Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Promptification of the Search Bar: Evidence of a Behavioral Spillover from Generative AI

Driss Masaoudi

driss.masaoudi@um6p.ma

Africa Business School, Mohammed VI Polytechnic University (UM6P)

Ravi Prakash Ranjan

Africa Business School, Mohammed VI Polytechnic University (UM6P)

Hugo Trad

Africa Business School, Mohammed VI Polytechnic University (UM6P)

Article

Keywords: Promptification, Generative AI, Large language models, Web search behavior, Google Trends, Google AI Overviews

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10917815/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

The Promptification of the Search Bar: Evidence of a Behavioral Spillover from Generative AI

Authors:

Driss Masaoudi*
Africa Business School, Mohammed VI Polytechnic University (UM6P), Morocco

Ravi Prakash Ranjan
Africa Business School, Mohammed VI Polytechnic University (UM6P), Morocco

Hugo Trad
Africa Business School, Mohammed VI Polytechnic University (UM6P), Morocco

*** Corresponding author:**

Driss Masaoudi
Africa Business School, Mohammed VI Polytechnic University (UM6P), Morocco
Email: driss.masaoudi@um6p.ma

ABSTRACT

Large language models (LLMs) have produced documented behavioral changes across information tasks, raising the question of whether they have also changed how users formulate requests to the legacy search bar. This study documents a durable structural shift in the syntactic register of web search queries following LLM diffusion, a shift termed "the promptification of the search bar" and characterized through panel co-movement and temporal precedence at population scale. Using 196,500 weekly US search-frequency observations from Google Trends across 25 semantic categories and 30 intent markers (2021 to 2025), this study classifies intent query markers into three classes of increasing conversational delegation and estimates their differential response to LLM diffusion. Conversational delegation rose by 58.7% after a structural break in June 2023, moving from no relation with LLM diffusion to tight co-movement, with temporal precedence running from diffusion to query syntax and not the reverse. An epoch decomposition shows the shift was established roughly eleven months before the platform deployed its own generative layer within the legacy tool, and the elasticity to LLM diffusion declined rather than increased after deployment, the opposite of what a platform-driven account predicts. The shift is strongest in domains serving cognitively complex needs, deepens over time as longer, more elaborate conversational forms standardize, and replicates worldwide. These findings are consistent with a spillover from LLM use into the legacy search bar: freed by exposure to conversational machines, users are returning to the natural register that keyword era had compressed, and the legacy interface is adapting to a behavior it did not create.

Keywords: Promptification, Generative AI, Large language models, Web search behavior, Google Trends, Google AI Overviews

INTRODUCTION

Large language models are changing how people seek, process, and produce information. Experimental evidence establishes that LLM use substantially improves productivity and substitutes for worker effort in professional writing, programming, and consulting tasks (Noy and Zhang 2023; Brynjolfsson et al. 2025), while significantly reducing task completion time in general information retrieval compared to traditional search engines (Xu et al. 2023). Studies on technology platforms indicate that LLM adoption reduces human contribution to public knowledge repositories; specifically, posting activity on Stack Overflow decreased by 25% within six months of ChatGPT's release (del Rio-Chanona et al. 2024). Longitudinal survey evidence documents widening demographic and socioeconomic divides in LLM usage over time (Angrisani et al. 2026; Bick et al. 2026). Experimental

comparisons of LLM-based and search-based learning demonstrate that the greater ease afforded by LLM synthesis reduces the depth of knowledge users develop and the originality of what they subsequently produce (Melumad and Yun 2025). Across these findings, a common behavioral mechanism emerges: LLMs alter how individuals externalize cognitive effort, increasing the delegation of mental work to machine interfaces (Risko and Gilbert 2016; Candrian and Scherer 2022; Skulmowski 2023; de Souza et al. 2024).

From the earliest days of web search, the interaction between user and engine imposed a behavioral adaptation on the human side. Although users initiate searches from what Belkin termed an "anomalous state of knowledge" (Belkin 1980; Belkin et al. 1982), representing a need inherently difficult to specify in advance. Consequently, they learn to accommodate index-based retrieval systems by omitting their diverse underlying intentions—whether informational, navigational, or transactional (Broder 2002; Rose and Levinson 2004; Marchionini 2006; Jansen et al. 2008). Instead, they formulate search-engine-friendly forms reduced to fragmented phrases. Empirical analyses of query logs systematically document this linguistic behavior: users exclude function words and rely heavily on nouns, producing queries that average fewer than three words (Barr et al. 2008). This brevity persists even for complex exploratory tasks(Jansen et al. 2008).

Even when an informational need is inherently interrogative, users formulate approximately 70% of these intents as compressed keyword queries because result relevance is historically indistinguishable across formats (Q. Liu et al. 2012; White et al. 2015), despite query formulation representing the most cognitively demanding stage of the search process (Gwizdka 2010; McGregor et al. 2023). This ingrained habituation manifests as a form of functional fixedness, which leaves non-experts struggling with prompt design in generative AI interactions (Zamfirescu-Pereira et al. 2023; J. Liu et al. 2025). Yet, keyword syntax is merely an acquired constraint rather than a cognitive default; developmental literature confirms that novice child searchers naturally use question-like conversational queries before adopting keyword compression (Kammerer and Bohnacker 2012; Bilal and Gwizdka 2018).

The history of search interface design reflects repeated attempts to recover natural language from this keyword compression. While the integration of voice search, semantic parsing, and knowledge graphs enabled systems to process question-formulated queries (Pang and Kumar 2011; Bondarenko et al. 2020; Aghaei et al. 2022), population-level keyword abbreviation persisted as the dominant behavioral norm (Barr et al. 2008; Athukorala et al. 2016; Aghaei et al. 2022). Because keyword search was deeply ingrained, users did not spontaneously abandon keyword syntax, demonstrating that natural language expression requires explicit conversational affordances (Papenmeier et al. 2021).

The public release of ChatGPT in November 2022 marked the first mass-market deployment of a Large Language Model (LLM) (Chatterji et al. 2025), introducing a widely accessible conversational interface that explicitly rewarded natural language input at population scale (Zamfirescu-Pereira et al. 2023; OpenAI et al. 2024; Zubair et al. 2025). By normalizing this register, LLMs established a new human-machine communication convention characterized by complex, contextual delegation (Hohenstein et al. 2023; Mannuru et al. 2024; Ye et al. 2024; Mo et al. 2025; Zhu et al. 2025; Siddiqui et al. 2026).

Despite these affordances, this cognitive delegation introduces significant informational risks. Because generative architectures synthesize outputs rather than retrieving verifiable documents, they expose users to hallucinations and obscure information provenance, limitations that persist even in Retrieval-Augmented Generation (RAG) systems (Lewis et al. 2020; Shah and Bender 2022; Memon and West 2024; Ravichander et al. 2025). Consequently, while LLM adoption has measurably reduced traditional search activity (Padilla et al. 2025)**,** a functional dichotomy has emerged in user behavior: individuals increasingly rely on LLMs for complex, analytical, and learning-oriented tasks (Caramancion 2024; Chatterji et al. 2025; Zubair et al. 2025), while legacy search engines, which still process billions of queries daily (Lolić et al. 2024; Ma et al. 2025), continue to be preferred for quick fact-finding, source verification, and navigational certainty (Hersh 2024; Kaiser et al. 2025).

Given this large and persistent reliance on legacy systems, it remains unknown whether the conversational behaviors normalized by LLMs are now generalizing to traditional search interfaces. We term this potential shift the *promptification* of the search bar: the progressive, population-scale emergence of conversational register forms in a context previously governed by keyword constraints.

While recent evidence demonstrates that LLMs displace specific information-seeking tasks (Memon and West 2024; del Rio-Chanona et al. 2024; Melumad and Yun 2025), testing whether they have structurally altered how users formulate requests to the search engine that remains is not straightforward. It requires population-scale behavioral

data with sufficient temporal resolution to detect structural breaks, a pre-treatment period long enough to establish credible reference distributions, and a panel-econometric strategy that distinguishes user-driven behavioral change from platform-driven interface change. Because no randomized or quasi-experimental counterfactual is available at population scale for an interface as ubiquitous as Google Search, our identification rests on observational panel-econometric tools (two-way fixed-effects, cointegration, and error-correction tests of temporal precedence) rather than on a strict counterfactual causal design.

