AI Overview or Overreach? Google's Strategic Deployment of Generative AI in Search*

Robin Ng† Michael Wessel‡

2nd April 2026

Abstract

In May 2024, Google introduced AI Overviews, which synthesize search results into direct answers on the search engine results page, contributing to the growing prevalence of zero-click searches. Notably, Google does not employ AI Overviews for every search query. The key to understanding this selectivity lies in the heterogeneity of search intent, and we argue that AI Overviews are best understood as a strategic instrument for maximizing average revenue per user. We develop a theoretical framework in which a monopolist search platform decides whether to deploy AI Overviews across queries that differ in their search intent along two dimensions: whether the search is exploratory or targeted, and whether it is monetizable. For each intent type, we derive conditions under which the platform benefits from deployment and generate testable hypotheses. To test these predictions, we construct a novel dataset of over 2,000 Google search queries and 15,118 search engine results page observations, where AI Overviews appeared in 31.2% of searches. Consistent with revenue maximization, we find that deployment patterns vary systematically across intent types: for exploratory queries, AI Overviews are deployed by default and withheld only when organic results already suffice or source quality is too low; for targeted queries, deployment is rare and occurs only when the platform lacks confidence in the organic match. Across all intent types, deployment exhibits an inverted-U relationship with source quality. Our findings provide empirical evidence that AI Overview deployment varies strategically with search intent and that AIOs can be characterized as a novel form of platform self-preferencing, with implications for content creators, advertisers, and regulators concerned with platform market power.

Keywords: Generative AI, search intent, platform design, search advertising, zero-click search, analytical model, empirical study

---

*Robin Ng gratefully acknowledges support from the Deutsche Forschungsgemeinschaft (DFG) through the CRC TR 224, Project B05.

†Department of Economics and MaCCI, University of Mannheim; robin@robinng.com.

‡Chair of Business Information Systems, esp. E-Commerce and Digital Business, Friedrich Schiller University Jena; michael.wessel@uni-jena.de.

1 Introduction

Search engines organize information on the search engine results page (SERP) through a combination of organic listings, sponsored advertisements, and supplementary features such as knowledge panels and shopping carousels. Each of these elements represents a design choice about how to allocate user attention across competing information formats. Prior work has established that the arrangement of information on the SERP has first-order consequences for which links receive attention and traffic (Xu et al., 2012; Yang and Ghose, 2010). In May 2024, Google introduced AI Overviews (AIOs), which synthesize search results into AI-generated answers displayed on the SERP. Unlike prior SERP features that reorder or highlight existing content, AIOs generate entirely new platform-created content that competes with organic results on the same page by resolving queries directly on the SERP. The platform thus simultaneously controls the creation of the AI-generated content, the ranking of organic alternatives, and the monetization of the surrounding advertising space.

Early research on the implications of AIOs suggests that this feature contributed to the growing prevalence of zero-click searches, in which users obtain what they are seeking directly on the results page. Chapekis et al. (2025) find that users who encounter an AI summary click a traditional result at roughly half the rate of users who do not, and Khosravi and Yoganarasimhan (2026) estimate that AIO reduced daily traffic to English Wikipedia by approximately 15%. Thus, AIOs contribute to growing concerns about declining publisher traffic and revenue.[1] Google has already been found to have engaged in illegal monopolization of the search market (*United States v. Google LLC*, 2024), and legal scholars have argued that the integration of AIOs into search may help extend this dominance into AI-mediated information access (Singh, 2025).

While this evidence documents the downstream consequences of AIOs, we lack insights into the equally consequential question of when and how AIOs are deployed in the first place. In practice, Google does not deploy AIOs for every search query. Some queries consistently trigger an AI-generated summary, while others never or sometimes do. The key to understanding this selectivity lies in the heterogeneity of search intent. Not all search queries are alike: a user researching a topic generates very different revenue opportunities than a user ready to complete a purchase. We argue that the intent behind search queries differs along at least two dimensions that matter for platform revenue. First, queries vary in whether the search is exploratory or targeted toward a specific destination or action. Second, queries vary in their monetization capacity: some searches sit along a path toward a revenue-generating event, while others do not. These dimensions interact to create distinct deployment tradeoffs. For instance, if a user is exploring a non-monetizable domain (e.g., "Why is the sky blue?"), an AIO can resolve the query faster, freeing the user for future monetizable activity. Conversely, if a targeted search query provides opportunities for monetization through paid advertisements (e.g., "Buy iPhone 17"), an AIO may displace attention from sponsored results and reduce advertising revenue. We argue that AIOs are best understood as a strategic instrument

[1] Penske Media (publisher of Rolling Stone, The Hollywood Reporter, and other outlets) and Chegg have filed suit alleging that Google's AIOs caused substantial traffic and revenue losses. In the EU, a group of independent publishers have brought a competition complaint seeking interim measures. See Singh (2025) for an overview.

employed by the platform to maximize average revenue per user (ARPU), and that the decision of whether and how to deploy an AIO is driven by the same considerations that govern other aspects of SERP design: the tradeoff between improving user experience to retain engagement and preserving advertising revenue. This motivates our research question:

How does search intent shape a search platform's strategic deployment of AI Overviews, and are deployment patterns consistent with revenue maximization?

To address this question, we build a theoretical framework in which a monopolist search platform decides whether to deploy AIOs across search queries that differ in their search intent. We classify queries into four intent types, informational, navigational, commercial, and transactional, each occupying a distinct position in the space defined by exploration and monetization capacity. Each intent type carries distinct implications for platform revenue, and deploying an AIO is costly but can improve intent fulfilment and accelerate the arrival of monetizable intents. For each intent type, we characterize when the platform benefits from deployment and derive comparative statics that generate testable hypotheses.

To test the model's predictions, we construct a novel dataset of over 2,000 Google Search queries spanning the four intent types. For each query, we collect multiple SERPs using a web scraper, alternating across device types, days, and times to capture within-query variation in deployment. The final dataset covers 15,118 observations nested within 2,057 query clusters, where AIOs appeared in 31.2% of searches.

Consistent with our theoretical predictions, we find that deployment patterns vary systematically across intent types in ways that align with revenue maximization rather than uniform fulfilment improvement. For exploratory, non-monetary queries, AIOs are deployed as the default strategy, withheld only when organic results are already sufficient or when source material is too poor to produce a reliable overview. For targeted, non-monetary queries, deployment is rare and triggered only when the platform lacks confidence in the organic match. For exploratory, monetary queries, deployment reflects a tradeoff between converting users toward transactions and preserving valuable advertising clicks. For targeted, monetary queries, deployment is rare and the dominant predictor is the platform's confidence in the intent classification itself. Across all intent types, deployment exhibits an inverted-U relationship with source quality: AIOs are withheld both when source material is too poor to produce a reliable summary and when high-quality organic results already satisfy the user, making AI synthesis redundant.

We contribute to the literature in several ways. First, to the best of our knowledge, this is the first paper to model and empirically examine the platform's strategic deployment of AI-generated content within search results. A growing literature studies how users respond to generative AI in search (Zhu et al., 2025; Spatharioti et al., 2025; Kaiser et al., 2025; Liang et al., 2025), but these studies treat AI deployment as exogenous. We show that the queries for which AI is deployed are not randomly selected, and that understanding user-side effects requires first understanding the platform's upstream deployment decision. Second, we contribute to the literature on platform design and directed search (Hagiu and Jullien, 2011; Fradkin, 2017; Lee and Musolff, 2025) by characterizing a novel mechanism through which platforms shape user outcomes: the selective provision of AI-synthesized information that competes with and potentially displaces organic results. Unlike conventional self-preferencing, in which platforms favor their own products through ranking or placement (Zou and Zhou, 2025; Long and Amaldoss, 2024), or copy third-party innovations (Choi et al., 2025), AIOs represent a form of self-preferencing in which the platform synthesizes third-party content into a new competing product displayed on the same page. Third, we provide the first large-scale empirical analysis of AIO deployment patterns, contributing SERP-level measures of organic fulfilment, domain concentration, and source quality that can be applied to future studies of search engine behavior.

2 Related Literature

2.1 Platform Design and Directed Search

A substantial body of work examines how platforms design their marketplaces to direct user search and shape market outcomes. Hagiu and Jullien (2011) characterize when a platform with affiliated stores would divert search to maximize its profits, showing that diversion is profitable when the platform can capture a larger share of the surplus from affiliated transactions. Fradkin (2017) and Lee and Musolff (2025) estimate structural search models on Airbnb and Amazon, respectively, finding that directed search significantly affects market outcomes: without directed search, accepted inquiries on Airbnb would fall by 68% (Fradkin, 2017), and 95% of consumers on Amazon respond to directed search mechanisms (Lee and Musolff, 2025). More recent work treats discoverability itself as a strategic platform choice, showing that platforms face tradeoffs in how much discovery to enable and how that affects transactions, participation, and platform value (Hagiu and Wright, 2024). Taken together, this literature shows that platform design choices are not merely presentational but shape the transactions that occur.

In the context of web search specifically, several studies examine how SERP design affects user behavior and market outcomes. Ahn et al. (2018) use an eye-tracking experiment to demonstrate the importance of result display order for capturing consumer attention, emphasizing the outsized returns to being displayed first. Agarwal et al. (2015) study how competition from organic results affects sponsored ad performance, finding that competition decreases click-through rates but increases conversion rates conditional on click, a result they attribute to competition-induced consumer learning. Recent work on self-preferencing similarly shows that platform-controlled rankings, recommendations, and sponsored placement need not operate neutrally across sellers or products, but can instead shape competition, platform profit, and consumer outcomes in strategically favorable ways (Zou and Zhou, 2025; Long and Amaldoss, 2024). These studies highlight that the arrangement of information on the results page has first-order consequences for which links receive attention and traffic. Relatedly, Choi et al. (2025) show that platforms may copy successful third-party innovations, reducing complementors' incentives to innovate, though Foerderer et al. (2018) find that platform entry into complementary markets can also increase innovation through attention spillover effects.

AIOs can be understood as an evolution of zero-click search features, in which the platform provides answers directly on the results page rather than directing users to external websites. The consequences of such features for the broader content ecosystem have been studied in related contexts. Using the shutdown of Google News in Spain, Calzada and Gil (2020) and Athey et al. (2021) find that news aggregation negatively affects smaller and local news outlets, reducing their traffic by 8% to 14%. Others explore zero-click purchase decisions, such as Amazon's BuyBox, highlighting their significant role in steering demand (L. Chen et al., 2016; Raval, 2022). However, AIOs differ from these predecessors in an important way: rather than reordering or aggregating existing content, the platform generates entirely new content that competes with organic results. This distinction is consequential because the platform simultaneously controls the creation of the AI-generated content, the ranking of organic alternatives, and the monetization of the surrounding advertising space. Emerging evidence confirms that the downstream consequences are substantial: Khosravi and Yoganarasimhan (2026) estimate that AI Overviews reduced daily traffic to English Wikipedia by approximately 15%, and Aral et al. (2026) find that AI search reduces both the diversity and the credibility of sources to which users are exposed. Understanding whether the deployment of this content reflects user welfare considerations or revenue optimization therefore requires modeling the platform's deployment decision explicitly, which is the gap our paper addresses.

2.2 Search Advertising and Platform Monetization

Our theoretical model centers on the tradeoff between fulfilment improvement and advertising revenue displacement. This tradeoff is grounded in a rich literature on search advertising economics. Edelman et al. (2007) and Varian (2007) characterize the generalized second-price auction that governs how advertisers bid for placement on the SERP, establishing the microeconomic foundations of search engine revenue. Ghose and Yang (2009) estimate the relationship between ad position and click-through and conversion rates, finding that advertisers benefit from more prominent placement but that the relationship is nonlinear. Agarwal et al. (2011) study position effects in sponsored search and show that ads in higher positions receive more clicks but that conversion rates may decline, implying that the revenue-maximizing position depends on the interaction between attention and purchase intent.

A closely related question is how organic and sponsored results interact on the same page. Yang and Ghose (2010) analyze this relationship and find that the interdependence between organic and sponsored search advertising can be positive, negative, or zero depending on the keyword and product category, establishing that changes to the organic environment have non-trivial consequences for advertising revenue. Xu et al. (2012) study the effects of organic listings on sponsored search performance directly, showing that the presence and quality of organic results affect how users engage with paid advertisements. J. Chen and Stallaert (2014) show that targeted advertising can either increase or decrease publisher revenue, and Moshary (2025) provides causal evidence that sponsored search can cannibalize organic listings yet remain profitable for the platform because advertising revenue offsets lost commissions. Together, these studies demonstrate that organic and sponsored results do not operate in isolation: any intervention that alters the organic environment, such as inserting an AI-generated summary, will have spillover effects on advertising performance.

These studies characterize how the platform monetizes user attention through advertising, but they treat the SERP layout as fixed: organic results, sponsored results, and their relative positions are taken as given. The introduction of AIOs disrupts this assumption. When the platform inserts an AI-generated summary at or near the top of the results page, it alters the attention allocation across all other SERP elements, including the sponsored links from which the platform derives revenue. This creates a tension that the existing advertising literature does not address: deploying an AIO may improve user experience and accelerate fulfilment, but it may simultaneously cannibalize the advertising clicks that generate per-query revenue. Our paper formalizes this tension by modeling the platform's choice to deploy or withhold AIOs as a function of the advertising value at stake, and tests whether observed deployment patterns are consistent with this revenue calculus.

