close to zero and insignificant in both halves. Figure 7 reports the contrasts and confidence intervals, some including zero. Equality across periods is not formally tested. The comparison describes adjustment within the 26-day experiment, not persistence beyond it.

Table A11: PPML Robustness Estimates

Outcome	No controls	With controls
Search sessions	.0752*** (.0195)	0.0775*** (0.0191)
Distinct queries	.0667*** (.0188)	0.0692*** (0.0185)
Active search days	.0370*** (.0113)	0.0399*** (0.0108)
Unique platform categories	0.0317*** (0.0107)	0.0344*** (0.0100)
Unique semantic topics	0.0449*** (0.0143)	0.0476*** (0.0138)
Factual/explanatory queries	.1161*** (.0285)	0.1171*** (0.0281)
Procedural/how-to queries	.0733*** (.0278)	0.0746*** (0.0275)
Evaluation/recommendation queries	.0949*** (.0323)	0.0965*** (0.0320)
Experience/opinion queries	.0713** (.0317)	0.0744** (0.0315)
Exploratory/open-ended queries	.1581*** (.0519)	0.1598*** (0.0517)
Current/local/navigational queries	.0523** (.0203)	0.0549*** (0.0200)
Total UGC previews viewed	.0678*** (.0229)	0.0698*** (0.0225)
Total distinct direct UGC openings	.0515** (.0246)	0.0533** (0.0242)
Total direct UGC opening events	.0513** (.0245)	0.0532** (0.0241)
Sessions with downstream UGC openings	.0623*** (.0238)	0.0637*** (0.0233)
Downstream UGC openings	.0659* (.0363)	0.0674* (0.0359)

Notes: Each regression uses 100,000 randomized accounts over 26 days. Entries are log-link coefficients, not percentage changes. Parentheses show heteroskedasticity-robust standard errors. Controls are the same as in Equation (1). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

D.3 Difference-in-Differences Robustness Check

We supplement the experimental-period comparisons with a difference-in-differences specification for eight user-level count and total outcomes. These outcomes retain users with no relevant activity as zeros, without conditioning inclusion on searching or opening UGC. For each outcome, we estimate

$$Y_{ip} = \alpha + \gamma \,\mathrm{Treat}_i + \lambda \,\mathrm{Post}_p + \beta \,(\mathrm{Treat}_i \times \mathrm{Post}_p) + \varepsilon_{ip}. \tag{39}$$

Here, Treat_i follows the main analysis: it equals one for the group assigned to retain Generative AI Search and zero for the control group. The indicator Post_p identifies the experimental period. The coefficient β is the post-period treatment–control difference minus the pre-period treatment–control difference. We use OLS with one observation per randomized user in each period, equal user weights within each period, and user-clustered standard errors.

Table A12 reports positive interaction estimates for all eight outcomes, each significant at the 1% level. Their directions are consistent with the main findings on search scale and aggregate demand for direct and downstream UGC.

Table A12: Difference-in-Differences Estimates for Search and Aggregate UGC Demand

Outcome	DiD estimate	Clustered SE
Search scale		
Search sessions	0.3201***	(0.0696)
Distinct queries	0.2447***	(0.0553)
Active search days	0.0652***	(0.0147)
Aggregate direct UGC demand		
UGC previews viewed	3.1981***	(0.8941)
Distinct direct UGC pages opened	0.4141***	(0.1528)
Direct UGC opening events	0.4192***	(0.1540)
Aggregate downstream UGC demand		
Sessions with downstream UGC openings	0.1225***	(0.0332)
Downstream UGC openings	1.5975***	(0.5985)
Pre-period users	100,000	
Post-period users	100,000	
User-period observations	200,000	

Notes: The pre period is March 15–29, 2025 (15 days), and the post period is March 30–April 24, 2025 (26 days). Outcomes are counts or totals over the indicated period. Both periods include all 100,000 randomized users: 50,000 treatment and 50,000 control users. Users with no relevant activity remain in the sample with zero counts. All eight regressions use Equation (39), without additional covariates. Distinct direct openings are counted within sessions and summed over the period. Parentheses contain user-clustered standard errors. Stars use unadjusted two-sided p values: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

E Rank-Specific Content Demand

Sample and measurement. The randomized roster contains 100,000 accounts (50,000 Control and 50,000 Treatment). The observed search panel contains 487,940 searches: 234,805 in Control and 253,135 in Treatment, made by 26,155 and 26,556 users, respectively. Count outcomes over the experiment include the full account roster and assign zero when a user has no qualifying event. Per-search outcomes first average within each searching user, so frequent searchers do not receive greater weight in the between-arm comparison.

The ranking-specific browsing, clicking, CTR, search-state, and intent measures use the same UGC definition as the main browsing and clicking analysis. Table A16 reports rank-specific per-search counts under this definition, and Table 14 reports experiment totals in the main text.

Restoring original rank. Let d denote the recorded display position of a UGC result and a the recorded position of the AI answer in that search. Its original pre-insertion position is

$$r = d - \mathbf{1}\{d > a\}.$$

When no AI position is recorded, the search is coded as having no AI answer and $r = d$. The recorded position coordinates are preserved; observed UGC rows are not renumbered after filtering. Top 5 comprises $r = 0, \ldots, 4$ and Non-top 5 comprises all $r \geq 5$, not only the next five positions or a measured scroll page. According to the platform, relevance and quality inform the ranking. These ranges capture that prioritization, not independently measured high- and low-quality content categories. Invalid or missing ranks remain unassigned and are included only in Overall when the content record is otherwise usable.

Within each search and range, P counts distinct recorded result identifiers and C counts those with an opening record. If an identifier is missing, the record retains a row-specific fallback identifier rather than being merged with other missing identifiers. Among UGC records, 25,649 Control and 26,979 Treatment records have an unassigned rank; 158,011 and 166,982, respectively, use fallback identifiers. These are record counts, not numbers of users. A content identifier can appear in more than one rank range; therefore distinct Top 5 and Non-top 5 counts need not sum to the Overall distinct count. Previews measure logged exposure, not attentive reading, and C differs from a count of repeated opening events.