To characterize this phenomenon, we pose three empirical questions: (i) Has the syntactic composition of legacy web search structurally shifted toward conversational delegation following mass LLM adoption? (ii) Does this behavioral shift exhibit predictive temporal precedence and long-run co-movement with generative AI diffusion? (iii) Is the magnitude of this shift uniform across domains, or is it ordered by the cognitive complexity of the underlying informational need? We address these questions using Google Trends as a population-scale behavioral corpus, extracting a weekly panel of 196,500 observations (2021–2025) spanning 30 intent markers across 25 semantic categories (comprising 24 substantive domains and an 'All Categories' aggregate), and tracking the progressive nesting cascade of high-imbrication conversational expressions.

We report four main findings. First, the syntactic composition of legacy web search shifted substantially toward conversational delegation following a June 2023 structural break: Conversational Delegation frequency rose by 58.7%, and a nesting cascade of high-imbrication conversational forms, absent from the pre-LLM baseline, emerged progressively thereafter. Second, and central to our identification, this behavioral shift preceded the platform's own deployment of generative search by approximately eleven months. Decomposing the timeline into three epochs, separated by the June 2023 LLM-adoption break and the May 2024 rollout of Google AI Overviews, we find that the conversational shift was already established during the intermediate window, when a conversational interface existed only outside Google and not yet within the search bar; when Google subsequently deployed its own generative interface, it added no detectable change to the response regime already in place. Third, because LLM adoption and conversational query frequency are I(1) and cointegrated (Phillips–Ouliaris), we test precedence within an error-correction model: LLM adoption is weakly exogenous and conversational query frequency adjusts to it, yielding one-directional short- and long-run precedence from diffusion to syntax. Fourth, domain-level shift magnitudes cluster into distinct profiles ordered by the cognitive complexity of the underlying task rather than by search volume.

Data and Methods introduce our corpus and statistical framework. *Results* presents the findings. *Discussion* examines implications for information-seeking behavior, web visibility, and the epistemics of human-machine communication.

DATA AND METHODS

Search Frequency Index Corpus

To measure the syntactic composition of US web search behavior, we extracted weekly Search Frequency Index (SFI) trajectories for 30 English-language intent markers from Google Trends for the United States from January 2021 through December 2025 (262 weeks). To test the generalizability of our findings, we conducted an independent replication on the Google Trends Worldwide corpus using the exact same markers, semantic categories, and LLM proxy (SI Appendix I). Because Google Trends normalizes search volume relative to the peak within the requested spatio-temporal window, this extraction yields a relative frequency measure that is strictly comparable across terms within the same sampling frame (Google 2026; Springer et al. 2026).

Our operationalization departs from the conventional use of search data. Rather than deploying search frequencies as indirect proxies to nowcast external epidemiological, economic, or sociological phenomena (Ginsberg et al. 2009; Choi and Varian 2012; Lolić et al. 2024), we treat the query syntax as the direct behavioral signal. This design isolates our dependent variables (the syntactic intent markers) from standard proxy measurement criticisms.

For our independent variable, we use the ChatGPT topic index as a real-time proxy for population-level LLM adoption. Extracting the predefined "topic" rather than a raw "search term" captures the unified semantic concept while filtering out vocabulary fragmentation and sampling noise (Djorno et al. 2026). This choice leverages established technology forecasting principles, which demonstrate that search traffic for a dominant product brand provides a highly accurate proxy for the diffusion curve of a new technology (Jun et al. 2014). Because both the

independent and dependent variables are drawn from the same sampling frame, their temporal alignment is exact by construction, and the cross-platform measurement friction that typically constrains conventional observational designs does not arise.

To construct this temporally aligned dataset, all 30 intent markers were extracted independently on the exact same day. This protocol neutralizes the risk of unannounced platform updates and the day-to-day sampling variability caused by Google Trends' 24-hour cache resets (Djorno et al. 2026). This extraction yields a panel of 196,500 marker-category-week observations; aggregated to class-level weekly series and restricted to the 24 substantive domains (excluding the All-Categories aggregate), the primary two-way fixed-effects (TWFE) specification is estimated on 18,864 class-domain-week observations. To ensure robustness to syntactic operationalization, we evaluate three extraction specifications. Each label combines a **match type** with a **marker set**. The match type is either EXACT (exact-match: the marker queried in quotation marks, so that Google Trends enforces strict sequential word order) or BROAD (broad-match: the unquoted marker, relaxing word order so that the component words may appear in any order and interspersed with others). The marker set is either 10kw (all ten keyword markers, designated as "kw", per class) or Top5 (the five highest-frequency markers per class). This yields **EXACT 10kw** (the primary and most conservative baseline), **EXACT Top5** (high-volume restricted), and **BROAD 10kw** (semantic upper bound). All main-text results rely on the EXACT 10kw baseline.

Because our identification relies on detecting longitudinal structural breaks, distinguishing behavioral shifts from platform-level measurement artifacts is critical. A January 2022 Google Trends data-collection update introduced a well-documented discontinuity in SFI series (K. Liu 2024; Lolić et al. 2024; Djorno et al. 2026; Springer et al. 2026). In our Bai-Perron (Bai and Perron 2003) estimation, this artifact precisely coincides with the first detected break in higher-order intent classes. Consequently, we isolate this event as a measurement artifact and anchor all behavioral interpretations to the June 2023 break, the first structural shift consistent with LLM adoption dynamics.

Finally, we account for two inherent measurement constraints explicitly documented by the platform: (i) sampling variability producing apparent zero values for low-frequency queries, mitigated by our robust Bai-Perron break structure and the P2 null result in our nesting analysis; and (ii) the platform's explicit filtering of internal queries generated within Google's own conversational products, such as Gemini, AI Mode, and AI Overviews (including modern voice inputs directed at these AI interfaces) (Reid 2025). By isolating our analysis to the legacy search bar, we ensure that our findings reflect directional and structural shifts in user query formulation rather than absolute volume changes or platform-redirected interactions.

Intent Classification

We operationalize the promptification phenomenon through a five-register syntactic scale (**Fig. 1**) that classifies query markers by their structural complexity and relational orientation toward the search engine. This scale rests on the established distinction in computational linguistics between *content units* (the core topical elements) and *intent units* (the relational structures articulating how the user wants the system to engage with the content) (Saha Roy et al. 2015).

Because legacy index-based systems historically forced users to compress naturally rich informational needs into bare content units, our scale measures the progressive re-introduction of the intent units that keyword compression had suppressed. The scale captures a continuous syntactic expansion, moving from *L1 Telegraphic* (single noun, no intent structure) through *L5 Agentive Delegation* (modal + pronoun + verb, full conversational structure).

To preclude arbitrary selection, we constructed a balanced panel of 30 intent markers (10 per class) strictly anchored to established information-seeking taxonomies (see **SI Appendix C.1, Supplementary Table 2**), for the complete inventory detailing their register, structural form, and behavioral extraction basis). As illustrated in **Fig. 1**, these markers operationalize three distinct behavioral regimes:

- **Class I (Factual Retrieval; L1):** Establishes the legacy keyword baseline. Structurally composed of isolated nouns or noun-phrases, this class captures the historical norm of treating the machine merely as a document index. The marker extraction is strictly derived from foundational web taxonomies: evaluative and ranked retrieval modifiers (e.g., *top, best, cheap*) are anchored in Broder (Broder 2002); transactional and fact-checking intents (e.g., *buy, hours, review*) follow Rose and Levinson (Rose and Levinson 2004); and relational-free navigational queries (e.g., *near me, location*) operationalize the framework of Jansen et al. (Jansen et al. 2008).