2.3 Generative AI in Search

A growing body of work examines how users interact with generative AI search tools. In a field experiment on an ecommerce platform, Zhu et al. (2025) show that deploying generative AI leads consumers to refine their search queries and reduces browsing between merchants. Spatharioti et al. (2025) show that generative AI search can increase the speed of user decision-making without sacrificing outcomes when the AI output is correct; however, users overrely on AI output, leading to worse decisions when the output is incorrect. Kaiser et al. (2025) and Liang et al. (2025) compare task completion behavior of users of generative AI versus standard web search, finding that generative AI users completed tasks more quickly but were less likely to validate results against primary sources.

These studies contribute important evidence on the demand side of AI search, demonstrating that AI features meaningfully alter how users process information and make decisions. However, they share a common limitation: the deployment of AI is treated as exogenous. The AI feature is either available to the user or it is not, and the research question concerns user responses. Yet the platform's decision of when and where to deploy generative AI is itself strategic, shaped by the same revenue considerations that govern other aspects of SERP design. This matters empirically because the queries for which AI is deployed are not randomly selected; they are chosen precisely because the platform expects deployment to be profitable. Understanding the downstream effects of AI search on users and content creators therefore requires first understanding the upstream deployment decision, which is the focus of our paper.

2.4 Search Intent and Query Heterogeneity

Early research on web search assumed that online search was primarily driven by informational needs, but observed that different search queries led to vastly different outcomes, presenting a puzzle that motivated work on search query taxonomy (Byrne et al., 1999; Broder, 2002). Broder

(2002) classifies search into three categories: informational queries to research and acquire information, navigational queries to reach a specific webpage, and transactional queries to complete a task such as a purchase. Jansen et al. (2008) operationalizes these definitions and provides an algorithm for automated classification. Dai et al. (2006) focuses on commercial search intents, distinguishing between an exploration phase, in which users research products and compare options, and a commitment phase, in which users seek to complete a specific transaction.

These taxonomies have proven useful for understanding user behavior and for designing search algorithms, but they were developed to classify user needs rather than to analyze platform strategy. In particular, existing classifications do not map cleanly onto the revenue dimensions that determine whether the platform benefits from deploying AI content. Broder's three-way classification distinguishes queries by what the user wants but does not distinguish between queries that generate advertising revenue and those that do not. Similarly, Jansen et al.'s operationalization provides a useful algorithm for automated labeling but does not consider how each category relates to the platform's monetization structure.

Building on this literature, we propose a two-dimensional classification of search intent along exploration and monetization capacity. Navigational and transactional intents reflect targeted searches where the user has a specific destination or action in mind, while informational and commercial intents capture exploratory searches involving research and information acquisition. Orthogonally, commercial and transactional intents involve directly monetizable activity, while informational and navigational intents do not generate direct transaction revenue. This classification bridges the gap between the user-centric taxonomy literature and the platform strategy question we study. It is central to our theoretical framework because the platform's optimal deployment strategy differs qualitatively across quadrants: the tradeoffs governing deployment for an exploratory, non-monetary query are different from those for a targeted, monetary query. In the former case, AIO accelerates the arrival of monetizable intents at the cost of production; in the latter, AIO risks obstructing a revenue-generating transaction. Table 1 summarizes this classification.

Table 1: Search intent classification by exploration and monetization capacity.

	Exploratory	**Targeted**
Non-monetary	Informational	Navigational
Monetary	Commercial	Transactional

The decision to deploy AIOs is also related to the quality of search results and the nature of the query. We draw on Search Quality Evaluator Guidelines for Google, which describe multiple dimensions along which search quality is assessed.[2] This is especially relevant as Gong et al. (2018) shows the same observed query may still reflect heterogeneous underlying user interests, making ambiguity economically consequential. To address this, we consider the Needs Met framework, which rates how well results satisfy the user's query; the importance of user location for locally

[2] Available at `https://static.googleusercontent.com/media/guidelines.raterhub.com/en/searchqualityevaluatorguidelines.pdf`, accessed November 2025.

relevant searches; Your Money or Your Life (YMYL) topics, where incorrect information poses elevated risk of harm and Google applies a higher quality bar (Google, 2025); and the challenge of mixed-intent queries where classification confidence is low. These quality dimensions inform both our measurement strategy and our control variables, ensuring that our empirical tests of strategic deployment account for the platform's own stated quality considerations.

3 Model

We consider an infinite-horizon discrete-time model ($t = 0, 1, 2, \ldots$) in which a monopolist search platform serves a unit mass of users who arrive with one of the following search intents: **Informational (I)**, where the user seeks to learn something (e.g., "Why is the sky blue?"). **Navigational (N)**, where the user seeks to reach a specific website (e.g., Netflix). **Commercial (C)**, where the user seeks to research products/services before a potential transaction (e.g. "Best phone camera"). **Transactional (T)**, where the user seeks to complete a transaction (e.g., Buying an iPhone 17 Pro Max 2 TB). Let $\mathbb{H} = \{I, N, C, T\}$ denote the set of possible intent states, let 0 be the state in which the user draws a new intent, and let Z denote the terminal state in which the user exits the platform. Each intent $h \in \mathbb{H}$ arrives with some probability P_h that is independent of the previous intent, where $\sum_{h\in\mathbb{H}} P_h < 1$.[3] With complementary probability, $1 - \sum_{h\in\mathbb{H}} P_h$, the user exits the platform. An intent h is fulfilled with probability $f_h \in (0, 1)$. If the intent is unfulfilled, in the next period, the user continues their search. This occurs until the intent is fulfilled. Users may only perform one search per period and only generate a new intent if the existing intent is fulfilled.

The platform discounts future payoffs at the rate $\delta \in (0, 1)$. Its actual payoff in each period depends on user search intent: Informational searches generate no revenue. Transactional searches generate an acquisition fee $s > 0$ when fulfilled. Navigational searches generate a payoff $e > 0$ when fulfilled and the user exits the platform.[4] Commercial searches generate a per-click advertising revenue $a > 0$, where $a < s$, in every period where the user has such an intent.[5] When the user exits, the state becomes Z and the continuation payoff is $V_Z = 0$. Table 2 summarizes the payoffs and presents the value functions for each search intent.

The expected payoff of the platform when a new intent is generated is $V_0 = \sum_{h\in\mathbb{H}} P_h V_h = \frac{e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+\lambda_T+\lambda_C)}$, where $\lambda_h = \frac{P_h f_h}{1-\delta(1-f_h)}$.

Lemma 1. *In equilibrium, the expected payoff is increasing in P_h for all $h \in \mathbb{H}$; and increasing in f_I, f_N, f_T, and increasing in the payoffs e, s, a. Further, $\frac{\partial V_0}{\partial f_C} \geq 0 \Leftrightarrow V_0 \geq \frac{a}{1-\delta}$.*

Lemma 1 establishes baseline results for the model. The results are largely intuitive: a plat-

[3] In practice, search intents might be correlated, we make the independence assumption for tractability.

[4] The payoff e is the expected payoff equivalent to a user navigating to a different part of the search engine's ecosystem and hence a continuation payoff or outside the ecosystem leading to zero payoff. In either case, the user stops their search activity.

[5] Having $a < s$ reflects how per-click fees are typically lower than acquisition fees. If this condition is violated, results about when the platform adopts AI for commercial transactions hold only under a set of stricter conditions. All other results are unaffected.

Table 2: Platform value functions by user search intent.

Intent	Payoff	Next user intent	Value function
Informational	0	new intent or exit	$V_I = \delta((1-f_I)V_I + f_I V_0)$
Navigational	e if fulfilled	user exits	$V_N = f_N e + \delta(1-f_N)V_N$
Commercial	a each period	new intent or exit	$V_C = a + \delta((1-f_C)V_C + f_C V_0)$
Transactional	s if fulfilled	new intent or exit	$V_T = f_T s + \delta((1-f_T)V_T + f_T V_0)$

form's payoff increases when the arrival probabilities, payoffs and fulfilment rates increase, all of which creates more opportunities for monetizable intents. The exception is commercial fulfilment. Commercial fulfilment creates more opportunities for monetizable intents to occur, a positive effect on payoffs. However, because the platform earns a positive payoff in every period where users have a commercial intent, fulfilment of commercial intents can have a negative effect on payoffs. The net effect depends on whether the expected value of a new intent is larger than the perpetuity value of the commercial payoff.

3.1 AI Overview

The platform may additionally choose to deploy an AI Overview feature for each search intent at some marginal cost $d > 0$. When users interact with the AI Overview, the probability of fulfilment is f_h^{AI}. Additionally, AI Overview affects outcomes differently depending on the user's intent: For informational searches, AI fulfils the intent with some probability f_I^{AI}. For navigational searches, AI does not change the probability of fulfilment, $f_N^{AI} = f_N$.[6] For transactional searches, deploying AI alters the probability of the first fulfilment to f_T^{AI}. If this fails, then the user continues searching beyond the AI Overview and the probability of fulfilment returns to f_T. For commercial searches, AI Overview generates no advertising revenue but can direct a user's intent towards transactional with some probability f_{ct}, otherwise a new intent is drawn.[7] Table 3 summarizes the effects of AI and the resulting value functions.

To understand the effects of adopting AI Overview, consider each change in isolation.

Informational AI When the platform adopts AI Overview for informational search intents, its expected payoff is $V_0^I = \frac{-d\frac{P_I}{1-\delta(1-f_I^{AI})}+e\lambda_N+s\lambda_T+\frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I^{AI}+\lambda_T+\lambda_C)}$. The platform adopts AI if and only if $V_0^I > V_0$.

Proposition 1. *The platform adopts AI Overview for informational search intents if and only if*

$$d < V_0 \frac{\delta(1-\delta)(f_I^{AI} - f_I)}{1-\delta(1-f_I)}.$$

[6]In practice, AI Overview may change the probability of fulfilment. However, individuals end up exiting the platform which means, in equilibrium, the AI Overview would never be employed. This assumption helps with interpretation, but we do not formally require it.

[7]It is possible for the AI Overview to fail to generate a transactional intent after fulfilling the commercial intent. Likewise, it is possible that the platform earns some advertising revenue from the AI Overview. Including these does not qualitatively affect the predictions in our main hypotheses, and we discuss their effects with Hypothesis 3.

Table 3: Effects of deploying AI on value function.

Intent	**AI effect**	**Value function**
Informational	fulfilment improves, f_I^{AI}	$V_I^{AI} = -d + \delta((1 - f_I^{AI})\hat{V}_I + f_I^{AI}\hat{V}_0)$
Navigational	no change, $f_N^{AI} = f_N$	$V_N^{AI} = -d + \hat{V}_N$
Commercial	no payoff; move to state T w.p. f_{ct}	$V_C^{AI} = -d + \delta((1 - f_{ct})\hat{V}_0 + f_{ct}\hat{V}_T)$
Transactional	first fulfilment w.p. f_T^{AI} subsequent fulfilment w.p. f_T	$V_T^{AI} = -d + (1 - f_T^{AI})\delta\hat{V}_T + f_T^{AI}(s + \delta\hat{V}_0)$

Notes: $\hat{V}_h$ $\forall h$ represents the continuation value under the relevant platform policy. $f_{ct} \in (0, 1)$.

This means AI Overviews are only deployed if $f_I^{AI} > f_I$.

The right-hand side of the condition represents the benefit of accelerating the fulfilment of informational searches. Without AI, failing to fulfil informational searches delays the arrival of monetizable intents. AI eliminates this delay. When f_I is larger, that is, the probability of fulfilment increases, then the right-hand side decreases, which means the platform is less likely to adopt AI Overview for informational search intents.

Hypothesis 1. *AI Overviews are more likely for informational queries with lower fulfilment rates.*

Navigational AI When the platform adopts AI Overview for navigational search intents, its expected payoff is $V_0^N = \frac{-P_N d + e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+\lambda_T+\lambda_C)}$.

Proposition 2. *The platform never adopts AI Overview for navigational search intents.*

Observe that we can write $V_0^N = \frac{-P_N d}{1-\delta(\lambda_I+\lambda_T+\lambda_C)} + V_0$. Since $d > 0$, $V_0^N < V_0$. In practice, it is possible that AI Overviews are deployed when users have navigational search intents due to imperfect intent classifications by the platform.

Hypothesis 2. *AI Overviews are unlikely for navigational queries.*

Commercial AI When the platform adopts AI Overview for commercial search intents, its expected payoff is $V_0^C = \frac{e\lambda_N + s\lambda_T + P_C(-d + s\delta\frac{f_{ct}f_T}{1-\delta(1-f_T)})}{1-\delta(\lambda_I+\lambda_T+P_C(1-f_{ct}+\delta\frac{f_{ct}f_T}{1-\delta(1-f_T)}))}$.

Proposition 3. *The platform adopts AI Overview for commercial search intents if and only if* $V_0^C > V_0$,

$$d < \underbrace{\delta V_0(1 - f_{ct} + \frac{\delta f_{ct} f_T}{1-\delta(1-f_T)} - \frac{f_C}{1-\delta(1-f_C)})}_{\text{indirect effect}} + \underbrace{\frac{\delta s f_{ct} f_T}{1-\delta(1-f_T)} - \frac{a}{1-\delta(1-f_C)}}_{\text{direct effect}}.$$

The direct effect shows the net effect of the loss in advertising revenue due to users interacting with the AI Overview rather than organic links and the potential gain of directing users directly to a transactional intent. The indirect effect captures the change in potential continuation value because of the probability of new intents being drawn.