For counts per search and the three search-state indicators, all observed searches enter before averaging within a user. For CTR, C/P is defined only when $P > 0$ in the relevant range; the user's mean is taken over these eligible searches. A user without an eligible search does not enter that CTR cell. Thus CTR is neither a pooled opening-to-preview ratio nor a ratio with no-preview searches coded as zero. Intent-specific browsing and clicking average all observed searches within each user and original intent category, including zero counts. Intent-specific CTR uses the positive-preview rule within the same categories. Users enter a category when they have an eligible search in it.

Inference. Rank and intent results are estimated jointly from user-level summaries, with user-clustered CR1 covariance retaining dependence across cells; direct contrasts use that covariance. Intent-specific count outcomes use six jointly estimated intent cells. CTR uses the joint rank-by-intent cells. Tests use the corresponding user-cluster degrees of freedom. All significance statements use unadjusted two-sided p values. Intervals are pointwise 95% intervals, not simultaneous intervals.

Table A13: User Samples for Intent-Specific Content Demand

Intent	Browsing and Clicking Control / Treatment	CTR Control / Treatment
Factual/explanatory	11,698 / 11,994	11,698 / 11,971
Procedural/how-to	5,752 / 5,998	5,752 / 5,981
Evaluation/recommendation	6,396 / 6,729	6,396 / 6,727
Experience/opinion	6,073 / 6,283	6,073 / 6,278
Exploratory/open-ended	1,811 / 1,928	1,811 / 1,926
Current/local/navigational	14,361 / 14,672	14,361 / 14,659

Notes: Browsing and clicking include all observed searches within each user and intent. CTR includes searches with a positive number of recorded UGC previews. Each eligible user receives equal weight within an outcome and intent and can contribute to multiple intent categories.

E.1 Full Rank-by-Intent Results and Direct Comparisons

Table A13 reports the user samples for the three intent-specific outcomes in Table 12. Joint tests of equality of the six intent effects yield $F(5, 45200) = 1.531$ for browsing ($p = .176$) and $F(5, 45200) = 2.429$ for clicking ($p = .033$). Table A14 reports every rank-by-intent CTR effect. Joint tests of equality of the six treatment effects yield $F(5, 45177) = 4.882$ for Top 5 ($p < .001$), $F(5, 45177) = 3.259$ for Non-top 5 ($p = .006$), and $F(5, 45177) = 5.007$ Overall ($p < .001$). These tests compare treatment effects across intents. The comparisons condition on observed queries and, for CTR, recorded previews.

E.2 Additional Rank-Specific Per-Search Counts

Table A16 decomposes the main UGC counts by original rank, using the same user weighting as the other rank results. These are per-search outcomes among searching users. Table 14 in the main text instead reports experiment totals, including zero activity, and the direct comparisons between the two rank ranges.

Table A14: Complete Rank-by-Intent Click-Through Results

Intent	Effect (pp)	SE	p	Users (Control / Treatment)
A. Top 5				
Factual/explanatory	-1.064^{***}	(0.227)	$< .001$	11,636 / 11,909
Procedural/how-to	-1.884^{***}	(0.336)	$< .001$	5,718 / 5,944
Evaluation/recommendation	-1.217^{***}	(0.319)	$< .001$	6,369 / 6,683
Experience/opinion	−0.401	(0.339)	0.236	6,047 / 6,248
Exploratory/open-ended	0.649	(0.658)	0.324	1,803 / 1,916
Current/local/navigational	−0.340	(0.222)	0.126	14,293 / 14,592
B. Non-top 5				
Factual/explanatory	−0.107	(0.173)	0.536	8,762 / 8,806
Procedural/how-to	0.058	(0.276)	0.833	4,366 / 4,460
Evaluation/recommendation	-0.811^{***}	(0.253)	0.001	4,968 / 5,223
Experience/opinion	0.574^{**}	(0.277)	0.038	4,184 / 4,343
Exploratory/open-ended	0.349	(0.516)	0.499	1,240 / 1,349
Current/local/navigational	−0.251	(0.168)	0.136	9,309 / 9,321
C. Overall				
Factual/explanatory	-0.889^{***}	(0.168)	$< .001$	11,698 / 11,971
Procedural/how-to	-1.105^{***}	(0.252)	$< .001$	5,752 / 5,981
Evaluation/recommendation	-1.150^{***}	(0.238)	$< .001$	6,396 / 6,727
Experience/opinion	−0.123	(0.253)	0.628	6,073 / 6,278
Exploratory/open-ended	0.630	(0.485)	0.194	1,811 / 1,926
Current/local/navigational	-0.351^{**}	(0.163)	0.031	14,361 / 14,659

Notes: All cells use the same UGC definition as the main analysis. Each user receives equal weight within each rank-by-intent cell. CTR is undefined when there is no preview in that cell. A user may contribute to several cells; inference preserves the cross-cell covariance. All effects are Treatment minus Control. Standard errors are in parentheses. Significance uses unadjusted two-sided p values: $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$.

Table A15: Direct Treatment-Effect Comparisons

Contrast	Effect, SE	95% CI	p
Browsing: Top 5 minus Non-top 5	0.083 (0.071)	$[-0.056, 0.221]$	0.241
Clicking: Top 5 minus Non-top 5	-0.023^{**} (0.010)	$[-0.043, -0.003]$	0.025
CTR (pp): Top 5 minus Non-top 5	-0.535^{***} (0.143)	$[-0.816, -0.254]$	$< .001$
No preview (pp): Top 5 minus Non-top 5	−0.470 (0.330)	$[-1.116, 0.176]$	0.154
Preview, no opening (pp): Top 5 minus Non-top 5	1.694^{***} (0.369)	$[0.970, 2.418]$	$< .001$
Any opening (pp): Top 5 minus Non-top 5	-1.224^{***} (0.339)	$[-1.888, -0.559]$	$< .001$
CTR, Top 5: factual minus experience (pp)	-0.663^{*} (0.395)	$[-1.437, 0.112]$	0.094
CTR, Non-top 5: factual minus experience (pp)	-0.681^{**} (0.320)	$[-1.308, -0.053]$	0.033
CTR, Overall: factual minus experience (pp)	-0.767^{***} (0.295)	$[-1.345, -0.188]$	0.009
Browsing: factual minus experience	-0.3705^{*} (0.2108)	$[-0.7836, 0.0426]$	0.079
Clicking: factual minus experience	-0.1121^{**} (0.0456)	$[-0.2015, -0.0226]$	0.014

Notes: All contrasts use the same UGC definition as the main analysis. Browsing and clicking are counts per search; CTR and search-state contrasts are percentage points. Factual minus experience compares factual/explanatory with experience/opinion queries. Contrasts retain within-user covariance across cells. All effects are Treatment minus Control. Standard errors are in parentheses. Significance uses unadjusted two-sided p values: $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$.