- **Class II (Directed Informational; L2):** Captures the intermediate transition to natural language. It expands into structured interrogatives (Wh- + noun/verb) and comparative operators that specify a response type, treating the engine as an information database rather than a conversational agent. Marker selection is driven by process and temporal query frameworks (e.g., *how to, when is*) from Pang and Kumar (Pang and Kumar 2011); comparative operators (e.g., *difference between, vs*) from Bondarenko et al. (Bondarenko et al. 2020); and complex factual exploration (e.g., *what is, meaning of*) from White et al. and Marchionini (Marchionini 2006; White et al. 2015).
- **Class III (Conversational Delegation; L3–L5):** Operationalizes the target promptification behavior. This class groups three advanced registers: cognitive operations acting on information (L3: e.g., explain, summarize), personal intent signaling (L4: personal pronoun + verb, e.g., help me with), and full agent interlocution (L5: modal + pronoun + verb, e.g., can you). In this regime, the user addresses the system as a responsive agent. The theoretical foundation for this cognitive offloading and agentic role-assignment stems from pre-LLM frameworks of conversational search by Radlinski and Craswell (Radlinski and Craswell 2017) and Risko and Gilbert (Risko and Gilbert 2016). To ensure empirical validity, the syntactic markers populating these registers (e.g., act as, write for me, generate a) are drawn from recent empirical studies of LLM-prompt formulation by Liu et al. (J. Liu et al. 2025) and Zamfirescu-Pereira et al. (Zamfirescu-Pereira et al. 2023).

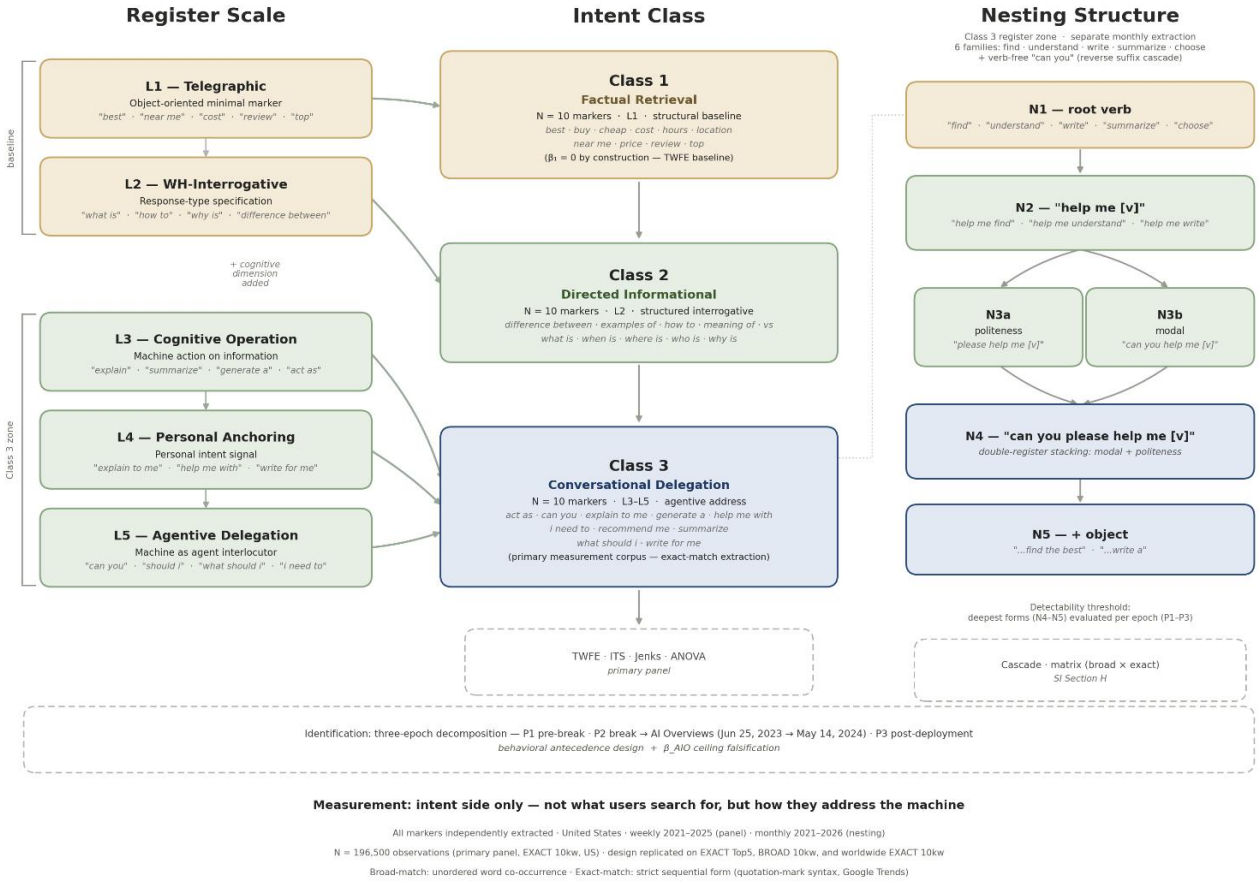


Fig. 1. Conceptual framework: five-register syntactic scale (L1–L5) mapped onto three intent classes, the three-epoch identification design (P1–P3: behavioral precedence and ceiling falsification), and the nesting cascade (six imbrication levels N1–N5; five intention-verb families plus the verb-free "can you" family).

Break-date anchoring

Endogenous structural break detection via the Bai-Perron algorithm identifies the primary inflection point for the LLM adoption proxy at June 11, 2023 (**SI Appendix E.6, Supplementary Table 13**). Exactly one week later (June 18, 2023), the broad "ChatGPT" semantic topic series crossed over the exact search term series in relative frequency, marking the population's cognitive maturation from querying a novel product string to engaging with a consolidated

generative AI concept (**SI Appendix B**). To capture this fully secure adoption baseline and to ensure a strictly symmetric 262-week observation window for our primary panel estimates, we formalize the treatment anchor at the week of June 25, 2023. This aligns the dataset into exactly 131 pre-treatment weeks and 131 post-treatment weeks. Substantively, this June 25 baseline closely aligns with the macroscopic plateau of global generative AI diffusion following its initial rapid growth, as documented by the World Bank (Y. Liu et al. 2025; Y. Liu and Wang 2026). Sensitivity analyses confirm that panel elasticity estimates are highly robust to anchor displacements; shifting the anchor backward to the exact Bai-Perron date, or extending the displacement up to ±8 weeks, yields highly stable coefficients; anchoring as far back as the week of the ChatGPT public release (December 1, 2022) leaves the estimates directionally stable and, if anything, slightly larger (**SI Appendix E.7, Supplementary Table 14**). Consequently, June 2023 serves as our exogenous treatment anchor, distinctly preceding the user-led behavioral break in Conversational Delegation endogenously detected in January 2024.

The Two-Period Panel Specification

We estimate the differential behavioral response of each intent class to LLM adoption using a two-way fixed-effects (TWFE) panel specification, complemented by an interrupted time-series (ITS) analysis at the macro level. All marker search frequency index (SFI) series are z-standardized relative to the pre-break parameters (μ_{pre}, σ_{pre}) of their own intent class, so that each coefficient measures deviation from that class's historical behavioral norm.

Our primary specification uses a global adoption proxy, $Z_{LLM,t}$, defined as the standardized "ChatGPT (Topic)" series, which is common to all domains and varying only over time. Stacking the three class-specific standardized series, with $Z_{Intent,i,c,t}$ the frequency of intent class $i \in \{I, II, III\}$ in category **c** at week **t**, we estimate

$$Z_{Intent,i,c,t} = \alpha_c + \gamma_t + \beta_2 (Z_{LLM,t} \times 1[i=II]) + \beta_3 (Z_{LLM,t} \times 1[i=III]) + \varepsilon_{i,c,t},$$

where α_c are category fixed effects absorbing time-invariant domain differences in baseline query composition, and γ_t are week fixed effects absorbing temporal shocks common across domains. Because $Z_{LLM,t}$ varies only over time, it is collinear with the week effects γ_t and is absorbed by them; β_2 and β_3 are therefore identified as the responsiveness of Classes II and III to LLM adoption relative to the Class I keyword baseline (the omitted reference). Standard errors are Driscoll–Kraay (Driscoll and Kraay 1998), robust to heteroskedasticity, serial correlation, and cross-sectional dependence, with bandwidth h = 4 weeks.