Unlike the previous cases, the sign of the continuation value is ambiguous. This means the platform may deploy AI Overviews even when the continuation value decreases. A sufficient condition for this is when the direct effect is sufficiently large, i.e., the acquisition fee s is sufficiently large relative to the advertising revenue a.

Hypothesis 3. *AI Overviews are more likely for commercial search intents when the ratio of acquisition fee to advertising revenue is sufficiently large.*

In practice, there are other considerations when deploying commercial AI Overviews. First, the AI Overview may fail to generate a transactional intent after fulfilling the user's commercial intent. This decreases the value of deploying AI Overviews, and the platform will only deploy AI for a larger $s : a$ ratio. Second, if it is possible to embed advertising into the AI Overview, this would decrease the loss in advertising revenue due to users interacting with the AI Overview. This means the opportunity cost of deploying commercial AI Overview decreases, and the platform is more likely to do so. Therefore, the platform will deploy AI for a smaller $s : a$ ratio.

Transactional AI When the platform adopts AI Overview for transactional search intents, its expected payoff is $V_0^T = \frac{e\lambda_N + s(P_T f_T^{AI} + (1-f_T^{AI})\delta\lambda_T) - P_T d + \frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I + P_T f_T^{AI} + (1-f_T^{AI})\delta\lambda_T + \lambda_C)}$.

Proposition 4. *The platform adopts AI Overview for transactional search intents if and only if*

$$d < \frac{(1-\delta)(f_T^{AI} - f_T)}{1-\delta(1-f_T)}(s + \delta V_0).$$

Notice that a necessary, but not sufficient, condition for transactional AI Overview is $f_T^{AI} > f_T$. Intuitively, if AI Overview does not increase the chances of fulfilment, then it only delays the arrival of other monetization opportunities. However, this alone is not sufficient, and the likelihood of success must be high enough such that the expected gains from future monetization exceeds the cost of deploying the AI Overview.

Hypothesis 4. *AI Overviews are more likely for transactional search intents when deployment improves the probability of fulfilment.*

3.2 Extension

According to Google, AI Overviews are designed to present only information corroborated by high quality web results, explicitly distinguishing them from standalone LLM responses (Google, 2025). This retrieval augmented architecture means that AIO quality is constrained by the quality of available source material. Google further states that it withholds AIOs when high quality web content is unavailable, a situation it refers to as "data voids." To understand this, suppose that $f_h^{AI} \equiv \Psi(f_h)$, where $\Psi'(f_h) \geq 0$. This captures how having good underlying search results, which are more likely to fulfil the search intent, can help AIOs perform better. Otherwise, there is a data void and AIOs could garble information and potentially lead to lower fulfilment than organic search would provide.

This has two key implications: First, when the underlying search results are sufficiently poor, there is a data void and the deployment of AIOs leads to a low fulfilment probability. This reduces the platform's profits and it prefers not to deploy AI Overviews. Second, if the underlying search results are sufficiently good, then the benefit of deploying AI Overviews is too small, which would be unable to justify the cost of doing so. This provides the following hypothesis.

Hypothesis 5. *The deployment of AI Overviews follows an inverted-U shape with respect to the quality of the underlying search results.*

4 Empirical Analysis

Our theoretical model generates predictions about when Google should strategically deploy AI Overviews based on the tradeoff between user engagement benefits and advertising revenue displacement. We now turn to empirical analysis to test these predictions. In the model, deployment depends on organic fulfilment f_h, AIO capability f_h^{AI}, and revenue parameters a and s. Empirically, we proxy fulfilment using Needs Met ratings of organic results, AIO capability using average domain authority of source material, and revenue parameters using cost per click.[8] As the model derives distinct deployment logic for each intent type, we estimate separate specifications for each, with predictors that reflect the theoretical mechanism relevant to that quadrant of our framework (Table 2).

4.1 Data Collection

Testing our hypotheses requires observing AI Overview deployment decisions across search queries that vary in their search intent, organic result quality, advertising value, and user context. We construct a novel dataset combining varying search query characteristics with corresponding search engine results pages (SERPs) collected from Google in November 2025. Importantly, for our research question, what matters is coverage of the intent space, not representativeness of query frequency. We estimate conditional deployment probabilities within each intent type, not unconditional prevalence rates, so the sample must span the relevant dimensions of intent, topic sensitivity, and query format with sufficient variation.

A key challenge in studying search engine behavior is constructing a suitable query sample. Google does not publish query logs, and user generated datasets may exhibit selection bias toward popular or salient queries. We therefore employed a structured approach to generating search queries: we seeded a LLM (Claude Haiku 4.5, Anthropic) with high frequency keywords from a commercial search analytics API[9] covering the United States and Ireland, then prompted it to gen-

[8]Empirically, we are unable to have clean distinct measures for a and s, we discuss our identification strategy in Section 5.3.

[9]We use an industry leading provider for SEO data, see `https://dataforseo.com/apis`.

erate queries varying across search intent (informational, navigational, commercial, transactional), YMYL status (health, finance, or safety topics versus general), local relevance (queries containing geographic modifiers such as "near me"), question format, and length (one to ten words)[10]. This process yielded 33,082 candidate queries (see Appendix B for additional details on the query generation and validation process).

Each candidate was validated through a search analytics API, which classifies search intent based on Google's internal signals and returns an intent classification confidence score. We ranked candidate queries by intent classification confidence and selected those with the highest confidence scores within each intent category. The resulting sample exhibits high classification confidence across all categories, with mean confidence of 0.92 for both informational and navigational queries, and slightly lower values of 0.90 and 0.88 for commercial and transactional queries, respectively. We required minimum monthly search volume of 10 to ensure that each query reflects an actual search phrase used by real users rather than a synthetic combination without genuine search activity. After deduplication and filtering, 2,057 unique queries remained across the four intent categories: informational (567), navigational (600), commercial (364), and transactional (526). For each query, we also collected the average cost per click (USD) and average monthly search volume over the preceding 12 months, which we use as proxies for advertising value in our empirical specifications. Table 9 in Appendix B presents illustrative queries from each intent type, showing variation in topic sensitivity, local relevance, and observed AIO deployment rates.

For each query, we collected SERPs using a web scraper for Google search results. AI Overview deployment is not fixed at the query level; the same query may or may not trigger an AIO across repeated searches depending on Google's real time assessment of helpfulness and response quality (Google, 2025).[11] We therefore collected each query eight times across the study period, alternating between mobile and desktop device contexts as well as day of the week and time of day.

For local queries, we specified geographic coordinates through Google's UULE parameter (Los Angeles for US queries, Dublin for Irish queries). This procedure yielded 15,118 query device time observations nested within 2,057 query clusters. For each SERP, the scraper returns the full HTML page as well as structured data including organic results (URLs, titles, snippets, and positions), paid results, and related queries. Our analysis focuses on the deployment decision itself, that is, whether and where in SERP an AIO appears, rather than on the content of the AI-generated summary, which we do not observe.

4.2 Measures

Our primary dependent variable is *AIO deployment*, a binary indicator for whether Google displayed an AI Overview on the SERP (AI Overview = 1 in 31.2% of observations). AI Overviews were detected through HTML parsing of Google's distinctive CSS markers; detection accuracy was

[10]Ireland was selected as the representative for the EU GDPR region as it ensured a native English speaking population for a cleaner comparison.

[11]In our sample, AIOs were never generated for 47% of queries, sometimes generated for 44%, and generated in every instance for 9%.

verified through manual inspection. For supplementary analyses of the intensive margin, we record each AIO's position within the SERP hierarchy, classified into four ordinal categories by counting its rank among top level result containers: first (the very first element on the results page; 78.5% of deployed AIOs), top (appearing below a product carousel or featured snippet but above organic results; 10.8%), middle (appearing below paid results or below the first organic result; 7.9%), and low (embedded within the organic search results; 2.9%). This ordinal measure allows us to examine not only whether Google deploys an AIO but how prominently it features it, an intensive margin that reflects the platform's confidence in the AIO's value for a given query.

Our theoretical model requires a measure of baseline organic fulfilment f_h, capturing how well organic results satisfy the user's intent, independent of any AI Overview. We derive this measure following the Search Quality Evaluator Guidelines for Google, which specify detailed rating criteria and examples for each search intent type. We extracted all rules and illustrative examples from these guidelines and implemented a two stage classification approach. In the first stage, a LLM (Claude Haiku 4.5, Anthropic) analyzes each SERP and codes structured characteristics of the organic results relative to the inferred user intent. For instance, for a navigational query such as "netflix login," the model determines at what position in the organic results the actual Netflix login page appears, whether the target domain is present at all, and whether intermediate pages or competing sites obstruct direct access. For informational queries, the model assesses coverage completeness, source quality, and whether the results address the query from multiple relevant angles. In the second stage, a deterministic rule engine, developed based on the criteria specified in the guidelines, maps these coded features to a five point Needs Met scale: Fails to Meet (1), Slightly Meets (2), Moderately Meets (3), Highly Meets (4), and Fully Meets (5). This two stage design separates subjective judgment (feature extraction by the LLM) from the rating decision (deterministic rules), improving consistency and reproducibility (Carlson and Burbano, 2025). Crucially, the rating evaluates only organic results, excluding the AI Overview itself, ensuring the measure is not mechanically affected by AIO deployment (see Appendix C for additional details).

To capture the platform's confidence in the organic match, we compute a Herfindahl Hirschman Index (HHI) of domain concentration across the top 10 organic results for each SERP. We extract the base domain from each result URL and calculate HHI $= \sum_i \xi^2$, where ξ is domain i's share of the top 10 results. High concentration indicates organic results converge on a single domain, signalling a clear match for the user's query, while low concentration indicates dispersed results. We also compute the average domain authority across the top 10 organic search results, and use this as a measure of organic source quality. Domain authority predicts how likely a website is to rank in search engine results, with higher scores indicating more authoritative and widely-referenced sources.[12]

The search analytics API provides keyword-level advertising metrics that proxy for the revenue parameters in our model. Cost per click (CPC) is the realized average price in USD per click and

[12]We use data from an industry leading SEO tool which computes domain authority using backlink data, see `https://moz.com/learn/seo/domain-authority`.

thus captures the value advertisers place on a query. Average monthly search volume captures query popularity and the scale of the platform's revenue opportunity.

We control for mobile device (vs. desktop), YMYL classification, local query status, question format, and GDPR region (Ireland versus United States). YMYL (Your Money or Your Life) identifies queries where information quality is critically important because incorrect answers could affect users' health, financial stability, or safety; Google itself applies a higher quality bar for these queries (Google, 2025). To capture the presence of Google's own properties (e.g., YouTube) in organic results, we compute a weighted ecosystem score as $\sum_{j \in \mathcal{G}} 1/\text{rank}_j$ for Google owned domains $\mathcal{G}$ appearing in the top 10 results, representing an alternative fulfilment channel that may reduce the marginal value of deploying an AIO. When these properties rank prominently, Google may be more likely to retain users within its own ecosystem, reducing the need to deploy an AIO.

Intent classification confidence measures how clearly a query belongs to a single intent category. Some queries are unambiguously informational or navigational, while others sit at the boundary between categories. As confidence increases, the platform can more reliably apply the deployment strategy that our theoretical model predicts for that intent type; lower confidence implies intent ambiguity, which may cause the platform to hedge by blending strategies across intent types.

Table 4 reports summary statistics for the full sample and by intent type.

4.3 Empirical Model

Our theoretical model derives distinct deployment strategies for each intent type, determined by the interplay between organic fulfilment, AIO capability, and revenue considerations. We test these predictions using logistic regression, where the binary dependent variable indicates whether Google deployed an AIO on a given SERP. Because the same query is observed multiple times across device contexts and collection periods, we cluster standard errors at the query level for all specifications to account for within-query correlation. For each intent type, we estimate a specification with predictors that reflect the theoretical mechanism relevant to that quadrant of our framework (Table 1).

For informational queries (H1), our model predicts that AIO is the platform's default fulfilment strategy, deployed as long as the costs are sufficiently low. Proposition 1 shows that deployment is less likely as organic fulfilment f_I increases, because higher organic fulfilment narrows the gap between what organic results achieve and what AIO could add. The specification includes organic fulfilment as a linear term and average domain authority as a proxy for AIO production capability $\Psi(f_h)$:

$$\Pr(\text{AIO}_{it} = 1) = \Lambda\left(\beta_1 f_{it} + \beta_2 \text{DA}_{it} + \mathbf{X}_{it}' \boldsymbol{\gamma}\right) \tag{1}$$

where $\Lambda(\cdot)$ denotes the logistic function, f_{it} is organic fulfilment for query i at observation t, DA_{it} is average domain authority, and $\mathbf{X}_{it}$ is the vector of controls. We predict $\beta_1 < 0$ (AIO withdrawn when organic results are sufficient) and $\beta_2 > 0$ (deployment requires high-quality source material).

For navigational queries (H2), Proposition 2 establishes that the platform should never deploy

AIO, since deployment incurs cost d without improving fulfilment. In practice, the 10% deployment rate we observe suggests that deviations from this default occur when the platform lacks confidence in the organic match. To explain this conditional variation, we include two independent signals of organic confidence: fulfilment and domain concentration (HHI):

$$\Pr(\text{AIO}_{it} = 1) = \Lambda\left(\beta_1 f_{it} + \beta_2 \text{HHI}_{it} + \mathbf{X}'_{it}\boldsymbol{\gamma}\right) \tag{2}$$

We predict $\beta_1 < 0$ (poor organic results trigger the safety net) and $\beta_2 < 0$ (dispersed results signal an unclear match).