Table A16: UGC Browsing and Clicking per Search by Original Rank

Outcome	Top 5	Non-top 5	Overall
Per search: 52,711 searching users			
Browsing	−0.0349***	−0.1176	−0.1487**
	(0.0064)	(0.0715)	(0.0727)
Clicking	−0.0338***	−0.0110	−0.0447***
	(0.0058)	(0.0107)	(0.0139)

Notes: This table uses the same UGC definition as the main analysis. We average each user's search-level count before comparing users, giving each of the 52,711 searching users equal weight. Counts are distinct result identifiers, not repeated opening events. Original rank is restored before forming the ranges. All effects are Treatment minus Control. Standard errors are in parentheses. Significance uses unadjusted two-sided p values: $^{***}p < .01$, $^{**}p < .05$, $^{*}p < .10$.

F Structural Measurement, Estimation, and Diagnostics

This appendix details measurement, estimation, and diagnostics for the 33-parameter model in Table 16.

F.1 Linked Sample and AI-Opening Measurement

The linked structural panel is drawn from the 100,000 randomized accounts and covers March 30–April 14, 2025. It contains 304,306 search sessions, 3,070,457 distinct result-slot previews, and 28,232 AI previews. The balanced panel retains all 16 daily observations per included account, including days without search. The daily likelihood therefore uses both activation and inactivity.

The session-level AI-opening count determines whether full content was opened. A positive count records an opening. Zero and missing counts record no opening under the platform's convention. This rule yields 22,363 opened AI answers, or 79.2115% of AI previews viewed. It preserves the distinction between viewing a preview and opening the answer.

Session-level records determine whether an AI answer was opened. The current product places an available AI answer at display rank zero, but does not force it to be the first preview. Within-session previews follow their original recorded times, including AI `cardshow_time`, with the retained tie rule for equal timestamps. Each session contains at most one AI preview; UGC may be encountered before or after it.

The result's source display-slot index identifies a UGC candidate within a session. Repeated records of the same slot are merged at their first preview, with an opening recorded if any merged record reports one. This merges 293 extra records in 226 sessions and six repeated opening events. It retains every user and session. The resulting data contain 506,752 UGC openings and 22,363 AI openings.

Missing rank values are assigned to unused nonnegative integer slots within their session by minimizing squared distance to the retained rank approximations. Observed source ranks are preserved. When AI is observed at source index a, its source slot is reserved during completion. For a UGC source index r, we remove AI from its source position and reinsert it at display rank zero:

$$r^{\text{org}} = r - \mathbf{1}\{r > a\}, \qquad r^{\text{display}} = r^{\text{org}} + 1.$$

AI receives display rank zero. This conversion preserves UGC order and unviewed gaps; it does not densely rerank the viewed subset or change preview times. When no AI row is observed, recorded UGC display positions are retained and possible AI-present and AI-absent layouts use their documented conditional rank completions. Imputed positions are not treated as independent evidence that AI is absent. Rank completion and the assumed number of available results are separate: the completed path does not reveal each user's full set of unviewed results.

F.2 Text-Based Information and Expected Length

Semantic units, information amount, and density. The unit of measurement is a query–answer text pair. We normalize the visible full-answer text with Unicode NFKC, remove hidden reasoning blocks, and exclude whitespace when counting characters. The resulting length L retains visible Markdown or citation syntax. It measures normalized logged text rather than characters rendered in a particular screen layout.

A semantic unit is a minimal, self-contained piece of query-relevant content: a proposition, instruction, recommendation, explanation, condition, or comparison. Independent substantive steps count separately; restatements and redundant summaries do not. Distinct qualifications, causal claims, and scope restrictions are retained, along with attribution and uncertainty. For ambiguous queries, coding records the assumed interpretation without using answer-only details to establish the intended object.

The distinction between information expressed and the amount of text draws on propositional-density measurement and semantic content-unit evaluation (Brown et al., 2008; Lex et al., 2012; Nenkova and Passonneau, 2004). Our unit is tailored to query relevance. Retained units receive equal weight, independently of subsequent behavior. The measure captures expressed content rather than factual accuracy, novelty to a user, or individual learning.

Model-assisted extraction identifies candidate units with exact source evidence; a second pass checks relevance, support, and duplication. Omissions are corrected against the full source. Creative tasks count substantive requested outputs but exclude greetings and generic wrappers.

The corpus contains 25,150 distinct query–answer pairs, including supplemental texts used only to construct missing inputs. Of these, 25,139 are measured and 11 remain missing: ten annotation failures and one unresolved task interpretation. Within the measured set, 78 pairs receive source-backed corrections. No independent human benchmark has been applied, leaving measurement error in relevance, scope, and granularity unquantified.

For measured pairs, N is the number of retained distinct units, $D = 100N/L$ is density, and $\nu = N/10$ is the model's AI information amount. The scale of ten is a fixed linear normalization. The largest measured stock is 21.4, and values above one are retained. AI encounters with measured $N = 0$ are also retained. A separate opened indicator records whether an answer has been opened, even when its measured information stock is zero. Opening such an answer still incurs reading cost.

Missing measurements and maintained model inputs. Some observed AI encounters cannot be assigned an observed information amount. Full-answer text is unavailable for 990 encounters. A further 11 encounters have text but no resolved semantic measurement; nine of these answers were opened. For the latter group, observed L remains available, while observed N, D, and $N/10$ remain missing. Neither group is deleted from the behavioral sample.

The likelihood substitutes mean information amount from distinct measured answers to the same query, weighted equally. Without such answers, it uses the within-intent mean of exactly matched measured AI encounters, weighted by display frequency. Missing-text cases retain their donor links; cases with text but missing N use the fixed measurement corpus. Unresolved donors are excluded, and every instance of an answer receives the same input. Imputed information amounts remain distinct from observed counts and do not enter pre-opening beliefs.