This first specification yields a contrast and, by its structure, cannot itself establish whether the Class I baseline responds to adoption: with a purely time-varying proxy, any baseline co-movement is folded into γ_t. To test rather than assume the behavior of the keyword baseline, we re-estimate the model with a category-internal proxy, $Z_{LLM,c,t}$, utilizing the "ChatGPT (Topic)" series extracted and standardized within each category (Z_{LLM_intra} in SI Appendix D). Because this proxy varies across domains within each week, it is no longer collinear with γ_t, and an absolute LLM elasticity is identified for every class, Class I included:

$$Z_{Intent,i,c,t} = \alpha_c + \gamma_t + \beta_I (Z_{LLM,c,t} \times 1[i=I]) + \beta_{II} (Z_{LLM,c,t} \times 1[i=II]) + \beta_{III} (Z_{LLM,c,t} \times 1[i=III]) + \varepsilon_{i,c,t}.$$

Here the Class I elasticity, though the smallest, is itself positive and significant ($\beta_I = 0.72$, $P < 0.001$): the keyword baseline is not inert. The three absolute elasticities form a strict monotone gradient ($\beta_I = 0.72$, $\beta_{II} = 1.23$, $\beta_{III} = 2.05$; all $P < 0.001$), with the conversational register responding roughly three times as strongly to LLM adoption as the keyword baseline. The implied differential of Class III over Class I ($\beta_{III} - \beta_I = 1.338$) reproduces the primary global-proxy estimate ($\beta_3 = 1.326$; equivalently, the intra-proxy differential $\beta_{3,intra} = 1.34$, $P = 1.37 \times 10^{-15}$). The headline differentials are therefore not an artifact of the omitted-baseline structure: estimating the baseline rather than treating it as a fixed reference leaves both the register gradient and the conversational-delegation differential intact.

To evaluate temporal dynamics, we first test the order of integration of Z_{LLM} and Class III frequency (augmented Dickey–Fuller and Zivot–Andrews (Zivot and Andrews 2002); SI Appendix E.2) and their long-run equilibrium relationship using the cointegration framework of Phillips and Ouliaris (Phillips and Ouliaris 1990). Because both series are I(1) and cointegrated, the Granger representation theorem requires that predictive precedence be tested within an error-correction model rather than on the raw levels. We therefore estimate a two-equation Engle–Granger error-correction model (Engle and Granger 1987) (p = 4 lags, HAC Newey–West standard errors (Newey and West 1987)) in which the lagged disequilibrium (error-correction term) enters both equations: short-run precedence is the joint Wald test on the lagged differences of the putative cause, and long-run precedence is the significance of the

error-correction (adjustment) coefficient, weak exogeneity of the forcing variable being indicated by a non-significant adjustment coefficient in its own equation. Levels-based Granger tests (Granger 1969) are reported as a naive cross-check only (SI Appendix E.4).

Domain-level heterogeneity in structural shift magnitude is analyzed using a one-dimensional Jenks natural-breaks classification (k = 3) applied to domain-level structural break magnitudes λ. Partition validity is assessed by silhouette score and confirmed by a one-way ANOVA on domain-level LLM synchronization r_intra. This synchronization, defined as the Pearson correlation between domain-level Z_Intent and the category-internal proxy Z_LLM,c,t post-break, serves as a convergent behavioral criterion withheld from the initial classification.

The Three-Epoch Precedence Design

The central identification question is directional: did the conversational shift in legacy search precede or follow the platform's deployment of a generative interface? A randomized experiment, withholding generative AI from part of the population, is not feasible for an interface as ubiquitous as Google Search. We therefore exploit a natural sequencing of two dated events to separate a user-led account from a platform-led one. The public diffusion of ChatGPT, external to Google, anchors a first event in June 2023; Google's own rollout of AI Overviews inside the search bar anchors a second, independent, platform-side event in May 2024. These two dates partition the observation window into three epochs: a pre-LLM baseline (P1, before June 2023), an intermediate window in which a conversational interface existed only outside Google (P2, June 2023 to May 2024), and a post-deployment window in which Google itself offered a generative interface (P3, after May 2024). The platform's data-collection protocol explicitly filters out internal searches executed by its own products, specifically AI Mode and AI Overviews (Google 2026). Because subsequent conversational turns within these generative interfaces are excluded, this structural filtering guarantees that the continued expansion of Conversational Delegation observed in P3 captures genuine, initial human input to the search bar rather than algorithmic artifacts.

The logic is straightforward. If the platform's interface change drove users toward conversational formulation, the behavioral shift should appear in P3, after Google deployed its interface. If instead users imported conversational habits acquired from external LLM exposure, the shift should already be present in P2, before any conversational affordance existed inside the search bar. We estimate the level shift of each intent class at each transition using two boundary-isolated regressions on the standardized class series. The first estimates the jump from the pre-LLM baseline to the intermediate window,

$$Z_Intent,t = a + \lambda_1 \cdot D_P2,t + u_t\ ,$$

where D_P2,t = 1 for observations in P2, so that λ_1 captures the behavioral shift that occurred before any platform-side change. The second estimates the jump from the intermediate to the post-deployment window,

$$Z_Intent,t = a' + \lambda_2 \cdot D_P3,t + u'_t\ ,$$

where D_P3,t = 1 for observations in P3, so that λ_2 captures any additional shift coinciding with the platform's deployment. A user-led account predicts $\lambda_1 > 0$ for conversational delegation, the shift being already present before Google acts; a platform-led account predicts that the shift is concentrated in λ_2, appearing only once Google deploys. The design would have falsified the promptification hypothesis under either of two observable patterns: if the conversational shift had been absent in P2 and appeared only in P3, or if the response regime had changed discontinuously at the platform-deployment boundary. The panel falsification specification interacts the LLM-adoption proxy with the Conversational Delegation indicator and a post-deployment indicator anchored at Google's AI Overviews rollout (May 14, 2024) (Reid 2024), estimated by two-way fixed effects with Driscoll-Kraay standard errors over the full observation window under the three-epoch partition (SI Appendix F). The full epoch decomposition, including joint multi-intervention and regime-switching specifications and formal Wald tests of regime continuity, is reported in SI Appendix F.

While the primary panel estimates rely on the core weekly dataset (ending December 2025) and use the core EXACT 10kw specification, the nesting cascade analysis employs an extended observation window (January 2021 to April 2026) aggregated at the monthly level. This temporal extension and monthly aggregation were specifically designed to maximize detection sensitivity for sparse, emergent high-imbrication conversational expressions. The nesting cascade tracks exact-match emergence across six imbrication levels (N1–N5, the third split into politeness and modal variants) for six root conversational verb families, measuring the lag from the June 2023 LLM adoption break to the first population-level detectability of high-imbrication forms.

Scope of identification. The panel-econometric design adopted here characterizes population-scale temporal precedence and panel co-movement between LLM adoption and conversational query frequency. We use the term *identification* in this paper in the panel-econometric sense, denoting the separation of differential responses from time-invariant category effects and from common temporal shocks via fixed effects. We do not use it in the strict Rubin causal model, instrumental-variable, or difference-in-differences counterfactual sense, which is structurally not available for an interface as ubiquitous and non-randomized as Google Search. Error-correction and Granger results are tests of predictive temporal precedence and adjustment dynamics, not of structural causation; we report them as such throughout.

Use of Large Language Models

During the preparation of this manuscript, the authors used Claude (Anthropic, Sonnet 4.6 and Opus 4.8) and Gemini (Google, 3.1) in a strictly limited assistive capacity. These tools helped write and debug the R programming code required for the econometric analyses. The authors wrote an initial Python/matplotlib script to plot the conceptual schematic in **Fig. 1**, and the tools iteratively enhanced and debugged this code. The authors manually wrote the initial Python/Graphviz code for an internal study architecture map visualizing the workflow from data extraction to final outputs, and AI assisted in enhancing the visualization parameters. The authors iteratively reviewed, debugged, and executed this final code locally. The tools corrected grammatical structures and refined English phrasing. For the Supporting Information, the authors manually integrated all data and wrote all text before using Claude to optimize the structure and numbering. The authors continuously refined and manually modified all AI textual outputs to produce the final version. AI generated no research ideas, data, empirical results, scientific analyses, or visual images. The authors extracted all search volume data manually via gtrendsR and produced all data visualizations via local code. The authors prepared all references manually using Zotero. The authors executed all computational scripts locally and reviewed all AI-assisted code and textual edits. A complete detailed declaration of AI usage across the research process, including the specific role of each tool at each stage, is provided in (**SI Appendix J**).
Authors take full responsibility for the integrity of the content and conclusions of this article.