For commercial queries (H3), the deployment decision involves a tradeoff between the value of converting the user toward a transaction and the advertising revenue sacrificed by displacing organic clicks. Our model predicts that this tradeoff is captured by the interaction between organic fulfilment and cost per click:

$$\Pr(\text{AIO}_{it} = 1) = \Lambda\left(\beta_1 f_{it} + \beta_2 \ln \text{CPC}_i + \beta_3 f_{it} \times \ln \text{CPC}_i + \mathbf{X}'_{it}\boldsymbol{\gamma}\right) \tag{3}$$

The interaction β_3 is the key parameter. At low CPC, transaction value is modest and fulfilment should have little effect on deployment; at high CPC, higher organic fulfilment should reduce deployment because good organic results generate valuable advertising clicks that AIO would cannibalize. We predict $\beta_3 < 0$.

For transactional queries (H4), we estimate the same specification as for commercial queries to test whether the revenue tradeoff mechanism operates differently when the user already has transactional intent:

$$\Pr(\text{AIO}_{it} = 1) = \Lambda\left(\beta_1 f_{it} + \beta_2 \ln \text{CPC}_i + \beta_3 f_{it} \times \ln \text{CPC}_i + \mathbf{X}'_{it}\boldsymbol{\gamma}\right) \tag{4}$$

Our model predicts that AIO deployment should increase when AI improves the probability of fulfilment (Proposition 4), but unlike commercial queries, there is no exploration phase generating per-period advertising revenue. Whether this difference in the underlying mechanism produces a different empirical pattern is the key test.

Finally, the extension in Section 3.3 predicts that AIO deployment exhibits an inverted-U relationship with source quality (H5). We test this pooled across all intent types using a quadratic specification in average domain authority, with intent-type fixed effects to absorb the different baseline deployment rates:

$$\Pr(\text{AIO}_{it} = 1) = \Lambda\left(\beta_1 \text{DA}_{it} + \beta_2 \text{DA}^2_{it} + \boldsymbol{\alpha}'\mathbf{D}_i + \mathbf{X}'_{it}\boldsymbol{\gamma}\right) \tag{5}$$

where $\mathbf{D}_i$ is a vector of intent-type indicators (with informational as the omitted category). We predict $\beta_1 > 0$ and $\beta_2 < 0$, yielding an interior maximum: deployment first rises with source quality as AIO becomes feasible, then declines as high-quality organic results render AIO redundant. A significant quadratic term alone is not sufficient to establish an inverted-U (Haans et al., 2016);

we therefore apply the Lind and Mehlum (2010) procedure, which tests whether the slope is significantly positive at the lower bound and significantly negative at the upper bound of the data, and whether the estimated turning point lies within the observed range.

In addition to these deployment regressions (the extensive margin), we examine AIO positioning (the intensive margin) for informational and commercial queries using ordered logistic regression on the subsample of SERPs where an AIO was deployed. The dependent variable is the ordinal position (1 = first, 4 = low). If deployment and positioning reflect the same underlying confidence calculus, the predictors that increase deployment probability should also predict more prominent placement, allowing us to assess whether Google makes a single integrated decision or two distinct ones.

5 Results

Table 4 reports summary statistics for the full sample and by intent type. AI Overviews appear on 31.2% of SERPs overall, but deployment varies sharply across intent types: informational (50%) and commercial (51%) queries receive AIOs at roughly five times the rate of navigational (10%) queries and more than twice the rate of transactional (20%) queries. This pattern is consistent with the theoretical framework's central prediction: the default for exploratory queries is to deploy AIOs, while the default for targeted queries is to withhold them.

The summary statistics also reveal patterns in the explanatory variables that align with the theoretical framework. Domain concentration is highest for navigational queries (mean HHI = 0.31), where organic results naturally converge on a single target domain, and lowest for informational queries (mean HHI = 0.15), where many sources can address the same question. Cost per click is highest for commercial queries ($7.75), consistent with the high advertising value of users in the exploration phase of a purchase decision, and lowest for informational queries ($1.63), where direct monetization opportunities are limited.

Several control variables exhibit consistent effects across specifications in Table 5. These merit a brief discussion before turning to the hypothesis tests. Mobile device context increases AIO deployment across all intent types. This is consistent with the more constrained attention environment on mobile devices, where limited screen real estate amplifies the importance of the first result displayed (Ahn et al., 2018; Ghose et al., 2013). A deployed AIO can occupy half or more of the visible area even before the user clicks "show more" to expand it, displacing organic results below the fold and substantially increasing the likelihood of zero-click engagement. Local queries strongly reduce AIO deployment across the board, reflecting the difficulty of generating reliable AI summaries for geographically specific needs. GDPR region is associated with lower deployment, possibly reflecting regulatory caution or differences in source material availability. Intent classification confidence, although not always statistically significant, consistently points in the predicted direction: positive for informational and commercial queries, negative for navigational and transactional queries. As the platform becomes more certain about the user's intent, it more reliably applies the deployment strategy our model prescribes. We perform subsample analyses by intent

Table 4: Summary statistics.

	All	Info.	Nav.	Comm.	Trans.
AIO deployment	0.31	0.50	0.10	0.51	0.20
	(0.46)	(0.50)	(0.31)	(0.50)	(0.40)
AIO position (1-4)	1.35	1.29	1.25	1.37	1.56
(1=First; 4=Low)	(0.75)	(0.70)	(0.63)	(0.78)	(0.84)
Organic fulfilment (1-5)	3.44	3.53	3.63	3.37	3.18
(1=Fails; 5=Fully meets)	(0.95)	(0.71)	(1.18)	(0.94)	(0.80)
Domain concentration (HHI)	0.21	0.15	0.31	0.18	0.18
	(0.20)	(0.091)	(0.29)	(0.15)	(0.13)
Domain authority	68.52	72.70	65.97	71.01	65.19
	(16.94)	(12.49)	(21.24)	(13.78)	(16.42)
Cost per click (USD)	4.51	1.63	6.57	7.75	2.66
	(11.74)	(2.73)	(16.19)	(14.90)	(6.54)
Search volume ('000)	308.18	41.64	1010.38	11.95	8.66
	(6,137.09)	(250.14)	(11,374.56)	(62.56)	(66.33)
Mobile device (1=Yes)	0.48	0.48	0.47	0.48	0.48
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
YMYL classification (1=Yes)	0.35	0.29	0.42	0.43	0.29
	(0.48)	(0.45)	(0.49)	(0.50)	(0.45)
Local query (1=Yes)	0.40	0.46	0.48	0.16	0.41
	(0.49)	(0.50)	(0.50)	(0.37)	(0.49)
Question query (1=Yes)	0.27	0.61	0.047	0.17	0.23
	(0.44)	(0.49)	(0.21)	(0.38)	(0.42)
GDPR region (1=Yes)	0.44	0.44	0.47	0.45	0.41
	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)
Google ecosystem score	0.059	0.051	0.12	0.039	0.013
	(0.27)	(0.19)	(0.44)	(0.16)	(0.075)
Intent classification confidence	0.91	0.92	0.92	0.90	0.88
	(0.070)	(0.037)	(0.042)	(0.10)	(0.086)
Observations	15,118	4,113	4,376	2,739	3,890

Notes: Mean values with standard deviations in parenthesis. AIO position is only for the subsample of search queries with AIO, total observations 4,715. CPC was unavailable for 41% of informational, 34% of navigational, 29% of commercial and 28% of transactional observations; total observations with CPC data: 10,071.

classifications to study each hypothesis in turn.

5.1 Hypothesis 1: Informational Queries

Column (1) of Table 5 reports results for informational queries. Consistent with the theoretical prediction, organic fulfilment has a negative coefficient (-0.20, $p = 0.048$), indicating that AIO deployment declines as organic results improve. However, the marginal effects in Figure 1 reveal that this decline is gradual and AIO deployment remains above 45% across the entire fulfilment

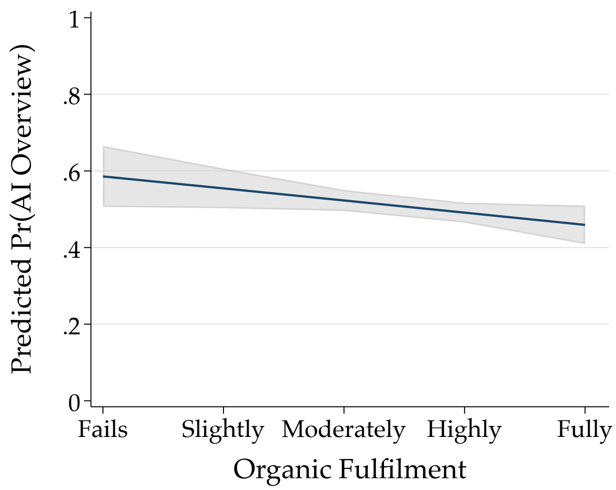
Table 5: Probability of AIO deployment.

	(1)	(2)	(3)	(4)
	Probability of AI Overview			
	Informational	**Navigational**	**Commercial**	**Transactional**
Organic fulfilment	-0.20**	-0.45***	0.21	0.18
	(0.10)	(0.086)	(0.19)	(0.17)
Domain authority	0.029***			
	(0.0059)			
HHI		-4.25***		
		(0.99)		
ln(CPC)			0.98***	0.32
			(0.37)	(0.39)
ln(CPC) $\times$ organic fulfilment			-0.19**	-0.030
			(0.097)	(0.12)
ln(search volume)	-0.018	-0.061**	-0.032	-0.021
	(0.029)	(0.029)	(0.049)	(0.044)
Mobile device	0.84***	0.32***	0.16*	0.58***
	(0.097)	(0.10)	(0.087)	(0.10)
YMYL	0.70***	0.13	0.29	0.19
	(0.19)	(0.22)	(0.28)	(0.24)
Local query	-2.22***	-2.02***	-0.78	-1.66***
	(0.19)	(0.26)	(0.53)	(0.37)
Question query	0.18	-0.42	1.34***	1.26***
	(0.16)	(0.49)	(0.31)	(0.25)
GDPR region	-0.40**	-0.98***	-0.22	-.88***
	(0.19)	(0.23)	(0.22)	(0.25)
Google ecosystem score	-0.73**	0.23	-0.87	1.17
	(0.33)	(0.21)	(0.69)	(0.98)
Intent classification confidence	2.93	-6.34**	2.25	-4.75**
	(1.82)	(2.47)	(1.71)	(1.93)
Constant	-3.40**	7.67***	-3.13*	2.19
	(1.69)	(2.39)	(1.86)	(1.93)
Observations	4,111	4,374	1,939	2,817

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Logistic regression results for probability of AIO deployment. Standard errors clustered at the query level.

Figure 1: Predicted probability of AIO deployment for informational queries.



Notes: Predicted probabilities from logistic regression (Table 5, column 1).

range. This is consistent with our prediction that the platform treats AIO as its default fulfilment strategy for informational queries, withdrawing it only when organic results render it redundant.

Average domain authority, our proxy for AIO production capability $\Psi(f_h)$, is strongly positive (0.029, $p < 0.001$). This confirms the capability gate implied by the extension in Section 3.2: the platform requires high-quality source material to produce a reliable AIO. Google itself describes AI Overviews as retrieval-augmented summaries that depend on high-quality web content (Google, 2025), and the data bear this out. A one-standard-deviation increase in domain authority (12.5 points) is associated with a 0.35-unit increase in the log-odds of AIO deployment.

The Google ecosystem score is negative and significant (-0.73, $p = 0.026$), indicating that when Google's own properties (e.g., YouTube, Google Maps) already appear in the organic results, AIO deployment decreases. This suggests that Google's own properties serve as an alternative fulfilment channel: when the platform is already providing direct answers through its ecosystem, AIO is partially redundant.

Column (1) of Table 6 examines AIO positioning conditional on deployment. Higher organic fulfilment pushes AIO further down the SERP (0.31, $p = 0.067$), indicating that when organic results are strong, Google still deploys AIO but positions it less prominently, letting organic results take priority. This two-stage decision, first whether to deploy, then how prominently to feature it, reflects a consistent confidence calculus: the same factors that reduce deployment probability also reduce placement prominence. Being on a mobile device also pushes down the position of AIOs, on average AIOs appear in the second position, above organic search results but below advertisements. Together with findings on probability of AIO on mobile devices, this highlights the strategic implementation of AIOs when screen real estate is limited.

Table 6: AIO positioning.