Observed choices and histories remain in the sample. Because choices are nonlinear in information amount, substitution by a mean need not equal averaging over missing information amounts; standard errors condition on these inputs.

Pre-opening information beliefs. We estimate $\mathcal{H}^A(\nu \mid k, b^{\text{ref}})$ from the 27,231 exactly matched AI encounters with measured information amounts. Supplemental query donors and imputed values are excluded.

The seven intents and four reference-count groups yield 28 cells. Each contains at least 37 measured encounters, so no cell requires broader pooling. We retain the complete empirical distribution in each cell. Identical values are combined using their encounter frequencies, producing 1,539 weighted support points across the cells. The largest cell contains 3,138 encounters.

The user does not know the realized full-answer information amount before opening. It enters the observed post-opening stock and later decisions. Reference count is visible in the preview; AI position does not enter the information belief. A position-only policy therefore changes when the user might encounter an answer, not the information distribution assigned to unchanged content.

UGC stock and its scale. Higher-ranked UGC receives a larger information weight within the same intent. Let $F^{\text{org}}_{k,0}(r^-)$ be the fraction of control-group viewed UGC results with display-slot indices strictly smaller than r. The reverse percentile and opening weight are

$$w_k(r) = 1 + q_k(r), \qquad q_k(r) = 1 - F^{\text{org}}_{k,0}(r^-). \tag{40}$$

The UGC-only rank starts at zero and excludes AI; a smaller index is a higher placement among UGC results. The reverse percentile $q_k(r)$ lies between zero and one. Each distinct opening therefore contributes one baseline unit plus its rank component. Ranks zero through four keep their individual weights. For ranks at least five, we first average the raw weights over control-group tail previews within intent. Every tail opening uses this common weight. The decay is $\exp(-\bar{w}_{kT}/\tau_k)$, rather than a mean of individual exponential decays. Summing the grouped opening weights gives the model's raw stock S^U.

The fixed scale used to convert this sum into bounded stock is

$$\tau_k = \frac{S^U_{k,.90}}{\log 10}, \tag{41}$$

where $S^U_{k,.90}$ is the retained 90th percentile of positive terminal raw stock under the individual-rank no-AI reference construction. The rank CDF and this scale are fixed before behavioral estimation. Grouping the tail changes individual increments and terminal stocks; it does not recalibrate τ_k to force the new terminal percentile to equal 0.9.

Expected length. The raw inputs are cleaned visible-character count for AI and platform word count for UGC. Write R^j for the raw count from source $j \in \{A, U\}$. We reduce the influence of very long answers with a log transformation and divide by the pooled median:

$$\ell^j = \frac{\log(1 + R^j)}{\widetilde{R}}, \qquad \widetilde{R} = \text{median}_{\text{pooled}}\{\log(1 + R)\}. \tag{42}$$

The transformed length ℓ^j is dimensionless and has a pooled median of one. The same denominator $\widetilde{R}$ is used for both sources. The model assumes a common effort coefficient on their normalized lengths within class.

Separate five-fold ridge regressions predict length from a compact query embedding, query-text features, and intent. The query features include length, digit, Latin-character, whitespace, and punctuation shares, and an indicator for a question mark. The AI model also uses visible reference-count groups. All observations

Table A17: Saved Query-Held-Out Expected-Length Predictions

Source	Records	Observed mean	Predicted mean	RMSE	R^2
AI	27,242	0.9842	0.9841	0.0880	0.2003
UGC	2,536,278	0.9793	0.9793	0.3058	0.0165

Notes: Lengths use Equation (42). RMSE is root mean squared prediction error. R^2 compares squared error with the observed sample-mean predictor. Observations receive equal record weight. The checks use saved continuous predictions before compression into expected-length groups. Mean-bias inference clusters by user. These diagnostics concern observed length, not reading time or semantic accuracy.

with the same normalized query belong to one fold. Each query is predicted by a model trained on the other four folds. This yields the source-specific expectation $\mu_L^j(z) = \mathbb{E}[\ell^j \mid z]$, where z contains the corresponding pre-opening information. Eventual opening and stopping choices are excluded.

The length regressions use 2,536,278 UGC records and 27,242 matched AI records with observed length. Missing attributes use fixed completed inputs and predictions, without zero-length coding or extra utility coefficients. Separate predictors allow expected AI and UGC lengths to differ.

The UGC predictor assigns every query an expected length, including queries starting with AI or without a UGC preview. We assign similar predictions to one of nine within-intent expected-length groups b^L for computation; the assigned group remains fixed throughout the query. Initial session values average over this group, possible layouts, and the first encountered display rank, source, and AI preview profile.

Measurement coverage and length prediction. The 28,232 AI encounters comprise 27,231 with measured information amounts and 1,001 with missing inputs. Stock is $N/10$ and density is $100N/L$ for $L > 0$; imputed inputs are excluded from belief donors.

Query-held-out predictions have mean observed-minus-predicted errors of 3.2522×10^{-5} for AI and 3.2431×10^{-5} for UGC (Table A17). None of fourteen source-by-intent mean-bias tests rejects after Holm adjustment. These results support average length calibration; interpreting predictions as users' reading-effort expectations remains an assumption.

F.3 Common Organic Supply and the Remaining Search Results

The platform's confirmation of non-personalized UGC retrieval and ranking, unchanged by AI availability, motivates Assumption 2. We model a common distribution rather than identical realized lists. Initial display-position weights are calibrated from actual first previews in the control environment. They distinguish display positions zero through four and equal per-result weights for later positions, conditional on query intent and the expected-length group that remains fixed within a query.

State of the remaining search results. Each of the top five UGC-only ranks is a distinct candidate. The tail initially contains 217 exchangeable UGC candidates, covering UGC-only ranks 5–221. AI, when available, is a separate candidate and consumes no UGC slot. The state records viewed results and the number of tail results remaining, whether previews were opened or skipped. At a current preview, that result is already included in this history. A result can never return. Exact display positions whose source, UGC information, or existence differs across layouts remain explicit; only the common exchangeable remainder is counted. For

Current, UGC-only ranks zero through four occupy display ranks one through five with AI and zero through four without AI.