RESULTS

Macro-Level Distributional Shift

Class III (Conversational Delegation) frequency increased by 58.7% following the June 2023 structural break, representing a distributional shift structurally distinct from any prior trend in the pre-LLM period (**Fig. 2**). Concurrently, Class I (Factual Retrieval) exhibits a weak leftward compression ($\Delta\mu$ = -3.2%), reflecting a slight decrease in the historical mean of the baseline telegraphic markers, in marked contrast to the steep rise of higher-register forms (**SI Appendix E.1, Supplementary Table 4**). This asymmetric pattern, characterized by Class III expanding sharply, Class I compressing modestly, and Class II (Directed Informational) shifting intermediately ($\Delta\mu$ = +15.3%), is directly visible in the macro ITS and constitutes the population-level behavioral signature of the promptification phenomenon.

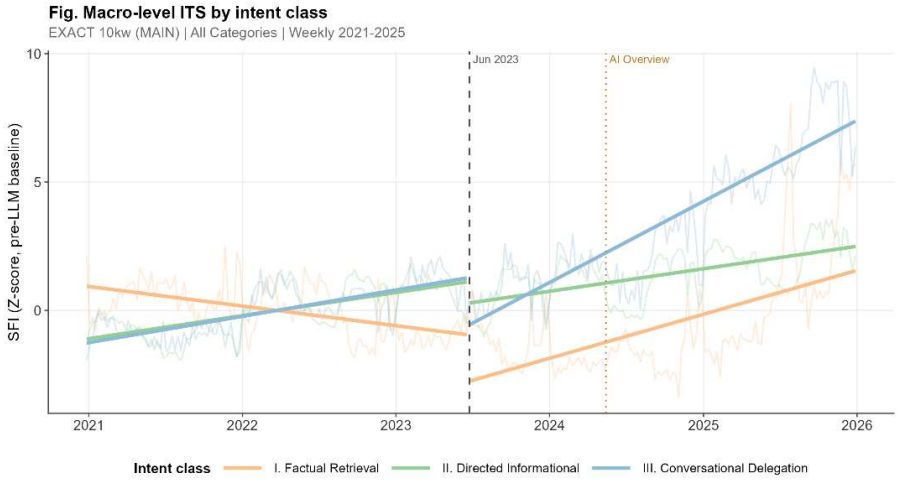


Fig. 2. Macro interrupted time-series: weekly SFI for three intent classes (EXACT 10kw, All Categories, US, January 2021–December 2025). Vertical dashed line: June 25, 2023 LLM adoption break anchor. January 2022 GT artifact break annotated. Shaded bands: 95% CIs on pre- and post-break means.

The Kolmogorov-Smirnov statistic for Class III reaches D = 0.657 (D^+, rightward shift), confirming that the post-break distribution cannot be characterized as a rescaled version of the pre-break distribution (**Fig. 3**). Distributional volatility expanded by 160.5%, reflecting a population that is not uniformly shifting but differentially adopting higher-register forms across domains. Pre-break co-movement between Class III frequency and the LLM adoption proxy was negligible (r_{pre} = -0.04), a mathematical reflection of the proxy's near-zero variance prior to mass adoption, confirming that historical Class III fluctuations were entirely independent of AI trends. Post-break correlation rises to r_{post} = 0.951, documenting tight co-movement between conversational syntax adoption and generative AI diffusion. Class II (Directed Informational) exhibits a consistent but more moderate shift ($\Delta\mu$ = +15.3%, D = 0.477, D^+). Class I (Factual Retrieval) shows a weak leftward compression ($\Delta\mu$ = -3.2%, D = 0.414, D^-) (**SI Appendix E.1, Supplementary Table 4**).

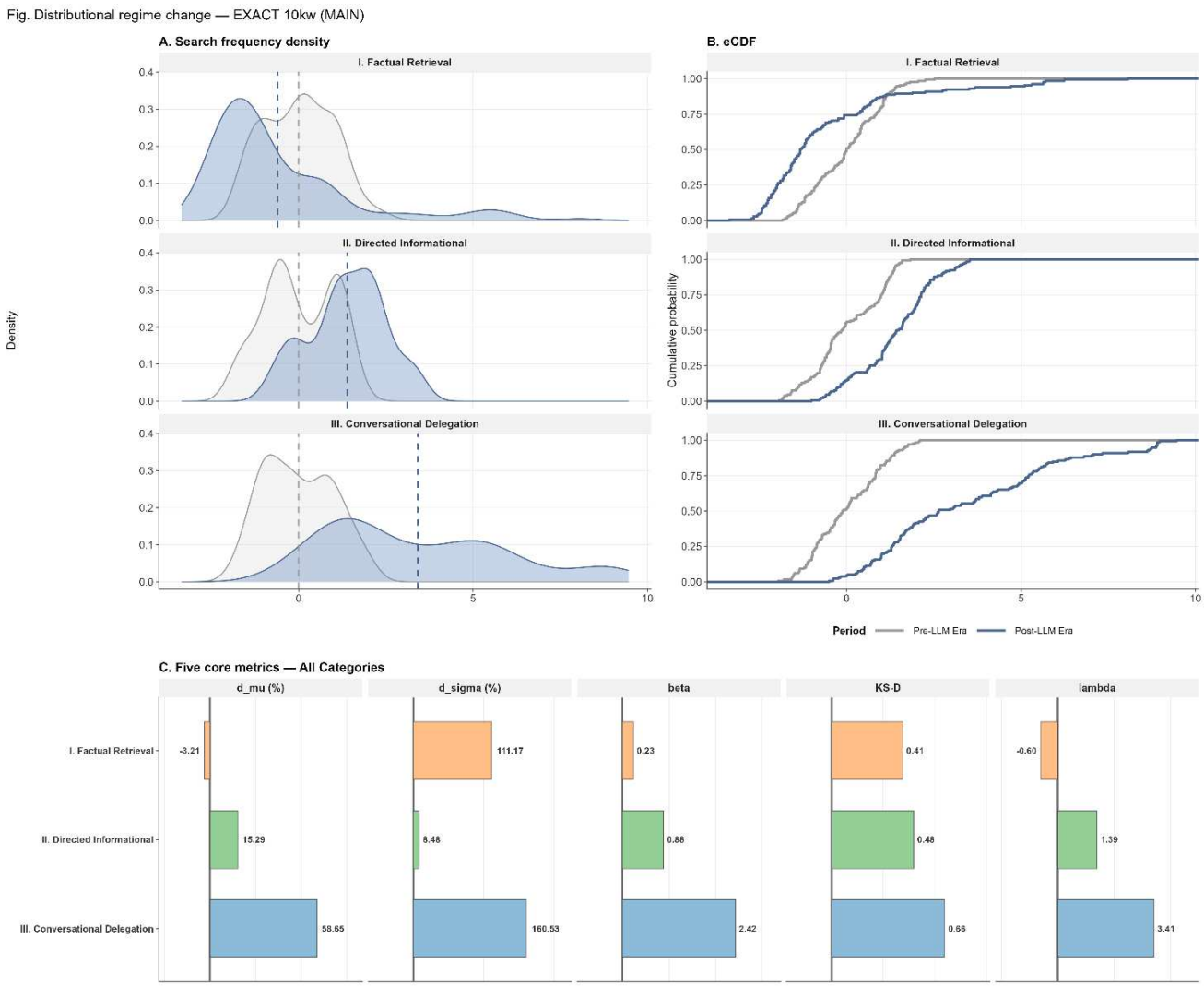


Fig. 3. Distributional dashboard: density plots (pre/post), empirical CDF, Δμ, Δσ, KS-D directional decomposition (D, D^+, D^-) for all three intent classes. All Categories, EXACT 10kw, US weekly 2021–2025.