	(1)	(2)
	Position of AI Overview	
	Informational	**Commercial**
Organic fulfilment	0.31*	0.72**
	(0.17)	(0.28)
Domain authority	0.011	
	(0.0086)	
ln(CPC)		0.77**
		(0.34)
ln(CPC) × organic fulfilment		-0.35***
		(0.12)
ln(search volume)	-0.027	-0.12**
	(0.043)	(0.053)
Mobile device	0.97***	0.25
	(0.22)	(0.18)
YMYL	-1.00***	-1.08***
	(0.31)	(0.36)
Local query	1.96***	-2.24***
	(0.31)	(0.76)
Question query	-1.14***	-0.37
	(0.29)	(0.29)
GDPR region	-0.46	-0.66**
	(0.29)	(0.29)
Google ecosystem score	1.32*	-0.51
	(0.70)	(0.67)
Intent classification confidence	7.02**	-3.25
	(2.87)	(2.13)
Observations	2,076	987

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

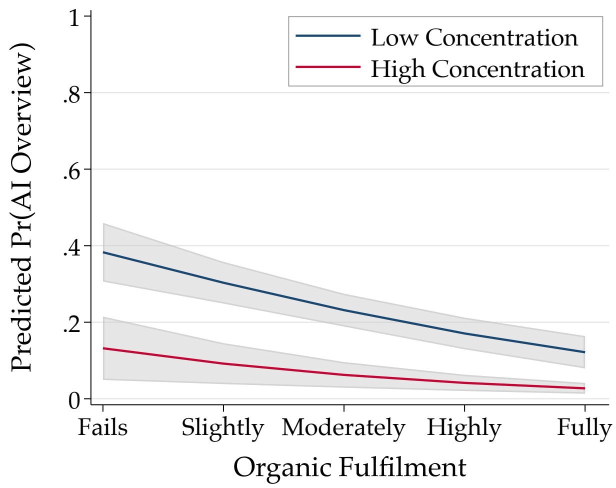
Notes: Ordered logistic regression results for position of AIO on SERPs. A larger position represents that the result appeared lower on the page. Standard errors clustered at the query level.

5.2 Hypothesis 2: Navigational Queries

Column (2) of Table 5 reports results for navigational queries. Proposition 2 predicts that the platform should never deploy AIO for navigational queries. Empirically, the baseline deployment rate is only 10%, far below informational (50%) or commercial (51%), confirming that the default is indeed no deployment.

When AIOs do appear, two independent signals explain the deviation. Organic fulfilment is strongly negative (-0.45, $p < 0.001$): when organic results fail to deliver the target website, the platform deploys AIO as a safety net. Domain concentration (HHI) is also strongly negative (-4.25, $p < 0.001$): when no single domain dominates the SERP, the match between query and destination is unclear, and the platform hedges by offering an AI summary.

Figure 2: Predicted probability of AIO deployment for navigational queries.



Notes: Predicted probabilities from logistic regression (Table 5, column 2). High domain concentration (bottom line): HHI = 0.5. Low domain concentration (top line): HHI = 0.1.

Figure 2 illustrates these effects jointly. At low domain concentration (HHI = 0.1, dispersed results), AIO probability declines from 0.38 at Fails to Meet to 0.11 at Fully Meets. At high domain concentration (HHI = 0.5, one domain dominates), AIO probability is near zero across all fulfilment levels, ranging from 0.14 to 0.03. Critically, both lines converge toward zero at the highest fulfilment level. When organic results clearly deliver the navigational target, whether through high fulfilment or high domain concentration, the platform steps aside entirely.

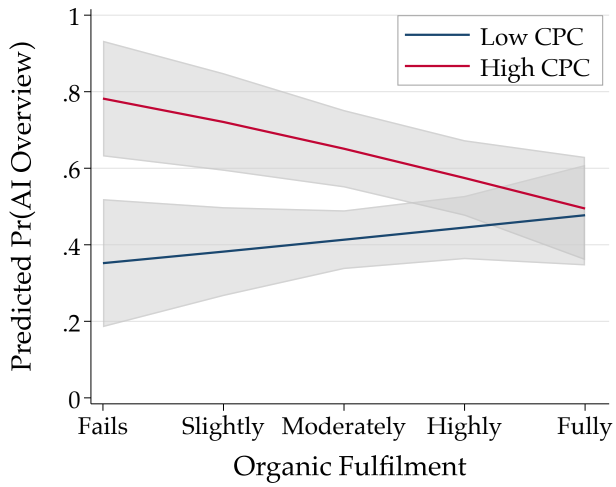
Intent classification confidence is strongly negative (-6.34, $p = 0.010$), indicating that when the platform is more confident the query is truly navigational, it is less likely to deploy AIO. This is consistent with the theoretical prediction: unambiguous navigational queries should never receive AIO, and deviations from this rule occur primarily when intent is uncertain.

5.3 Hypothesis 3: Commercial Queries

Column (3) of Table 5 reports results for commercial queries. Proposition 3 predicts that AIO deployment depends on the tradeoff between the value of converting the user toward a transaction and the advertising revenue displaced. The key test is the interaction between organic fulfilment and CPC.

The interaction is negative and significant (-0.19, $p = 0.049$). Figure 3 shows the implied pattern. At low CPC (10th percentile), AIO probability rises modestly with fulfilment, from 0.35 at Fails to Meet to 0.48 at Fully Meets. The transaction value is too low to justify conversion, so the

Figure 3: Predicted probability of AIO deployment for commercial queries.



Notes: Predicted probabilities from logistic regression (Table 5, column 3). High CPC (top line): 90th percentile. Low CPC (bottom line): 10th percentile.

platform lets users browse largely regardless of organic quality. At high CPC (90th percentile), AIO probability starts high (0.78 at Fails to Meet) and declines steeply to 0.49 at Fully Meets. When the transaction is valuable, the platform deploys AIO aggressively at low fulfilment to convert the exploring user toward a purchase, but withdraws as organic results improve, consistent with the platform protecting existing advertising revenue from displacement.

This pattern is consistent with AIO serving as an intent conversion engine for commercial queries. Rather than simply synthesizing information, commercial AIOs present curated product recommendations with linked product names that route the user back into Google Search with transactional queries, transforming exploring intent into targeted purchase intent. The platform only executes this conversion when the prize at the end of the funnel justifies the advertising revenue at risk along the way.

Column (2) of Table 6 examines AIO positioning conditional on deployment, using the same fulfilment × CPC specification. The deployment and positioning decisions reveal complementary logic. The fulfilment main effect is positive and significant (0.72, $p = 0.010$): higher organic fulfilment pushes AIO further down the SERP. CPC also pushes AIO down (0.77, $p = 0.025$). However, the interaction between the two is negative and significant (-0.35, $p = 0.005$), reversing the direction: when both CPC and fulfilment are high, the platform places AIO more prominently. This apparent tension with the deployment results reflects a selection effect. The deployment model shows that the platform becomes increasingly selective about deploying AIO as both CPC and fulfilment rise, filtering out cases where the risk of displacing valuable advertising is not justified. The observations that survive this filter are ones where the platform is confident the AIO adds value, and it signals this confidence by placing it prominently.

5.4 Hypothesis 4: Transactional Queries

Table 7: AIO deployment and source quality (pooled).

	(1) Probability of AI Overview
	All
Domain authority	0.094***
	(0.019)
Domain authority squared	-0.00071***
	(0.00014)
Navigational intent	-1.95***
	(0.14)
Commercial intent	-0.29**
	(0.12)
Transactional intent	-1.55***
	(0.12)
Device type	0.47***
	(0.042)
YMYL	0.72***
	(0.087)
Local query	-1.69***
	(0.12)
Question query	0.70***
	(0.093)
GDPR region	-0.45***
	(0.079)
Google ecosystem score	-0.34*
	(0.19)
Intent classification confidence	-1.88***
	(0.63)
Constant	-1.06
	(0.87)
Observations	15,066

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Logistic regression results for probability of AIO deployment over the full sample. Standard errors clustered at the query level. Higher-order domain authority terms were tested but not significant.

Column (4) of Table 5 reports results for transactional queries. Proposition 4 predicts that AIO deployment should increase when AI improves the probability of fulfilment. We test this using the same fulfilment-by-CPC interaction that captures the deployment logic for commercial queries, since CPC proxies for the transaction value s that makes AIO deployment worthwhile.

The interaction is entirely insignificant (-0.03, $p = 0.80$), and neither fulfilment (0.18, $p = 0.29$)

nor CPC (0.32, $p = 0.41$) significantly predicts deployment as main effects. We find no support for Hypothesis 4. The revenue tradeoff mechanism that drives commercial deployment, where AIO converts exploring intent into transactional intent in exchange for sacrificed ad revenue, does not apply to transactional queries because the user already has transactional intent. There is no exploration to accelerate and no per-period advertising revenue to trade off against transaction value.

The dominant predictor of transactional AIO deployment is instead intent classification confidence (-4.75, $p = 0.014$): the more confident the platform is that the query is truly transactional, the less likely it deploys AIO. This suggests that for transactional queries, the platform's primary challenge is not evaluating the deployment condition derived in the model, but determining whether the condition even applies. When intent is certain, the platform confidently withholds AIO; when intent is ambiguous, it hedges by deploying AIO under the logic of intent types where deployment is the default. This reveals a gap between the theoretical model, which assumes intent is known, and the empirical setting, where intent classification itself is a first-order decision.

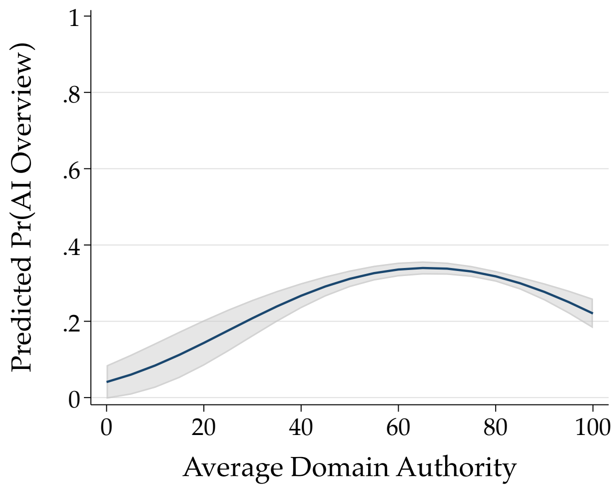
One possible explanation for the absence of a CPC effect is that our model's assumption that the platform captures a share of transaction value s may not hold for Google in the way it would for a vertically integrated marketplace such as Amazon. If Google monetizes transactional queries primarily through advertising rather than referral commissions, then from the platform's perspective the deployment decision for transactional queries resembles that for navigational queries: deployment incurs cost without generating additional revenue. The similar deployment rates for navigational (10%) and transactional (20%) queries are consistent with this interpretation.

5.5 Hypothesis 5: Source Quality and AIO Deployment

We proxy source quality using the variable domain authority, which corresponds to the $\Psi(f_h)$ function in our model, capturing the platform's capability to generate a reliable AI Overview from available sources. Column (1) of Table 7 reports a pooled specification across all intent types that tests the extension in Section 3.3. If AIO quality depends on the underlying data, $f_h^{AI} \equiv \Psi(f_h)$, then deployment should exhibit an inverted-U relationship with source quality: too low, and AIO cannot produce a reliable summary; too high, and AIO is redundant because organic results already fulfil the intent.

Average domain authority enters with a significant positive linear term (0.094, $p < 0.001$) and a significant negative quadratic term (-0.0007, $p < 0.001$). Though necessary, a significant quadratic coefficient alone is not sufficient to establish an inverted-U relationship (Haans et al., 2016). Following the procedure recommended by Lind and Mehlum (2010): first, we assessed whether the slope is sufficiently steep at both ends of the data range. The slope is positive and significant at the lower bound of domain authority ($b = 0.094$, $p < 0.001$) and negative and significant at the upper bound ($b = -0.048$, $p < 0.001$), confirming that the relationship rises and then falls within the observed range. Second, we computed the turning point using the delta method, which yields a peak at domain authority of 65.99 (95% CI: $[62.04, 69.93]$). This turning point is well within the observed data range and close to the sample mean (68.5), and its confidence interval lies entirely within the support of the data, confirming that the inverted U is not an artifact of extreme values. These results support Hypothesis 5.

Figure 4: Predicted probability of AIO deployment by source quality.



Notes: Predicted probabilities from logistic regression (Table 7, column 1). Intent-type fixed effects included.

Figure 4 plots the predicted probabilities. AIO probability rises from near zero at very low domain authority, peaks around 0.34 at domain authority of approximately 66, and declines at higher levels. The peak near the sample mean suggests that Google's deployment threshold is calibrated to the typical quality of web content: AIOs are deployed where source material is adequate but organic results alone are not fully satisfying. At the extremes, AIO is withheld either because sources are too poor to produce a reliable summary (the "data voids" that Google describes; Google, 2025) or because high-quality organic results already fulfil the intent, making AI synthesis redundant.

The intent dummies confirm the deployment hierarchy established in the descriptive statistics. Relative to informational queries (the omitted category), navigational queries have the lowest deployment (-1.95, $p < 0.001$), followed by transactional (-1.55, $p < 0.001$), and commercial (-0.29, $p = 0.015$). The gap between navigational and transactional on one hand and informational and commercial on the other mirrors the exploring-versus-targeted distinction in our theoretical framework.

6 Discussion and Conclusion

Our research develops a theoretical framework for how a monopolist search platform strategically deploys AI-generated content across search queries that differ in their intent, and tests these predictions using search engine results pages from Google. We find that AI Overview deployment is neither uniform nor random: it varies systematically with search intent in ways that are consistent with platform revenue maximization. For exploratory queries, AIOs are deployed by default and withheld only when organic results are sufficient or source quality is too low. For targeted queries, deployment is rare and occurs primarily when the platform lacks confidence in the organic match. For monetizable queries, deployment reflects a tradeoff between converting users toward transactions and preserving advertising revenue. These findings suggest that understanding user-side effects of AI in search requires first understanding the platform's upstream deployment decision. More broadly, these findings indicate that deployment is shaped not only by quality-related considerations, but also by monetization incentives embedded in SERP design.