Equation (7) uses the full observed set $\mathcal{B}_n$ to make the user's information explicit. For the value calculation, partition display positions into individually tracked positions $\mathcal{E}_g$ and the common exchangeable tail $\mathcal{T}_g$. The equivalent inventory coordinates are

$$B_n = \mathcal{B}_n \cap \mathcal{E}_g, \qquad c_n^{\text{tail}} = |\mathcal{B}_n \cap \mathcal{T}_g|. \tag{43}$$

These are the B and c used above: exact labels are retained where they affect information or availability, while the common tail enters through its remaining count. The common display tail need not contain all 217 UGC-only tail candidates, since boundary positions can differ across layouts. Simulation retains visited labels to prevent repeat previews.

Upon the next preview at rank r, the observed set becomes $\mathcal{B}_{n+1} = \mathcal{B}_n \cup \{r\}$, the current rank, source and visible profile are replaced, and the layout belief is updated using that observation. The AI-preview indicator remains one once AI has appeared. An opening can additionally reveal the layout through its information increment, as described below. The realized layout is not an extra observed state coordinate. Conditional on (x_n, m_n), the user does not need the full sequence of past opening choices to evaluate the next choice; direct-opening counts are recorded separately for the downstream-reading model.

For current display group j and intent k, baseline next-position weights ω_{kjr} are proportional to the exponential of a destination effect, an intent-by-destination effect, and a current-group-by-destination effect. Display positions zero through four are separate groups; later positions share a per-result weight. Tail destination and current-tail effects serve as reference categories. A current individually tracked position is unavailable, so its own-destination coefficient is omitted. The specification has 55 encounter coefficients and seven intent-specific AI-availability coefficients, with no AI-specific encounter coefficient or length interaction in availability. A quadratic penalty of strength 10 applies to intent and current-group deviations, leaving pooled destination effects unpenalized.

We estimate this environment from observed preview prefixes pooled across arms, through the first opening inclusive, retaining every session. Each possible layout has its own finite remaining-result denominator in Equation (6); unavailable layouts have zero likelihood. The AI-availability probability mixes these layout likelihoods. The first stage does not use later, post-opening transitions. Those later observations remain in the structural likelihood, which conditions on the fitted environment.

Removing previewed results accounts for availability mechanically; it does not by itself resolve selection into continuation. Assumption 3 requires the layout-conditional encounter law of continuing users to apply to comparable stoppers. Its weights depend on intent and current display position, but not stock, opening history, expected-length group, or latent class. The user's distribution over layouts can change with observed previews and information revealed by openings, changing the predicted next result. This distinction separates fixed encounter weights from updating beliefs. The model treats original UGC tail candidates as interchangeable in information and expected reading cost. The finite inventory is common across sessions and is not estimated from each user's stopping point.

AI placement and users' information. Under Current, AI occupies display rank zero when available, but users can encounter another display position first. AI and UGC enter the same position-based encounter law, with type-neutral weights normalized over the remaining results in each layout. Previewing any result removes it; previewing AI consumes the one AI opportunity whether it is opened or skipped.

Before availability is revealed, a displayed UGC result may have different UGC-only ranks under the two layouts. The user averages its opening value over these possibilities before the opening-choice shocks are realized. Opening reveals the information increment: distinguishable increments resolve the layout, while identical increments leave it unresolved. An AI preview reveals its reference-count profile, not its full semantic amount; the latter is revealed only on opening. Expected UGC reading length remains fixed for the query.

The researcher may infer the layout retrospectively from the full recorded session, but that later evidence is not given to the user at earlier choices. Sessions compatible with both layouts retain both in the likelihood. The initial value deducts the same class-specific preview cost as later previews. The Bellman calculation and path simulation use this same without-replacement timing and observable information.

Observed intent distributions. The first-intent distribution and conditional next-intent matrix are estimated from control queries and held fixed across policies. Joint experimental-group comparisons do not detect a difference in first-query intent ($p \approx 0.762$) or next-query intent conditional on the previous intent ($p \approx 0.765$). The first-query comparison uses 57,509 control and 59,870 treatment observations. The conditional next-query comparison uses 88,729 and 98,198 observations, respectively.

These checks concern observed searches and queries with successors. Holding their intent distributions fixed under new policies remains a modeling restriction.

F.4 Payoff Restrictions and the Value of Stopping

Opening and continuation shocks are conditionally independent Type-I extreme-value shocks, centered to remove Euler's constant from the expected maximum. Both scales are one. Fixing one scale establishes the utility unit; equalizing the two scales restricts their relative dispersion. The within-session discount factor and the coefficient on logarithmic information utility are also fixed at one. Together, these choices specify the scale and curvature of the decision problem.

The payoff in Section 5.4 counts each information gain once. If m_j is the stock after opening number j, then

$$\sum_j \{G_h(m_j) - G_h(m_{j-1})\} = G_h(m_{\text{terminal}}) - G_h(0,0) = G_h(m_{\text{terminal}}). \tag{44}$$

Subtracting incurred reading costs and the costs of all previews, including the first, gives Equation (23). Information benefit is retained upon stopping, while viewing an additional preview must justify a new cost.

The logarithmic utility function makes the marginal value of either source decrease with information obtained from the other. For positive weights, this relationship is

$$\frac{\partial^2 G_h}{\partial m^A \partial m^U} = -\frac{\chi_{Ah}\chi_{Uh}}{(1 + \chi_{Ah} m^A + \chi_{Uh} m^U)^2} < 0. \tag{45}$$

The negative cross-partial formalizes the maintained within-session substitution. Aggregate demand for UGC can still rise when users search more often. Its aggregate direction is determined by both levels of the model.

A UGC–AI–UGC path. Under a display rule with an available AI answer, consider a user who opens a UGC result, continues to AI and opens it, then continues to a second distinct UGC result and opens it before stopping. Let u_1 and u_2 be the UGC stocks after the first and second UGC openings, and let $\nu = N/10$ be the AI information revealed. The stock path is

$$(0,0) \longrightarrow (0,u_1) \longrightarrow (\nu,u_1) \longrightarrow (\nu,u_2).$$

The three information gains sum to $G_h(\nu,u_2)$. There are three reading costs and three preview costs. With the query's UGC length expectation fixed, total payoff excluding choice shocks is

$$G_h(\nu,u_2) - \gamma_{Lh}\{2\mu_L^U + \mu_L^A\} - 3\kappa_{Bh}. \tag{46}$$

The two UGC openings incur their common query-specific expected reading cost; AI incurs its source-specific cost. Viewing each preview, including the first, incurs κ_{Bh}.