Behavioral precedence: users led, the platform followed

The decomposition into three epochs isolates when the conversational shift occurred relative to the platform's own deployment of generative search (Methods; SI Appendix F). For Conversational Delegation, the shift is concentrated in the intermediate window, before Google deployed AI Overviews. The level jump from the pre-LLM baseline to the intermediate window is large and highly significant ($\lambda_1 = 1.276$, SE = 0.165, $P = 7.4 \times 10^{-13}$), establishing that the behavioral change was already in place while a conversational interface existed only outside Google. The subsequent jump across the platform-deployment boundary is also positive ($\lambda_2 = 3.312$, $P = 7.2 \times 10^{-15}$), but it represents the continued amplification of a shift already underway rather than its origin: the level of conversational delegation kept rising over time, while its responsiveness to further LLM adoption did not increase across the boundary. Estimated separately within each window, that responsiveness is statistically indistinguishable before and after deployment (3.58 in the intermediate window versus 2.91 afterward; difference not significant, P = 0.27), indicating that AI Overviews added no detectable change to the response regime already established (SI Appendix F).

The panel ceiling coefficient sharpens this falsification. In a two-way fixed-effects triple interaction, the marginal response of conversational delegation to LLM adoption after the deployment is significantly negative rather than positive ($\beta_{AIO} = -2.405$, SE = 0.354, $P = 1.15 \times 10^{-11}$ **in the US;** $\beta_{AIO} = -5.334$, SE = 1.272, $P = 2.8 \times 10^{-5}$ **worldwide**; SI Appendix F, Supplementary Table 18). A platform-led account predicts the opposite sign: an interface change that drove the shift should amplify, not depress, the marginal response to adoption.

The accompanying distributional moments trace the same two-stage process. In the intermediate window the mean of conversational delegation rises while its dispersion narrows ($\sigma = 0.87$ against the pre-LLM 1.00), the signature of a coherent, population-wide adoption of a new register rather than scattered experimentation. In the post-deployment window the mean rises far higher (to roughly +4.6 standard deviations above baseline) and dispersion widens ($\sigma = 2.49$), consistent with the broadening and diversification of conversational forms documented in the nesting cascade below. A user-led shift that consolidates first and diffuses later is exactly the pattern observed.

Two contrasts confirm that this is not an artifact of the instrument or a platform effect misread. First, the same decomposition applied to Factual Retrieval yields no stable regime: its level falls in the intermediate window and returns to its pre-LLM baseline in the post-deployment window (post-deployment level not significantly different from baseline, $P = 0.71$), fluctuating rather than establishing a durable shift. The model therefore detects a clear, datable break for Factual Retrieval when one exists, which makes the absence of a deployment-boundary regime change for Conversational Delegation a substantive finding rather than a limitation of sensitivity. Second, Directed Informational occupies the expected intermediate position, rising at both transitions but with its responsiveness to LLM adoption weakening after deployment. The resulting gradient, from Factual Retrieval through Directed Informational to Conversational Delegation, orders the three classes monotonically in both magnitude and regime stability, with only Conversational Delegation establishing and maintaining a continuous shift that began before the platform acted. The pattern is invariant across the EXACT 10kw, EXACT Top5, and BROAD 10kw specifications and replicates in the worldwide panel (SI Appendix F).

Differential Response and Panel Elasticity

To isolate this behavioral shift from time-invariant domain effects and common temporal shocks, we estimate the TWFE panel specification. The differential elasticity of Class III (Conversational Delegation) to LLM adoption is positive and highly significant ($\beta_3 = 1.326$, $P < 0.001$). Class II exhibits a smaller but significant differential sensitivity ($\beta_2 = 0.490$, $P = 0.015$), confirming a monotonic ordering of register complexity where the most advanced syntactic forms demonstrate the highest elasticity to AI diffusion (**Fig. 4** and **Table 1**). The structural break magnitude for Class III across all categories reaches $\lambda = 3.409$, documenting a large behavioral departure from the baseline. One-way ANOVA on domain-level structural break magnitudes confirms significant inter-class differentiation ($F = 73.09$, $P < 0.001$). Tukey HSD post-hoc tests establish significance for all three pairwise class comparisons: Class II vs. Class I ($\Delta\lambda = 1.357$, $P < 0.001$), Class III vs. Class I ($\Delta\lambda = 2.492$, $P < 0.001$), and Class III vs. Class II ($\Delta\lambda = 1.135$, $P < 0.001$). This systematic gradient between classes is thus statistically confirmed across domains by pairwise comparisons (**Fig. 5**).

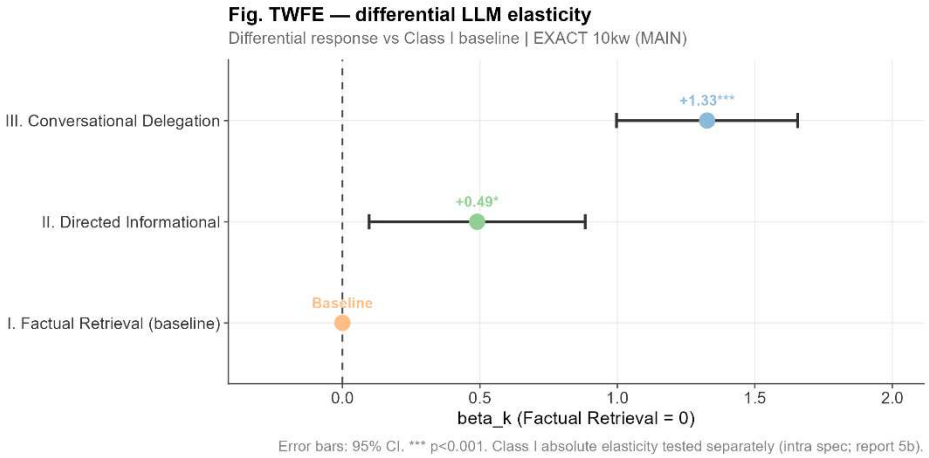


Fig. 4. TWFE differential LLM elasticity β by intent class (with 95% Driscoll-Kraay HAC CI), global and intra-specification panels side by side.

Table 1. TWFE differential LLM elasticity by intent class and syntactic specification.

Specification	Class	β	SE	P-value
EXACT 10kw (primary; most conservative)	I. Factual Retrieval	0 (baseline)	—	—
EXACT 10kw (primary; most conservative)	II. Directed Informational	0.490	0.200	0.015 *
EXACT 10kw (primary; most conservative)	III. Conversational Delegation	1.326	0.168	3.22×10^{-15} ***
EXACT 10kw (primary; most conservative)	Convergent intra-spec. β_3	1.338	—	1.37×10^{-15} ***
EXACT Top5 (high-volume restricted)	II. Directed Informational	0.708	0.147	1.39×10^{-6} ***
EXACT Top5 (high-volume restricted)	III. Conversational Delegation	1.545	0.135	2.76×10^{-30} ***
BROAD 10kw (semantic upper bound)	II. Directed Informational	0.691	0.187	2.19×10^{-4} ***
BROAD 10kw (semantic upper bound)	III. Conversational Delegation	1.640	0.164	1.41×10^{-23} ***

Driscoll-Kraay HAC-robust SE (Driscoll and Kraay 1998), bandwidth h = 4 weeks. The monotonic ordering $\beta_1 < \beta_2 < \beta_3$ holds across all three specifications and maps onto the gradient of prior syntactic constraint imposed by keyword retrieval. Convergent intra-specification estimate: β_3,intra = 1.338 ($P = 1.37 \times 10^{-15}$). * $P < 0.05$; *** $P < 0.001$.

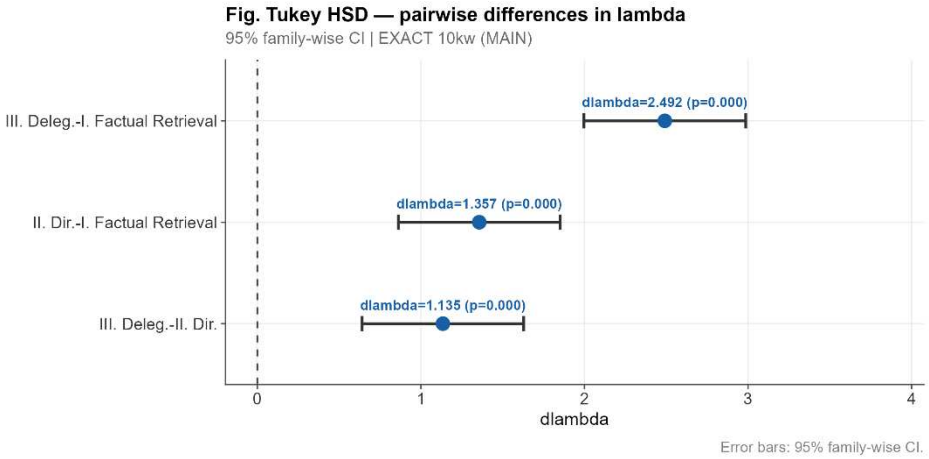


Fig. 5. Tukey HSD. Points = pairwise differences in mean λ between intent classes; Δλ and family-wise-adjusted P-values annotated; error bars = 95% family-wise CI. All three class pairs significant.