6.1 Contribution to Literature

Our findings deepen the understanding of how platform design choices shape user outcomes in search. A growing literature on platform design and directed search establishes that platforms actively steer users through ranking, display, and recommendation mechanisms, with significant consequences for market outcomes (Hagiu and Jullien, 2011; Fradkin, 2017; Lee and Musolff, 2025). Our research extends this literature by identifying the selective deployment of AI-generated content as a novel directed search mechanism. Unlike conventional ranking decisions that reorder existing results, AIO deployment introduces platform-generated content that competes with organic results on the same page. The platform thus simultaneously controls the creation of the competing content, the ranking of alternatives, and the monetization of the surrounding space. Our finding that deployment patterns differ qualitatively across intent types shows that this triple control is exercised with considerable strategic granularity, with the logic of deployment varying across intent types rather than only its intensity. This extends the self-preferencing literature, which has documented how platforms favor their own products through ranking and placement (Zou and Zhou, 2025; Long and Amaldoss, 2024) or by copying complementor innovations (Choi et al., 2025). AIOs represent a qualitatively different mechanism: the platform does not simply rank its own products higher or replicate existing innovations, but generates new content from third-party sources that competes with those sources on the same page.

Our findings also contribute to the literature on search advertising and platform monetization. Prior work establishes that organic and sponsored results interact on the SERP, with changes to the organic environment producing non-trivial consequences for advertising performance (Yang and Ghose, 2010; Xu et al., 2012). However, this literature treats the SERP layout as fixed. Our results show that when the platform itself introduces AI-generated content, the organic-sponsored interaction extends to a three-way relationship among organic results, AI content, and advertising. The fulfilment-by-CPC interaction we identify for commercial queries provides direct evidence that the platform internalizes this tension: deployment increases with advertising value when organic results are weak, but decreases when strong organic results are already generating valuable clicks. This tradeoff is absent for transactional queries, where there is no exploration-phase advertising revenue to sacrifice, suggesting that the mechanism is specific to the commercial advertising channel rather than a generic fulfilment effect.

A growing body of work examines how users interact with generative AI in search, demonstrating that AI features meaningfully alter information processing and decision-making (Zhu et al., 2025; Spatharioti et al., 2025; Kaiser et al., 2025; Liang et al., 2025). These studies share a common design in which AI deployment is treated as exogenous. Our findings highlight a selection problem in this literature: the queries for which AI is deployed are not randomly chosen but are selected precisely because the platform expects deployment to be profitable. Studies that estimate user-side effects of AI without accounting for this strategic selection risk confounding the effect of AI itself with the characteristics of queries that the platform chose to treat. This concern extends to recent work documenting downstream consequences of AIOs, such as reduced publisher traffic (Khosravi and Yoganarasimhan, 2026) and narrower source exposure (Aral et al., 2026), where the observed effects may partly reflect the platform's strategic query selection rather than the causal impact of AI content alone. Our theoretical model and empirical results provide a foundation for understanding this selection mechanism, and may inform identification strategies in future demand-side studies.

6.2 Practical Implications

Our findings carry implications for three groups of stakeholders: content creators, advertisers, and regulators.

For content creators and publishers, our results clarify the conditions under which AI Overviews are most likely to displace organic traffic. The inverted-U relationship between source quality and AIO deployment (Hypothesis 5) implies that mid-quality content faces the greatest displacement risk: it is good enough to serve as source material for AI synthesis but not so authoritative that organic results render the AIO redundant. Producers of highly authoritative content, such as established news organizations, government agencies, and domain-specific reference sites, may face less displacement precisely because their organic results already satisfy users, reducing the platform's incentive to deploy an AIO. However, this also means that such producers may supply the raw material from which AIOs are constructed without necessarily receiving the click-through traffic that historically sustained their business models. This tension sits at the heart of ongoing litigation between publishers and Google over the scraping and synthesis of copyrighted content (Singh, 2025). Our finding that the platform withholds AIOs when source quality is too low suggests that content creators cannot simply reduce quality to avoid displacement; doing so would also reduce their visibility in organic results.

For advertisers, the fulfilment-by-CPC interaction for commercial queries reveals a tension that performance marketing strategies must accommodate. When AIOs are deployed, they may redirect user attention away from the sponsored links through which advertisers acquire customers. Our results suggest that this displacement is most pronounced for high-CPC queries where organic results are weak, precisely the conditions under which advertisers historically captured the most value. Advertisers operating in product categories where users frequently engage in exploratory commercial search should anticipate that AIOs will increasingly mediate the path from exploration to purchase, potentially requiring shifts toward advertising formats that are integrated with or adjacent to AI-generated content.

For regulators, our findings contribute to the ongoing legal and policy debates surrounding Google's market power in search. The deployment patterns we document are consistent with a platform that strategically modulates AI content based on revenue implications, not solely on user benefit. This distinction matters for antitrust analysis: if AIO deployment were purely a quality improvement, uniform deployment across intent types would be expected. Instead, the systematic variation we observe suggests that AIO deployment is a strategic instrument for revenue optimization, and is therefore relevant to concerns about how a dominant platform may use AI-generated content to shape attention and monetization.

Our findings are also relevant to broader debates about self-preferencing and the extension of monopoly power into AI-generated answers. In this respect, AIOs differ meaningfully from the content reproduction at the center of news aggregation regulations. Those regulatory responses focused on platforms reproducing content without compensation, leading to frameworks that required platforms to negotiate licensing agreements with publishers. AIOs present a more complex challenge: Google is not simply reproducing content but selectively synthesizing it in ways that serve its own commercial interests. Addressing this may require regulators to develop new tools that go beyond those designed for aggregation.

6.3 Limitations and Future Research

Our theoretical model represents an early attempt at formally characterizing the platform's AIO deployment decision, and as such it rests on a number of simplifying assumptions. For example, it treats the platform's knowledge of user search intent as certain. This contrasts with our empirical results, which show that intent uncertainty is itself a central feature of the deployment problem: intent classification confidence emerges as the dominant predictor of transactional AIO deployment, rather than the fulfilment and revenue variables our model emphasizes. In other words, the platform's problem is not only whether to deploy an AIO conditional on intent, but also how confidently it can infer intent in the first place. Our empirical results therefore provide a basis for future theoretical work on the joint problem of intent classification and deployment under ambiguity.

Second, our data captures a snapshot of Google's deployment strategy in November 2025 and thus at a particular point in the technology's maturity. AIO deployment patterns are likely to evolve as the underlying language models improve, as Google refines its deployment algorithms, and as competitive pressures from AI-native search products intensify. Longitudinal studies tracking deployment decisions over time would reveal whether the patterns we identify are stable features of the platform's revenue calculus or transient artifacts of a particular stage in the technology's rollout.

Finally, our analysis focuses on the platform's own deployment decision and does not directly address its interactions with the broader ecosystem. Extending our supply-side deployment model with demand-side data on click-through rates and user engagement would allow estimation of the welfare effects of AIO deployment across intent types. Given our finding that deployment is strategically selected, such studies should account for the selection mechanism we document to avoid confounding the treatment effect of AI with the characteristics of queries selected for treatment. Our analysis also focuses on Google, leaving open how its AIO deployment strategy interacts with and shapes the strategies of competing platforms operating under different incentive structures and monetization constraints. Content creators may also adjust their production and SEO strategies in response to AIO deployment, advertisers may shift spending across formats, and competing platforms may alter their own AI strategies. Taken together, these strategic interactions represent an important direction for future research.

References

Agarwal, Ashish, Kartik Hosanagar, and Michael D. Smith (2011). "Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets". *Journal of Marketing Research* 48.6, pp. 1057–1073.

— (2015). "Do Organic Results Help or Hurt Sponsored Search Performance?" *Information Systems Research* 26.4, pp. 695–713.

Ahn, Jae-Hyeon, Yoon-Soo Bae, Jaehyeon Ju, and Wonseok Oh (2018). "Attention Adjustment, Renewal, and Equilibrium Seeking in Online Search: An Eye-Tracking Approach". *Journal of Management Information Systems* 35.4, pp. 1218–1250.

Aral, Sinan, Haiwen Li, and Rui Zuo (2026). "The rise of AI search: Implications for information markets and human judgement at scale". *arXiv preprint arXiv:2602.13415*.

Athey, Susan, Markus Mobius, and Jeno Pal (2021). *The Impact of Aggregators on Internet News Consumption*. Tech. rep. National Bureau of Economic Research.

Brand, James, Ayelet Israeli, and Donald Ngwe (2025). "Using LLMs for Market Research". Harvard Business School Marketing Unit Working Paper No. 23-062, revised October 2025.

Broder, Andrei (2002). "A Taxonomy of Web Search". *ACM SIGIR Forum*. Vol. 36. 2. ACM New York, NY, USA, pp. 3–10.

Byrne, Michael D., Bonnie E. John, Neil S. Wehrle, and David C. Crow (1999). "The Tangled Web We Wove: A Taskonomy of WWW Use". *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 544–551.

Calzada, Joan and Ricard Gil (2020). "What Do News Aggregators Do? Evidence from Google News in Spain and Germany". *Marketing Science* 39.1, pp. 134–167.

Carlson, Natalie and Vanessa Burbano (2025). "The Use of LLMs to Annotate Data in Management Research: Foundational Guidelines and Warnings". *Strategic Management Journal* 46.5, e70023. DOI: `10.1002/smj.70023`.

Chapekis, Athena, Anna Lieb, Sono Shah, and Aaron Smith (May 2025). *What Web Browsing Data Tells Us About How AI Appears Online*. Tech. rep. Pew Research Center. URL: `https://www.pewresearch.org/data-labs/2025/05/23/what-web-browsing-data-tells-us-about-how-ai-appears-online/`.

Chen, Jianqing and Jan Stallaert (2014). "An Economic Analysis of Online Advertising Using Behavioral Targeting1". *Mis Quarterly* 38.2, pp. 429–450.

Chen, Le, Alan Mislove, and Christo Wilson (2016). "An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace". *Proceedings of the 25th International Conference on World Wide Web*, pp. 1339–1349.

Choi, Jay Pil, Doh-Shin Jeon, and Byung-Cheol Kim (2025). ""Sherlocking" and Platform Information Policy". *Management Science* 71.4, pp. 3073–3092. DOI: `10.1287/mnsc.2023.01187`.

Dai, Honghua, Lingzhi Zhao, Zaiqing Nie, Ji-Rong Wen, Lee Wang, and Ying Li (2006). "Detecting Online Commercial Intention". *Proceedings of the 15th International Conference on World Wide Web*, pp. 829–837.

Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz (2007). "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords". *American Economic Review* 97.1, pp. 242–259.

Foerderer, Jens, Thomas Kude, Sunil Mithas, and Armin Heinzl (2018). "Does Platform Owner's Entry Crowd Out Innovation? Evidence from Google Photos". *Information Systems Research* 29.2, pp. 444–460. DOI: 10.1287/isre.2018.0787.

Fradkin, Andrey (2017). "Search, Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb". *Working Paper*.

Ghose, Anindya, Avi Goldfarb, and Sang Pil Han (2013). "How Is the Mobile Internet Different? Search Costs and Local Activities". *Information Systems Research* 24.3, pp. 613–631.

Ghose, Anindya and Sha Yang (2009). "An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets". *Management Science* 55.10, pp. 1605–1622.

Gong, Jing, Vibhanshu Abhishek, and Beibei Li (2018). "Examining the Impact of Keyword Ambiguity on Search Advertising Performance: A Topic Model Approach1". *MIS quarterly* 42.3, pp. 805–829.

Google (May 2025). *AI Overviews and AI Mode in Search*. Tech. rep. URL: https://search.google/pdf/google-about-AI-overviews-AI-Mode.pdf.

Haans, Richard F. J., Constant Pieters, and Zi-Lin He (2016). "Thinking About U: Theorizing and Testing U- and Inverted U-Shaped Relationships in Strategy Research". *Strategic Management Journal* 37.7, pp. 1177–1195. DOI: 10.1002/smj.2399.

Hagiu, Andrei and Bruno Jullien (2011). "Why Do Intermediaries Divert Search?" *The RAND Journal of Economics* 42.2, pp. 337–362.

Hagiu, Andrei and Julian Wright (2024). "Optimal discoverability on platforms". *Management Science* 70.11, pp. 7770–7790.

Jansen, Bernard J., Danielle L. Booth, and Amanda Spink (2008). "Determining the Informational, Navigational, and Transactional Intent of Web Queries". *Information Processing & Management* 44.3, pp. 1251–1266.

Kaiser, Carolin, Jakob Kaiser, Rene Schallner, and Sabrina Schneider (2025). "A New Era of Online Search? A Large-Scale Study of User Behavior and Personal Preferences during Practical Search Tasks with Generative AI versus Traditional Search Engines". *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–7.

Khosravi, Mehrzad and Hema Yoganarasimhan (2026). "Impact of AI Search Summaries on Website Traffic: Evidence from Google AI Overviews and Wikipedia". *Available at SSRN 6164926*.

Lee, Kwok Hao and Leon Musolff (2025). "Two-Sided Markets Shaped by Platform-Guided Search". *Working Paper*.

Liang, Yidong, Zhijing Wu, Fan Zhang, Dandan Song, and Heyan Huang (2025). "How Users Interact with Generative Information Retrieval Systems: A Study of User Behavior and Search Experience". *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 634–644.

Lind, Jo Thori and Halvor Mehlum (2010). "With or Without U? The Appropriate Test for a U-Shaped Relationship". *Oxford Bulletin of Economics and Statistics* 72.1, pp. 109–118. DOI: 10.1111/j.1468-0084.2009.00569.x.