Users anticipate possible paths rather than knowing this sequence in advance. After opening AI, the extra gross benefit from the second UGC is

$$\log\left(1 + \frac{\chi_{Uh}(u_2 - u_1)}{1 + \chi_{Ah}\nu + \chi_{Uh}u_1}\right). \tag{47}$$

The numerator is the weighted UGC increment. More AI information raises the denominator and reduces this benefit. The user may therefore stop before the second UGC opening. The order of encounters matters for choices, even though completing the same openings produces the same final gross information benefit.

F.5 From Initiation Probabilities to Expected Activity

Daily activation and repeated same-day querying are distinct sources of search volume. Write $p_h^D(g)$ and $p_h^Q(k,g)$ for the probabilities in Equations (28) and (26). Let $\mathcal{N}_h(k,g)$ be the expected number of sessions from a current query of intent k through the end of that day's search, including the current session. One session is already underway. With probability $p_h^Q(k,g)$, another begins with an intent drawn from $\mu_0(k' \mid k)$:

$$\mathcal{N}_h(k,g) = 1 + p_h^Q(k,g) \sum_{k'} \mu_0(k' \mid k)\mathcal{N}_h(k',g). \tag{48}$$

Expected active days over the 16-day window are $16p_h^D(g)$. Expected sessions are $16p_h^D(g)\sum_k \mu_0(k)\mathcal{N}_h(k,g)$. These identities translate the estimated initiation probabilities into counts; they introduce no additional preference parameter.

The two-level model solves the within-session problem and uses its prospective value in separate initiation equations. The zero stopping value inside a session is a residual value: information already acquired is retained. It excludes the value of a future query, which enters the separate initiation decision. Stocks reset at a new query, and feature expectations remain stable across queries and days. Within a query, users update both information stock and remaining AI availability.

F.6 Joint Estimation and Numerical Diagnostics

We maximize Equation (33) with bound-constrained L-BFGS-B (Byrd et al., 1995). At each trial parameter vector, backward induction solves both classes' session problems from the point at which no results remain. The resulting opening, continuation, and initiation probabilities, together with the downstream count distribution, give each class's history likelihood. Their prior-weighted mixture yields the objective and gradient.

All parameters are optimized jointly, integrating over user class at each trial vector. The computation follows the nested model-solution and parameter-optimization logic of Rust (1987). Here, finite-horizon backward induction replaces infinite-horizon fixed-point iteration. Coordinate scaling improves numerical conditioning without changing the objective or economic coefficients.

Tracking the remaining search results makes the session problem acyclic: each UGC preview either adds a viewed top rank or increases the viewed tail count, and AI can appear at most once. A UGC preview removes a result; an AI preview leaves UGC availability unchanged but consumes the single AI opportunity. Thus no sequence returns to the same full search state, and we solve backward from exhaustion. After exhaustion, continuation is unavailable ($C = -\infty$), and the continuation-or-stop value is zero ($M = 0$). At most 222 UGC results are available; an AI opportunity can add one preview. The endpoint therefore follows from exhausting the available results rather than an additional preview horizon.

The retained approximation uses eleven UGC-stock grid points, 25 AI-information knots, and 21 layout-belief nodes. The stock grids include zero and their respective upper endpoints. AI knots are spaced using $\log(1 + \nu/0.1)$ to place more knots near zero. This is a numerical spacing rule; economic information remains $N/10$. Values use cubic interpolation on these grids, with stock interpolation in original units. Expectations at AI use the full empirical information distribution. The layout-specific encounter weights, UGC information increments, and stock scale are held fixed during behavioral estimation.

Three nearby starting vectors were screened at a projected-score-per-user tolerance of 10^{-1}. The best was refined to a tolerance of 10^{-3}. The final largest projected native score per user is 9.6189417×10^{-4}. All 33 estimates are interior. Multiple nearby starts and local curvature establish numerical regularity of this solution, without establishing a global likelihood maximum.

Standard errors allow observations from the same user to be dependent. Let $s_i(\widehat{\theta})$ be the gradient of user i's log likelihood after integrating out class. Let $\widehat{H}$ be the negative Hessian of the total log likelihood at the estimate $\widehat{\theta}$. The covariance is

$$\widehat{\mathrm{Var}}(\widehat{\theta}) = \widehat{H}^{-1} \left[\sum_i s_i(\widehat{\theta}) s_i(\widehat{\theta})' \right] \widehat{H}^{-1}. \tag{49}$$

The middle term sums users' score outer products. The two inverse information matrices translate that variation into parameter uncertainty. We calculate the Hessian with central score differences, comparing relative steps of 2×10^{-5} and 10^{-5} with parameter-specific scaling. The reported covariance uses the finer step and is expressed in the original parameter units. It includes all 33 parameters and their cross-class and cross-block covariances.

The information matrix is positive definite with rank 33. In scaled coordinates, its smallest eigenvalue is 8.592768 and its condition number is 4.872644149×10^5. All 33 standard errors are finite and positive. The covariance retains the joint variation of the two classes and all behavioral equations when evaluating derived outcomes.

For a derived value $f(\theta)$, such as the gain in prospective query value, the delta-method variance is $\nabla f(\widehat{\theta})'\widehat{\mathrm{Var}}(\widehat{\theta})\nabla f(\widehat{\theta})$, where ∇f is the vector of derivatives with respect to all parameters. This inference conditions on semantic coding, assigned values for missing information amounts, expected-length predictions, intent laws, and the search environment. It excludes uncertainty in those first-stage inputs. Parameter uncertainty is also distinct from Monte Carlo error in the numerical calculation of a simulated outcome.