Temporal Precedence and Co-movement

Long-run behavioral equilibrium between LLM adoption and Class III query frequency is confirmed by Phillips–Ouliaris cointegration tests (P = 0.010) (SI Appendix E.3, Supplementary Table 6). Because both series are individually I(1) (ADF and Zivot–Andrews; **SI Appendix E.2**) and cointegrated, predictive precedence is identified within an error-correction model rather than on the raw series (Granger representation theorem). The error-correction term is significant only in the conversational-delegation equation (adjustment coefficient $\varphi = -0.155$, HAC SE 0.048, P = 0.002; adjustment half-life ≈ 4 weeks) and negligible in the adoption equation ($\varphi = -0.012$, P = 0.33): LLM adoption is weakly exogenous and conversational query frequency adjusts to it, establishing long-run precedence from diffusion to syntax. Short-run causality runs in the same single direction (joint Wald, LLM → syntax: F = 2.47, P = 0.045; reverse: F = 0.12, P = 0.98); a naive levels Granger test agrees (Direction A: F = 4.34, P = 0.002; Direction B: F = 0.51, P = 0.73) but is reported only as a cross-check given the unit roots (**SI Appendix E.4, Supplementary Table 7**). This temporal sequencing establishes that generative AI diffusion acts as the leading predictive indicator for the conversational shift in search behavior, a one-directional pattern that also replicates in the worldwide panel (**SI Appendix I.3**). Endogenous structural break detection via the Bai-Perron algorithm confirms the primary behavioral inflection point (Break 2) for Class III at January 14, 2024, structurally isolating it from the known January 2022 Google Trends data-collection artifact (**SI Appendix E.6, Supplementary Table 13**).

Domain-Level Behavioral Profiles

To determine whether the promptification phenomenon is an isolated artifact of specific topics or a broad-based behavioral change, we disaggregate the primary panel into 24 distinct semantic domains (the 25-category panel comprises these 24 substantive domains plus the "All Categories" aggregate). The endogenous structural break magnitude (λ) remains strictly positive across all 24 evaluated categories, and all domains record significant post-break correlations with the LLM adoption proxy (r_intra). This establishes that the conversational shift reflects varying intensities of adoption rather than a binary division of affected versus unaffected domains (**SI Appendix G.2**). Applying a one-dimensional Jenks natural-breaks optimization (k = 3, silhouette S = 0.543), we partition this continuous gradient into three distinct shift profiles: Dominant, Emerging, and Incipient (**Fig. 6**). This classification reflects the relative intensity of the behavioral response, which is ordered by the cognitive nature of the dominant informational need rather than simple search volume (Wildemuth et al. 2018). For instance, while specialist fields like Science register as Incipient (λ = 1.246), they still exhibit a substantial structural break and post-break synchronization. In contrast, consumer-oriented domains like Sports (λ = 3.628) and Shopping (λ = 3.555), where informational needs are often open-ended and benefit disproportionately from synthesis, exhibit the largest Dominant shifts (**Fig. 7**). The statistical validity of this emergent profiling is strongly confirmed by a one-way ANOVA on the convergent domain-level LLM synchronization criterion (r_intra) withheld from the initial classification (F = 19.43, P < 0.001) (**Fig. 8**).

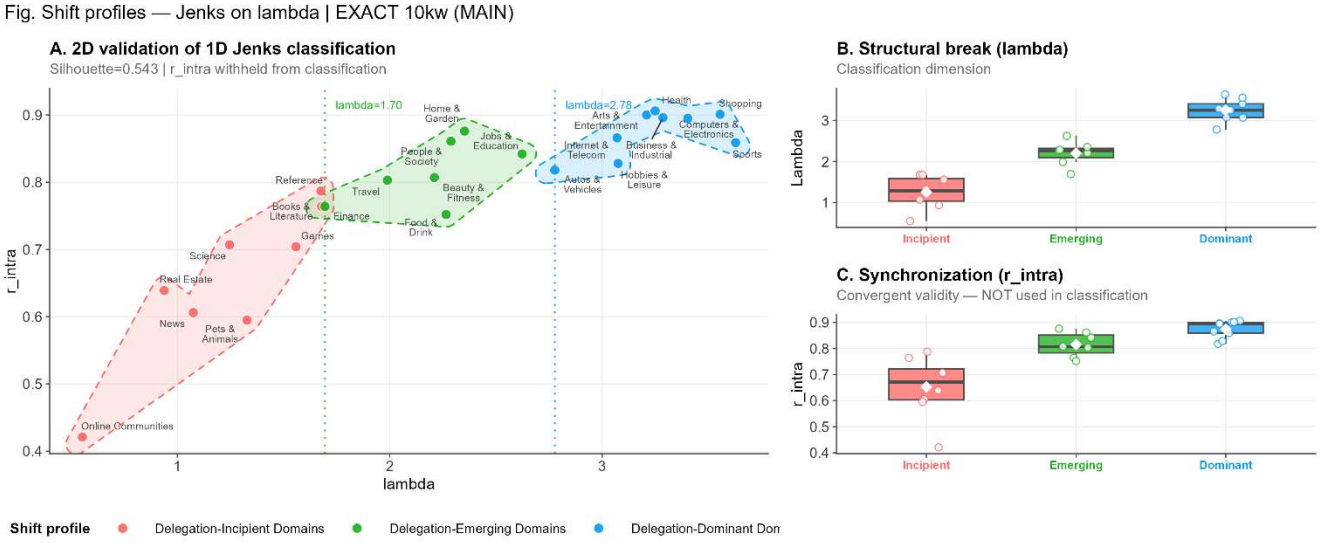


Fig. 6. Jenks shift profiles: λ vs. r_intra scatter for 24 domains, colored by profile. Boxplots of λ and r_intra by profile. Thresholds λ_IE = 1.695 and λ_ED = 2.777 annotated.

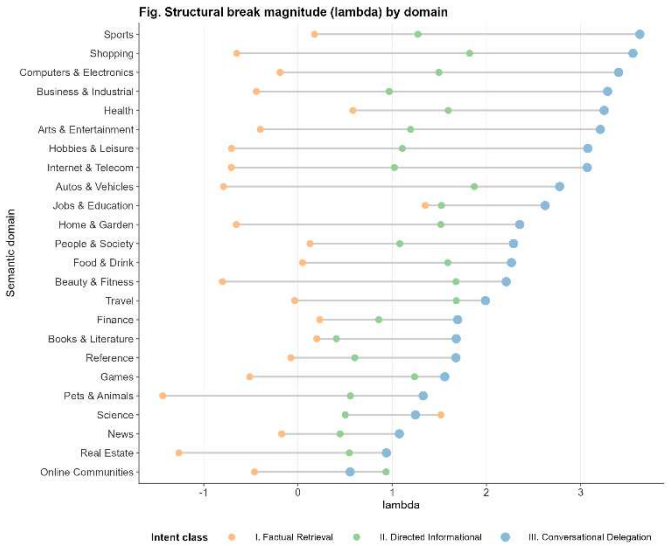


Fig. 7. Lambda lollipop*: 24 domains ordered by structural break magnitude λ (Class 3), colored by profile. Vertical lines at Jenks thresholds.*