Long, Fei and Wilfred Amaldoss (2024). "Self-preferencing in e-commerce marketplaces: The role of sponsored advertising and private labels". *Marketing Science* 43.5, pp. 925–952.

Moshary, Sarah (2025). "Does sponsored search advertising augment organic search? evidence from an e-commerce platform". *Management Science* 71.11, pp. 9687–9709.

Raval, Devesh (2022). "Steering in One Click: Platform Self-Preferencing in the Amazon Buy Box". *Unpublished Manuscript*.

Singh, Madhavi (2025). *The Antitrust Case Against AI Overviews*. Harvard Journal of Law and Technology Digest. October 6, 2025. URL: `https://jolt.law.harvard.edu/digest/the-antitrust-case-against-ai-overviews`.

Spatharioti, Sofia Eleni, David Rothschild, Daniel G. Goldstein, and Jake M. Hofman (2025). "Effects of LLM-based Search on Decision Making: Speed, Accuracy, and Overreliance". *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pp. 1–15.

Varian, Hal R. (2007). "Position Auctions". *International Journal of Industrial Organization* 25.6, pp. 1163–1178.

Xu, Lizhen, Jianqing Chen, and Andrew Whinston (2012). "Effects of the Presence of Organic Listing in Search Advertising". *Information Systems Research* 23.4, pp. 1284–1302.

Yang, Sha and Anindya Ghose (2010). "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" *Marketing Science* 29.4, pp. 602–623.

Zhu, Yuting, Shuang Zheng, Xin Ye, and Liang Shen (2025). "Generative Search: Evidence from a Large-Scale Field Experiment". *Available at SSRN 5297194*.

Zou, Tianxin and Bo Zhou (2025). "Self-preferencing and search neutrality in online retail platforms". *Management Science* 71.5, pp. 4087–4107.

Online Appendices

A Proofs

Proof of Lemma 1. First, recall that $V_0 = \frac{e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+\lambda_T+\lambda_C)}$, where $\lambda_h = \frac{P_h f_h}{1-\delta(1-f_h)}$. Further, $\frac{\partial\lambda_h}{\partial P_h} = \frac{f_h}{1-\delta(1-f_h)} > 0$, and $\frac{\partial\lambda_h}{\partial f_h} = \frac{P_h(1-\delta)}{(1-\delta(1-f_h))^2} > 0$.

To see that V_0 is increasing in P_h for all $h \in \mathbb{H}$: First, note that P_I only appears in the denominator through λ_I, and $\frac{\partial V_0}{\partial \lambda_I} > 0$. Therefore, $\frac{\partial V_0}{\partial P_I} > 0$. Second, P_N only appears in the numerator through λ_N, and $\frac{\partial V_0}{\partial \lambda_N} > 0$. Therefore, $\frac{\partial V_0}{\partial P_N} > 0$. Finally, observe that both P_T and P_C appear in both the denominator and numerator, and $\frac{\partial V_0}{\partial \lambda_T} > 0$ and $\frac{\partial V_0}{\partial \lambda_C} > 0$. Therefore, $\frac{\partial V_0}{\partial P_T} > 0$ and $\frac{\partial V_0}{\partial P_C} > 0$.

To see that V_0 is increasing in the payoffs e, s, a, note that these terms only appear positively in the numerator. Therefore V_0 increases in each of the payoffs.

To see that V_0 is increasing in f_I, f_N, f_T: First, f_I only appears in the denominator through λ_I, and $\frac{\partial V_0}{\partial \lambda_I} > 0$. Therefore $\frac{\partial V_0}{\partial f_I} > 0$. Second, f_N only appears in the numerator through λ_N, and $\frac{\partial V_0}{\partial \lambda_N} > 0$. Therefore, $\frac{\partial V_0}{\partial f_N} > 0$. Finally, f_T appears in both the numerator and denominator through λ_T, and $\frac{\partial V_0}{\partial \lambda_T} > 0$. Therefore, $\frac{\partial V_0}{\partial f_T} > 0$.

To see how V_0 changes in f_C, note first that the total derivative of V_0 with respect to f_C is

$$\frac{\partial V_0}{\partial f_C} = \frac{\delta P_C((1-\delta)(e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}) - a(1-\delta(\lambda_I+\lambda_T+\lambda_C)))}{(1-\delta(1-f_C))^2(1-\delta(\lambda_I+\lambda_T+\lambda_C))^2}.$$

Since the denominator is positive and δ and P_C are positive, then $\frac{\partial V_0}{\partial f_C} > 0$ if and only if $(1-\delta)(e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}) - a(1-\delta(\lambda_I+\lambda_T+\lambda_C)) > 0$, and $\frac{\partial V_0}{\partial f_C} = 0$ when the equation holds with equality. Further note that

$$(1-\delta)(e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}) - a(1-\delta(\lambda_I+\lambda_T+\lambda_C)) > 0 \Leftrightarrow$$
$$(1-\delta)V_0 - a > 0 \text{ dividing both sides by } 1-\delta(\lambda_I+\lambda_T+\lambda_C).$$

Therefore, $\frac{\partial V_0}{\partial f_C} > 0$ if and only if $V_0 > \frac{a}{1-\delta}$ and is zero when $V_0 = \frac{a}{1-\delta}$. □

Proof of Proposition 1. Recall that $V_0 = \frac{e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+\lambda_T+\lambda_C)}$ and $V_0^I = \frac{-d\frac{P_I}{1-\delta(1-f_I^{AI})} + e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I^{AI}+\lambda_T+\lambda_C)}$. The platform adopts AI Overview for informational searches if and only if

$$V_0 < V_0^I \Leftrightarrow$$
$$d\frac{P_I}{1-\delta(1-f_I^{AI})}(1-\delta(\lambda_I+\lambda_T+\lambda_C)) < (e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C})\delta(\lambda_I^{AI} - \lambda_I) \Leftrightarrow$$
$$d < V_0\frac{\delta(\lambda_I^{AI} - \lambda_I)(1-\delta(1-f_I^{AI})}{P_I} = V_0\frac{\delta(1-\delta)(f_I^{AI} - f_I)}{1-\delta(1-f_I)}.$$

When the inequality is satisfied, we show that the right-hand side of the inequality is strictly decreasing in f_I.

$$\begin{aligned}
\frac{\partial V_0 \delta \frac{(1-\delta)(f_I^{AI}-f_I)}{1-\delta(1-f_I)}}{\partial f_I} =& \frac{\partial V_0}{\partial f_I}\delta \frac{(1-\delta)(f_I^{AI}-f_I)}{1-\delta(1-f_I)} - \frac{(1-\delta)V_0\delta(1-\delta(1-f_I^{AI}))}{(1-\delta(1-f_I))^2} \\
=& \frac{\delta V_0}{1-\delta(\lambda_I+\lambda_T+\lambda_C)} \cdot \frac{P_I(1-\delta)^2\delta(f_I^{AI}-f_I)}{(1-\delta(1-f_I))^3} - \frac{(1-\delta)V_0\delta(1-\delta(1-f_I^{AI}))}{(1-\delta(1-f_I))^2} \\
=& \frac{\delta(1-\delta)V_0}{(1-\delta(1-f_I))^2}\left[\frac{P_I\delta(1-\delta)(f_I^{AI}-f_I)}{(1-\delta(\lambda_I+\lambda_T+\lambda_C))(1-\delta(1-f_I))} - (1-\delta(1-f_I^{AI}))\right].
\end{aligned}$$

We can now show that this is always negative. Since $\frac{\delta(1-\delta)V_0}{(1-\delta(1-f_I))^2} > 0$, then the sign depends on the terms in the square bracket.

$$\begin{aligned}
\frac{P_I\delta(1-\delta)(f_I^{AI}-f_I)}{(1-\delta(\lambda_I+\lambda_T+\lambda_C))(1-\delta(1-f_I))} - (1-\delta(1-f_I^{AI})) &< 0 \\
P_I\delta(1-\delta)(f_I^{AI}-f_I) < (1-\delta(1-f_I^{AI}))(1-\delta(1-f_I))(1-\delta(\lambda_I+\lambda_T+\lambda_C)&) \\
P_I\delta f_I^{AI}(1-\delta(1-f_I)) < (1-\delta(1-f_I^{AI}))(1-\delta(1-f_I))(1-\delta(\lambda_T+\lambda_C)&) \\
0 < 1-\delta(\lambda_I^{AI}+\lambda_T+\lambda_C&).
\end{aligned}$$

This is always true since $\lambda_h < P_h$ for all h and for all f with or without AI.

Therefore, when f_I increases, the inequality is less likely to be satisfied. □

Proof of Proposition 3. The platform adopts AI for commercial searches when $V_0^C > V_0$. First, recall that $V_0 = \frac{e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+\lambda_T+\lambda_C)}$ and $V_0^C = \frac{e\lambda_N+s\lambda_T+P_C(-d+s\delta\frac{f_{ct}f_T}{1-\delta(1-f_T)})}{1-\delta(\lambda_I+\lambda_T+P_C(1-f_{ct}+\delta\frac{f_{ct}f_T}{1-\delta(1-f_T)}))}$. Then observe that

$$\begin{aligned}
P_C(-d+s\delta\frac{f_{ct}f_T}{1-\delta(1-f_T)}) =& \frac{a\lambda_C}{f_C} + P_C(\frac{\delta s f_{ct}f_T}{1-\delta(1-f_T)} - d - \frac{a}{1-\delta(1-f_C)}). \\
\delta(P_C(1-f_{ct}+\delta\frac{f_{ct}f_T}{1-\delta(1-f_T)})) =& \delta\lambda_C + \delta P_C(1+\frac{\delta f_{ct}f_T}{1-\delta(1-f_T)} - \frac{f_C}{1-\delta(1-f_C)} - f_{ct}).
\end{aligned}$$

Then we know the platform adopts AI Overview if and only if

$$V_0^C > V_0 \Leftrightarrow \frac{e\lambda_N + s\lambda_T + P_C(-d + s\delta\frac{f_{ct}f_T}{1-\delta(1-f_T)})}{1-\delta(\lambda_I + \lambda_T + P_C(1 - f_{ct} + \delta\frac{f_{ct}f_T}{1-\delta(1-f_T)}))} > \frac{e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+\lambda_T+\lambda_C)} \Leftrightarrow$$

$$(1-\delta(\lambda_I+\lambda_T+\lambda_C))(e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C} + P_C(\frac{\delta s f_{ct}f_T}{1-\delta(1-f_T)} - d - \frac{a}{1-\delta(1-f_C)})) >$$

$$(e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C})(1-\delta(\lambda_I+\lambda_T+\lambda_C) - \delta P_C(1 + \frac{\delta f_{ct}f_T}{1-\delta(1-f_T)} - \frac{f_C}{1-\delta(1-f_C)} - f_{ct}))$$

$$\Leftrightarrow \delta(e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C})(1 - f_{ct} + \frac{\delta f_{ct}f_T}{1-\delta(1-f_T)} - \frac{f_C}{1-\delta(1-f_C)}) >$$

$$(1-\delta(\lambda_I+\lambda_T+\lambda_C))(\frac{a}{1-\delta(1-f_C)} + d - \frac{\delta s f_{ct}f_T}{1-\delta(1-f_T)}) \Leftrightarrow$$

$$d < \delta V_0(1 - f_{ct} + \frac{\delta f_{ct}f_T}{1-\delta(1-f_T)} - \frac{f_C}{1-\delta(1-f_C)}) + \frac{\delta s f_{ct}f_T}{1-\delta(1-f_T)} - \frac{a}{1-\delta(1-f_C)}.$$

□

Proof of Proposition 4. The platform adopts AI for transactional searches when $V_0^T > V_0$. First, recall that $V_0 = \frac{e\lambda_N+s\lambda_T+\frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+\lambda_T+\lambda_C)}$ and $V_0^T = \frac{e\lambda_N+s(P_T f_T^{AI}+(1-f_T^{AI})\delta\lambda_T)-P_T d+\frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+P_T f_T^{AI}+(1-f_T^{AI})\delta\lambda_T+\lambda_C)}$. Then we know the platform adopts AI Overview if and only if

$$V_0^T > V_0 \Leftrightarrow$$

$$(-P_T d + e\lambda_N + s(P_T f_T^{AI} + (1-f_T^{AI})\delta\lambda_T) + \frac{a\lambda_C}{f_C})(1-\delta(\lambda_I+\lambda_T+\lambda_C)) >$$

$$(e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C})(1-\delta(\lambda_I + P_T f_T^{AI} + (1-f_T^{AI})\delta\lambda_T + \lambda_C)).$$

Then note that

$$P_T f_T^{AI} + (1-f_T^{AI})\delta\lambda_T - \lambda_T = P_T f_T^{AI} - \lambda_T(1-\delta+\delta f_T^{AI}) = P_T\frac{(f_T^{AI}-f_T)(1-\delta)}{1-\delta(1-f_T)},$$

$$-P_T d + e\lambda_N + s(P_T f_T^{AI} + (1-f_T^{AI})\delta\lambda_T) + \frac{a\lambda_C}{f_C} = P_T(s\frac{(f_T^{AI}-f_T)(1-\delta)}{1-\delta(1-f_T)} - d) + e\lambda_N + s\lambda_T + \frac{a\lambda_C}{f_C},$$

$$1-\delta(\lambda_I + P_T f_T^{AI} + (1-f_T^{AI})\delta\lambda_T + \lambda_C) = 1-\delta(\lambda_I+\lambda_T+\lambda_C) - P_T\delta\frac{(f_T^{AI}-f_T)(1-\delta)}{1-\delta(1-f_T)}.$$