G Counterfactual Construction and Additional Results

This appendix details the three design analyses in Section 8, their numerical evaluation, and Monte Carlo error. Policy tables report the main simulation estimates at fitted parameters. The same two design objectives guide all three analyses: joint improvement in sessions and direct UGC openings per session, and higher total direct UGC openings per user when joint improvement is not established. These criteria organize the evaluation of fixed policies; they do not change the selection rule used for a policy after its main simulation results are observed.

G.1 Common Environment and Information Timing

Every policy uses the same transition law for the finite set of search results as the estimated model. A preview removes the encountered result before the opening decision. UGC-only ranks zero through four are tracked individually. Every tail preview depletes the 217-result UGC tail by one, whether opened or skipped. The display position controls encounters; the UGC-only rank controls information. Inserting AI does not remove any UGC result or change its information quality. The one AI opportunity is consumed when its preview is encountered, even if the answer is skipped. No second AI answer can arrive.

The user's query-specific UGC expected-length group remains fixed. AI stock is the semantic amount $N/10$, with reference-count-conditioned beliefs before opening. For information multiplier η,

$$N^p = \eta N, \qquad \nu^p = \eta N/10, \qquad L^p = L, \qquad \eta \in \{0.8, 0.9, 1, 1.1, 1.2\}. \tag{50}$$

The same support probabilities accompany the scaled amounts. Zero information remains zero; fractional amounts are a continuous policy index. Expected reading costs remain fixed. Allocation and disclosure retain $\eta = 1$.

The first- and conditional next-intent distributions remain fixed. Stocks reset at a new query. User classes are drawn from probabilities based on observed characteristics, rather than from posterior classification using treatment-period outcomes. Expected values for each policy determine initiation. Downstream UGC openings follow the estimated relationship with direct UGC openings and do not enter prospective session value.

G.2 The Current-Policy Benchmark

The common-population benchmark in Table A18 uses selective availability at display position zero for Current and zero availability for No AI. These are simulated policy outcomes, separate from the experimental intention-to-treat estimates and the observed-history conditional predictions described in Section 7.3.

Table A18: The Current Design Relative to No AI

Outcome	No AI	Current	Difference (MC SE)
Search sessions/user	2.8644	3.2034	0.3390 (0.0039)
Active search days/user	1.1455	1.2001	0.0546 (0.0008)
Direct UGC openings/session	1.6656	1.6402	-0.0253 (0.0015)
Total direct UGC openings/user	4.7714	5.2545	0.4830 (0.0080)
Sessions with downstream openings/user	1.1938	1.3172	0.1234 (0.0020)
Downstream UGC openings/user	13.3901	14.7836	1.3935 (0.0341)

Notes: Paired simulations over 16 days. Per-user means include inactive users; per-session intensity averages replication ratios. Standard errors measure Monte Carlo error at fitted parameters.

Table A19: Twenty Percent More Information at Alternative Positions

Position	Sessions/user	UGC openings/session	Total UGC openings/user
0	0.0325 (0.0013)	0.0001 (0.0003)	0.0538 (0.0022)
1	-0.0042 (0.0005)	0.0275 (0.0008)	0.0814 (0.0027)
2	-0.0402 (0.0012)	0.0505 (0.0011)	0.0938 (0.0043)
3	-0.0653 (0.0017)	0.0660 (0.0013)	0.0999 (0.0050)

Notes: Entries are differences from Current with paired Monte Carlo SEs. All positions retain Current availability and multiply AI information by 1.2 at fixed expected length. UGC openings are direct openings.

G.3 Additional Placement Results

Table A19 reports positions zero through three with 20% more information. Figure 11 compares these positions across five information multipliers, including the Current-equivalent position-zero, multiplier-one cell. At position zero, the per-session effect is not resolved from zero by its paired Monte Carlo interval, so joint improvement is not established. Delayed higher-information answers instead raise direct UGC openings per session while reducing sessions relative to Current. Under the aggregate-demand objective, the design-selected position-three rule has the largest total direct UGC point estimate in the evaluated grid, 1.9018% above Current. The stars in Figure 11 identify this design, not statistical significance or an optimum outside the evaluated grid.

G.4 Fixed-Budget Allocation: Selection and Exact Rules

The fixed query weights integrate the full-roster class priors, Current daily activation, and expected same-day query sequences. A single observable-intent policy applies to both latent classes. The weights remain fixed during policy selection and main evaluation.

Availability enters the user's beliefs before the first preview and after every observed preview. An available AI answer at display position zero need not be encountered first. Consequently a policy with an interior availability probability is solved at its actual probability; it is not a mixture of behavior under known NoAI and known AllAI layouts.

Sparse availability curves generate candidate allocations, using six initial knots and the exact Current probability for each intent. Candidates are then evaluated with their actual availability-specific beliefs, values, and behavior. These curves are a search device, not a claim of full-domain interpolation precision.

For the joint-improvement objective, let $Q(\lambda)$ denote expected sessions and $I(\lambda)$ the ratio of expected direct UGC openings to expected sessions. The balanced policy search uses the objective

$$\max_{\lambda,t} \; t \quad \text{subject to} \quad Q(\lambda)/Q(\lambda^0) - 1 \geq t, \quad I(\lambda)/I(\lambda^0) - 1 \geq t, \tag{51}$$

and the budget and probability restrictions in Equation (35). It also searches the two frontier objectives separately, maximizing one proportional gain subject to a nonnegative value of the other. Multiple feasible starting allocations provide numerical candidates. The balanced policy is a selected rule in this searched class, not a proved global maximizer. Its selection of Current does not imply optimality under a different objective; a policy with a trade-off can yield higher total direct UGC demand.

Candidate comparisons use the design sample, including local checks at actual availability probabilities. The balanced rule and both frontier rules retain Current. For the aggregate-demand objective, a separate search maximizes expected total direct UGC openings per user. An AI-opening-maximizing allocation provides a distinct engagement benchmark. Policies are selected using design calculations and fixed before an independent main simulation. Evaluation is conditional on the retained numerical approximation and the searched candidate set.

Table 20 in the main text reports the resulting availability probabilities.