Then we can rewrite $V_0^T > V_0$ as

$$
\begin{aligned}
(P_T(s\frac{(f_T^{AI}-f_T)(1-\delta)}{1-\delta(1-f_T)}-d)+e\lambda_N+s\lambda_T+\frac{a\lambda_C}{f_C})(1-\delta(\lambda_I+\lambda_T+\lambda_C)) > \\
(e\lambda_N+s\lambda_T+\frac{a\lambda_C}{f_C})(1-\delta(\lambda_I+\lambda_T+\lambda_C)-P_T\delta\frac{(f_T^{AI}-f_T)(1-\delta)}{1-\delta(1-f_T)}) \Leftrightarrow \\
\frac{(f_T^{AI}-f_T)(1-\delta)}{1-\delta(1-f_T)}(s(1-\delta(\lambda_I+\lambda_T+\lambda_C))+\delta(e\lambda_N+s\lambda_T+\frac{a\lambda_C}{f_C}))>d(1-\delta(\lambda_I+\lambda_T+\lambda_C)) \\
\Leftrightarrow d<\frac{(f_T^{AI}-f_T)(1-\delta)}{1-\delta(1-f_T)}(s+\delta\frac{e\lambda_N+s\lambda_T+\frac{a\lambda_C}{f_C}}{1-\delta(\lambda_I+\lambda_T+\lambda_C)})=\frac{(f_T^{AI}-f_T)(1-\delta)}{1-\delta(1-f_T)}(s+\delta V_0).
\end{aligned}
$$

Therefore, the condition $f_T^{AI} - f_T > 0$ is a necessary condition for the platform to adopt AI Overview for transactional searches, and $d < \frac{(1-\delta)(f_T^{AI}-f_T)}{1-\delta(1-f_T)}(s+\delta V_0)$ is the necessary and sufficient condition. $\square$

B Search Query Generation and Validation

Testing our theoretical predictions requires a query sample that spans the four intent types, varies along dimensions that could affect AIO deployment (e.g., YMYL status, local relevance, question format, length), and reflects queries that real users actually search. Google does not publish query logs, and publicly available datasets such as those derived from search engine competitions would likely overrepresent popular head queries and underrepresent the long tail of specific, intent-rich queries that are central to our analysis. We therefore employed a structured generation approach that combines LLM-based query production with external validation against a commercial search analytics API.

The key design principle, analogous to the separation of subjective judgment from deterministic rules advocated by Carlson and Burbano (2025), is that the LLM handles only generation, producing realistic query text that mimics natural search behavior, while all intent labels, confidence scores, and search volumes are determined externally by the API based on Google's own signals. This means that our final sample is not dependent on the LLM's judgment about intent, substantially mitigating concerns about classification accuracy or bias. More broadly, because the external validation step imposes intent and volume requirements that are independent of the prompt, prompt sensitivity, an important concern in LLM-based research (Carlson and Burbano, 2025), affects the candidate pool but not the final sample. This appendix describes the generation protocol, validation procedure, and filtering criteria in detail.

B.1 Generation Protocol

We used Claude Haiku 4.5 (Anthropic) accessed through the Anthropic API.[13] Each API call requested 180 candidate queries in structured JSON format, with the model instructed to return each query's text, whether it was phrased as a question, and its approximate length category (short: 2–4 words; long: 5–10 words).

To ensure the generated queries resemble actual search behavior rather than generic LLM completions, we seeded each prompt with *anchor examples*: high-frequency keywords extracted from the DataForSEO search analytics API for the United States and Ireland. These anchors serve a similar function to few-shot examples in classification tasks (Carlson and Burbano, 2025), grounding the model's output in the vocabulary and phrasing conventions of real searches. For instance, anchor examples for commercial queries included brand names and retail chains (e.g., "old navy," "papa john's pizza"), while anchors for navigational queries included login and service pages (e.g., "facebook login").

Each generation call was parameterized along four dimensions that define the combinatorial design of our query sample:

[13] For reproducibility, we document the model identifier (`claude-haiku-4-5-20251001`), temperature (0.9), maximum output tokens (8,000), and system prompt ("You are a data generator that strictly adheres to constraints and outputs valid JSON."). Prompt templates, anchor keyword lists, and the complete generation and validation code are available upon request.

1. **Search intent** (4 levels): informational, navigational, commercial, transactional. The prompt mapped each intent to Google's own query classification vocabulary, drawing on the Search Quality Evaluator Guidelines. For example, informational queries were described as "KNOW (User wants to know a fact, concept, or answer)," while transactional queries were described as "DO (User wants to buy, download, or complete an action NOW. Must use specific models/brands)."

2. **YMYL status** (2 levels): YMYL queries were constrained to health, financial, legal, and civic topics; non-YMYL queries covered entertainment, hobbies, technology, and general knowledge.

3. **Local relevance** (2 levels): local queries were instructed to include geographic modifiers ("near me," "in my area," "nearby," or a specific city name), with approximately 75% using generic location terms and 25% using the city name. Non-local queries were explicitly prohibited from containing any location terms.

4. **Geographic market** (2 levels): United States (Los Angeles) and Ireland (Dublin), each with market-specific anchor examples and location parameters.

This yields $4 \times 2 \times 2 \times 2 = 32$ unique configuration cells. To generate sufficient candidate volume for post-validation filtering, we repeated each configuration 8 times, producing $32 \times 8 = 256$ batches of 180 queries each, for a target of approximately 46,000 raw candidate queries.[14]

For commercial queries with local relevance (the most challenging category to generate with correct intent alignment), we provided detailed positive and negative examples within the prompt. For instance, the prompt specified that "best italian restaurants in my area" is commercial (comparison/research), while "chipotle near me" is navigational.

B.2 External Validation

A central concern with LLM-generated research data is that the generated items may reflect the model's training distribution rather than the target construct (Carlson and Burbano, 2025; Brand et al., 2025). We address this by treating LLM generation as a *candidate production* step and imposing external validation through the DataForSEO Keywords Data API, which returns Google's own signals for each query.

For each candidate query, the API returns:

- **Intent classification**: a primary intent label (informational, navigational, commercial, or transactional) based on Google's internal signals, along with a confidence score $\in [0, 1]$ reflecting how clearly the query maps to a single intent category.

- **Search volume**: the average monthly search volume over the preceding 12 months, indicating whether the query reflects actual user search behavior.

[14] The actual yield was 33,082 candidates after accounting for occasional API failures and malformed responses.

- **Cost per click**: the average CPC in USD, which we use as a proxy for advertising value of the search query in our empirical specifications.

B.3 Filtering and Final Sample

We applied three sequential filters to the raw candidate pool:

1. **Deduplication**: Exact-match deduplication on query text, retaining the first occurrence. Because the same configuration was repeated 8 times with high temperature, some queries appeared in multiple batches.
2. **Search volume**: We excluded queries with fewer than 10 average monthly searches, which ensures that each query in our sample corresponds to a real search phrase used by actual users, rather than a plausible but synthetic combination that no one actually searches.
3. **Intent confidence**: We ranked queries by the API's intent classification confidence and retained those with the highest scores within each intent category, ensuring strong alignment between our target intent and Google's classification.

After filtering, 2,057 unique queries remained. Table 8 reports the final distribution.

Table 8: Final query sample by intent type after validation and filtering.

Intent type	**Queries**
Informational	567
Navigational	600
Commercial	364
Transactional	526
Total	2,057

The smaller commercial sample reflects the inherent difficulty of generating queries that the validation API classifies with high confidence as commercial rather than informational or transactional. Commercial intent occupies a narrow region between exploratory research and purchase commitment, and many LLM-generated candidates were reclassified by the API into adjacent categories. We view this attrition as a feature of the validation procedure rather than a limitation: it ensures that the commercial queries in our sample genuinely reflect comparison and research behavior rather than mislabeled informational or transactional searches.

Table 9 presents illustrative queries from each intent type, selected to show diversity across topic sensitivity, local relevance, intent classification confidence, and observed AIO deployment rates.

Table 9: Sample search queries by intent type.

Intent	Example query	YMYL	Local	Conf.	AIO rate
Informational	Best beaches near me	No	Yes	0.87	71%
	Best comedy TV shows	No	No	0.88	0%
	Capital gains tax rates	Yes	No	0.93	75%
	DNA structure	No	No	0.94	88%
	Fun things to do nearby	No	Yes	0.91	86%
	How to tie a tie	No	No	1.00	38%
	What are menopause symptoms	Yes	No	0.94	86%
	What is potential energy	No	No	0.89	88%
Navigational	BBC Weather	No	No	0.90	0%
	CVS pharmacy in my area	Yes	Yes	0.85	0%
	Disney+ login	No	No	0.93	0%
	Grammarly official site	No	No	0.88	0%
	Ikea near me	No	Yes	0.87	0%
	Indeed jobs	No	No	0.89	0%
	Netflix login	No	No	0.94	0%
	PayPal sign in	Yes	No	0.98	0%
Commercial	Affordable homeowners insurance	Yes	No	0.97	100%
	Best coffee makers	No	No	0.98	50%
	Best video hosting platforms	No	No	1.00	88%
	Dental implant cost	Yes	No	0.92	100%
	Gaming monitor refresh rates	No	No	1.00	100%
	How to get a small business loan	Yes	No	0.96	100%
	Best DUI attorney in my area	Yes	Yes	0.89	57%
	What is the best credit card for rewards	Yes	No	0.95	100%
Transactional	DJI Mini 4 Pro	No	No	0.96	12%
	Eyeglass shop nearby	Yes	Yes	0.89	0%
	Fujifilm X-S20 camera	No	No	0.90	12%
	Glucose meter kit	Yes	No	0.97	38%
	iPhone 15 Pro Max	No	No	0.93	25%
	Microsoft Surface laptop	No	No	0.94	0%
	Oculus Meta Quest 3 VR headset	No	No	0.90	50%
	Where can I buy Nike shoes near me	No	Yes	1.00	0%

Notes: AIO rate is the share of SERP observations for each query in which an AI Overview was deployed. YMYL and Local are binary indicators for query characteristics. Conf. is the intent classification confidence score from the DataForSEO API. Queries are illustrative examples selected to show diversity across intent types and query attributes.

C Measuring Organic Fulfilment

Our theoretical model requires a measure of baseline organic fulfilment f_h, capturing how well organic results satisfy the user's intent absent any AIO. We therefore construct a dataset following the Search Quality Evaluator Guidelines for Google, which specify detailed rating criteria, decision rules, and illustrative examples for each search intent type. This appendix describes the measurement procedure, which implements the separation of subjective judgment from deterministic rules advocated by Carlson and Burbano (2025): a LLM extracts structured facts about each SERP, and a deterministic rule engine maps those facts to a Needs Met rating. This two-stage design ensures that the same fact profile always produces the same rating, that every rating can be audited by inspecting which rule fired, and that the rules can be revised without rerunning the LLM.

C.1 Fact Extraction (LLM)

For each SERP, Claude Haiku 4.5 (Anthropic) receives the query text, device type, geographic context, query metadata (YMYL status, local relevance, question format, average domain authority), and the top 10 organic results (title, URL, snippet, date, and position). The model is instructed to extract a structured JSON object containing objective, verifiable features of the SERP relative to the inferred user intent. These features include:

- **Result quality dimensions**: Overall result quality (0–3), friction or noise in reaching the answer (0–3), and coverage completeness relative to the query (0–3).
- **Browsing requirement**: Whether the user would need to consult multiple results to satisfy the query, as opposed to finding the answer in a single result.
- **Position-level assessment**: Which positions contain the most helpful results and which contain notably poor results.
- **Navigational signals**: Whether and at what position the target page appears, and whether the top result is the official target.
- **Geographic alignment**: Whether there is a mismatch between the user's location and the geographic orientation of the results, and whether this mismatch is severe enough to render most results unusable.

Crucially, the model is explicitly instructed not to assign a Needs Met rating. This ensures that the LLM performs only the task it is suited for (interpreting the semantic relationship between query and results) and does not perform the task that should be governed by fixed rules (mapping features to ratings).

C.2 Deterministic Rating (Rule Engine)

A rule engine maps the extracted facts to a five-point Needs Met scale: Fully Meets (5), Highly Meets (4), Moderately Meets (3), Slightly Meets (2), and Fails to Meet (1). The rules were developed based on the criteria and examples specified in the Search Quality Evaluator Guidelines. The principal decision branches are:

1. **Quality floor**: If overall result quality is zero, the rating is Fails to Meet regardless of other signals. If quality is low and friction is high, the rating is Slightly Meets.
2. **Browsing-dependent queries**: For queries requiring consultation of multiple results (common for exploratory queries with commercial or informational intents), the rating depends on the quality and coverage of the result set as a whole, with high quality and low friction yielding Highly Meets and partial coverage yielding Moderately Meets or below.
3. **Geographic mismatch override**: If the geographic mismatch is severe, the rating is capped. For YMYL queries with severe mismatch and low result quality, the rating is Fails to Meet; for non-YMYL queries, it is capped at Moderately Meets.
4. **Navigational queries**: If the dominant intent is navigational and the target page is present, the rating depends on the target's position and friction. Position 1 with the official target and low friction yields Fully Meets; positions 1–2 yield Highly Meets; positions 3–5 yield Moderately Meets; lower positions yield Slightly Meets.

Each rating decision is accompanied by a deterministic confidence score and a list of rule reasons, ensuring full transparency in how each SERP was rated.