G.5 Disclosure of AI Information Amount

Disclosure changes the information available before opening an unchanged AI answer. Let $H(N \mid x)$ be its semantic-unit distribution given what the user knows at the preview, including intent and visible reference count. This is the belief $\mathcal{H}^A$ expressed in raw counts N rather than scaled amounts $\nu = N/10$. The disclosure rule g_k assigns each count to a label s for intent k; here s indexes labels, not sessions. After observing label $s = g_k(N)$, the belief is

$$H(N \mid x, s) = \frac{\mathbf{1}\{g_k(N) = s\} H(N \mid x)}{\mathbb{P}(g_k(N) = s \mid x)}. \tag{52}$$

The formula retains and renormalizes the amounts consistent with the label. Averaging these posteriors over label probabilities recovers the original answer distribution. The label conveys no acquired stock; stock changes only after opening. Expected reading length remains fixed, so the label changes the expected information benefit but not the cost assigned to opening an unchanged answer.

Common labels use cutoffs 11.5 and 20.5 semantic units. The intent-specific candidate family combines boundaries associated with one-sixth, one-third, one-half, two-thirds, and five-sixths quantiles, together with the common rule. Empty and tied bins are excluded. Within the finite family, selection maximizes prospective query value under the fixed reference class–intent mix. Thresholds are frozen before the main UGC outcomes are evaluated. Full disclosure reveals exact information amount and is not restricted to three labels.

The three fixed disclosure designs are evaluated against both objectives, rather than re-optimized for either reported UGC outcome. Their main simulation raises sessions, but the Monte Carlo intervals for the per-session effects include zero. Joint improvement in both margins is therefore not established. Total direct UGC openings increase by 0.1903%, 0.2073%, and 0.2838% for common labels, intent-specific labels, and full disclosure, respectively. The latter is the largest aggregate gain among these three designs, not a claim of optimal disclosure over all possible signals.

Table A20: Intent-Specific Semantic-Unit Thresholds

Intent	Lower threshold	Upper threshold
Factual / Explanatory	7.5	15.5
Procedural / How-to	11.5	20.5
Evaluation / Recommendation	9.5	17.5
Experience / Opinion	9.5	16.5
Exploratory / Open-ended	11.5	20.5
Current / Local / Navigational	6.5	13.5
Ambiguous / Other / Unclassified	7.5	18.5

Notes: Labels indicate $N \le c_1$, $c_1 < N \le c_2$, and $N > c_2$. Common labels use thresholds 11.5 and 20.5 for every intent. These are semantic-unit counts, not stock $N/10$ or density.

Table A21: Disclosure Effects Relative to Current

Outcome	Common labels	Intent-specific labels	Exact disclosure
Search sessions/user	0.0052 (0.0005)	0.0059 (0.0005)	0.0082 (0.0006)
Active search days/user	0.0010 (0.0001)	0.0011 (0.0001)	0.0016 (0.0001)
Direct UGC openings/session	0.0005 (0.0003)	0.0004 (0.0003)	0.0005 (0.0003)
Total direct UGC openings/user	0.0100 (0.0010)	0.0109 (0.0011)	0.0149 (0.0014)
AI openings/user	-0.0050 (0.0006)	-0.0056 (0.0006)	-0.0068 (0.0006)
Downstream UGC openings/user	0.0283 (0.0047)	0.0310 (0.0050)	0.0419 (0.0055)

Notes: Absolute differences with paired Monte Carlo SEs. Availability, display position zero, answer content, and expected reading lengths remain at Current. Per-user outcomes cover 16 days and retain inactive users.

G.6 Outcome Aggregation and Monte Carlo Error

For a per-user outcome let $Y_b(p)$ be the replication-b mean under policy p. The $B = 100$ replications each contain 1,000 users followed for 16 days. Inactive users remain in all per-user denominators. For comparator r, the absolute effect and paired Monte Carlo SE are

$$\widehat{\Delta} = \frac{1}{B}\sum_b [Y_b(p) - Y_b(r)], \qquad \widehat{\mathrm{SE}}_{\mathrm{MC}}(\widehat{\Delta}) = \frac{\mathrm{sd}_b[Y_b(p) - Y_b(r)]}{\sqrt{B}}. \tag{53}$$

Common draws preserve cross-policy covariance without forcing identical actions or exposure after choices diverge.

For direct UGC intensity, each replication first forms its ratio of total openings $U_b(p)$ to sessions $Q_b(p)$:

$$\widehat{I}_{\mathrm{mean}}(p) = \frac{1}{B}\sum_b \frac{U_b(p)}{Q_b(p)}. \tag{54}$$

The paired SE applies to differences in these ratios. Every replication has positive sessions. This is neither the average of individual users' ratios nor the ratio of expected population totals.

For any policy means $\bar{Y}(p)$ and $\bar{Y}(r)$, reported percentage changes equal

$$\widehat{\delta} = 100\{\bar{Y}(p)/\bar{Y}(r) - 1\}. \tag{55}$$

The Monte Carlo influence for replication b is $100\{Y_b(p) - [\bar{Y}(p)/\bar{Y}(r)]Y_b(r)\}/\bar{Y}(r)$. Its standard deviation divided by $\sqrt{B}$ retains variation in both means and their pairing. Figures use this ratio-of-means calculation for percentage error bars. Intervals are pointwise normal intervals, not simultaneous bands over the policy search.

G.7 Numerical Checks and Extrapolation Boundaries

The implementation checks the transition probabilities over remaining search results and simulated paths against the same recursion over remaining results. Current places available AI at display rank zero without forcing its first preview, and no-AI policies remove the AI opportunity. Disclosure posteriors preserve the original information distribution. The main simulation panels contain all scheduled replications and do not hit the population query cap. These checks cover probability consistency, simulated paths, and completion of scheduled replications. Policy comparisons use the retained numerical approximation.

The simulations condition on semantic measurements, expected-length predictions, completed ranks, grouped UGC increments, initial and successor distributions, intent laws, and the assumed finite tail. Their Monte Carlo intervals do not include estimation or policy-selection error. The nonexperimental policy effects further rely on stable expectations, the value-to-initiation relationship, and an unchanged content supply. Additional AI opportunities inherit their intent's current answer distribution. The results do not establish long-run creator participation, platform profit, or welfare effects.

Research Support and Competing Interests

Designing Generative AI Search:
Search Activity and Demand for Creator Content

Hongxian Huang, Jinglong Zhang, Jason Chan, and Anindya Ghose

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.