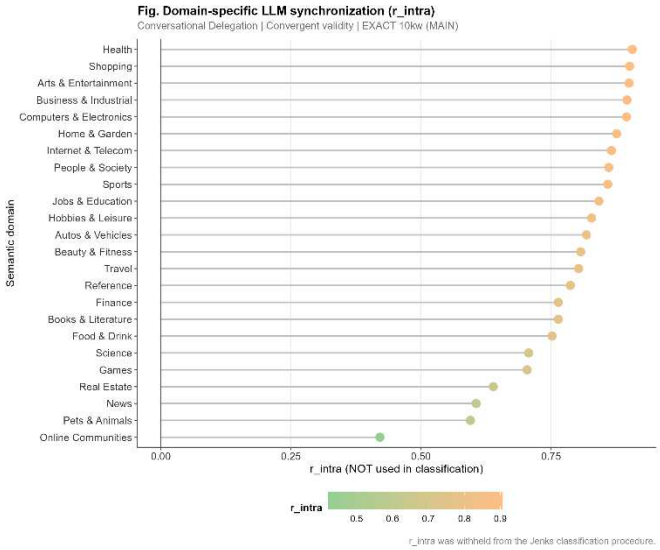


Fig. 8. r_intra synchronization*: 24 domains ordered by post-break LLM synchronization, colored by profile. Error bars: 95% CI.*

Progressive Syntactic Emergence

Having established the broad-based nature of the shift, we trace its micro-structural evolution through a nesting cascade analysis of six conversational verb families. We evaluate the standardization of promptification across three intersecting dimensions: temporal progression, syntactic imbrication depth (N1 to N5), and the convergence between broad-match semantic intent and exact-match syntactic formulation. For the five base verbs, the third imbrication level bifurcates to capture distinct politeness (N3a) and modal (N3b) trajectories, while the "can you" family follows a direct five-level progression. We decompose the observation window into three distinct behavioral periods anchored on the two pivotal events of the study. These consist of a 29-month pre-LLM baseline from January 2021 to May 2023 (P1), an 11-month window of mass LLM adoption from June 2023 to April 2024 (P2), and a 24-month post-AIO window from May 2024 to April 2026 (P3).

The combined detectability matrix (**Fig. 9**), supported by period-faceted heatmaps detailing exact-match and broad-match trajectories across all verb families (**SI Appendix H.3**), reveals a clear chronological and imbrication-stratified sequencing. To quantify this structural shift, we define detectability as the percentage of months within a given period registering non-zero search volume. For example, a 75% detectability in the 24-month P3 window indicates the expression successfully crossed the population-level measurement threshold in 18 distinct months. Among 18 high-imbrication exact syntactic forms completely absent in the P1 baseline (0% detectability), zero new forms achieved population-level detectability during the 11-month P2 window (**SI Appendix H.1, Supplementary Table _22**). Instead, the structural shift manifests exclusively in P3 through two concurrent mechanisms measured via distinct metrics. First, pre-existing conversational forms that already registered baseline volume in P1 intensify sharply. We measure this intensification using the ratio of the mean Search Frequency Index (SFI) between periods, with P3/P1 volume ratios reaching 9.6 for "help me write", 5.4 for "help me understand", and 4.1 for "can you help me find" (**SI Appendix H.1, Supplementary Table _24**). Second, 10 of the 18 advanced forms that were previously absent emerge in P3 following a strictly imbrication-stratified lag measured from the June 2023 anchor. The shortest stabilization lag is 13 months for forms like "can you please help me find", extending to 15 months for "can you help me write" and up to 24 months for "please help me write"; four additional forms register a first single-month detection only in the final month of the window (April 2026, lag 34 months), at the edge of detectability (**SI Appendix H.1, Supplementary Table 25**).

To ensure this structural absence in P2 represents a genuine behavioral void rather than an artifact of Google Trends algorithmically censoring low-frequency queries (K. Liu 2024; Google 2026; Springer et al. 2026), we deliberately treat the platform's volume-reporting rule not as a limitation, but as a strict measurement threshold for population-level detectability. By comparing broad-match and exact-match detectability (**SI Appendix H.2, Supplementary**

Table 26), we observe that general semantic intent successfully crossed this visibility threshold in P2, whereas its exact-match counterparts remained at absolute zero. This divergence provides direct evidence of the promptification mechanics: the underlying conversational intent existed and was actively searched post-LLM, but the specific multi-component syntactic structures required to express it had not yet achieved the critical mass of standardization necessary to register as a shared population-level norm.

This structural zero in P2 confirms that users adopted the conceptual intent of conversational delegation well before standardizing the exact syntactic formulations required to articulate it. We compare broad-match and exact-match detectability across periods. Historically, broad-match forms representing general semantic intent were already highly detectable, whereas their exact-match counterparts were almost completely absent. Consequently, a sharp reduction in this gap demonstrates that the exact word order of these expressions is gaining population-level traction, signaling a syntactic standardization toward queries characteristic of promptification. This comparison reveals a distinct convergence driven by imbrication depth. For intermediate levels, the detectability gap collapses significantly: the gap for "can you help me write" drops from 100 percentage points in P1 and P2 to just 25 percentage points in P3. Similarly, the gap for "can you help me understand" contracts from 97 to 29 percentage points. However, at higher imbrication levels (N4 and N5), the gap remains wide or diverges, indicating that deeper syntactic complexity requires longer habituation to standardize (Saha Roy et al. 2016). Inherently, prompt formulation is a highly personalized and idiosyncratic process, as users engage with generative systems through individualized natural language to satisfy specific cognitive needs (Zamfirescu-Pereira et al. 2023; Zubair et al. 2025). The delayed but decisive emergence of these multi-component expressions in P3, alongside the sharp contraction of the semantic-syntactic gap at intermediate levels, indicates that idiosyncratic natural language prompts are progressively coalescing into shared syntactic conventions.



Fig. 9. Combined nesting detectability matrix: six verb families × imbrication levels N1–N5. Cell value: % non-zero months across P1–P3. Color: ≥80% dark green → 0% grey. ★ = first detectable in P3 post-AIO (May 2024). Zero new emergence in P2 confirmed across all N3a–N5 cells (18 forms absent in P1 EXACT; 10 emerge in P3, 8 remain below threshold).

This late, ordered emergence of long high-imbrication forms is consistent with a continuing deepening of conversational use rather than a fresh wave of new adopters, a mechanism we develop in the Discussion.

Robustness to Syntactic Specifications

The primary panel findings are highly robust across alternative syntactic specifications. The monotonic ordering of differential elasticity ($\beta_1 < \beta_2 < \beta_3$) holds consistently under high-volume restricted (EXACT Top5) and semantic upper-bound (BROAD 10kw) specifications (**Table 1**). The EXACT 10kw specification provides the most conservative lower bound on the behavioral shift ($\beta_3 = 1.326$). When restricted to the highest-volume markers (EXACT Top5), the Class III elasticity expands ($\beta_3 = 1.545$), and it reaches its maximum magnitude under the semantic upper bound (BROAD 10kw, $\beta_3 = 1.640$). Endogenous structural break detection across these specifications reveals a clear chronological sequencing: the primary behavioral break for Class III emerges as early as September 3, 2023, under the BROAD 10kw specification, structurally preceding the January 14, 2024 break identified in the strict EXACT 10kw panel (**SI Appendix E.6, Supplementary Table 13**). This confirms that the observed promptification is not an artifact of low-frequency marker volatility, but a deep structural shift where broad semantic adoption systematically preceded strict syntactic standardization.

Worldwide Generalization and Domain Stability

The population-scale promptification phenomenon is not a US-specific platform artifact. A complete replication on worldwide search data preserves the core qualitative findings and confirms the universality of the domain gradient. The endogenous Bai-Perron algorithm identifies the primary behavioral structural break for Class III at the exact identical date (January 14, 2024), anchored against an identical LLM adoption plateau (June 25, 2023) (**SI Appendix I.5, Supplementary Table 30**). Across domains, all 24 evaluated semantic categories maintain a strictly positive structural break magnitude ($\lambda > 0$), with the aggregate Class III magnitude actually expanding in the worldwide panel ($\lambda = 4.349$). The Jenks domain ordering is qualitatively preserved, with fields like Health, Shopping, and Sports remaining Dominant, and News and Real Estate remaining Incipient, even as classification thresholds rescale upward (**SI Appendix I.6, Supplementary Table 31**).

While the aggregate worldwide panel elasticity is moderately attenuated ($\beta_3 = 0.994$, $P < 0.001$), reflecting heterogeneous global LLM exposure, the Class II differential effect collapses entirely ($P = 0.186$). This heterogeneity isolates the conversational register (Class III) as the sole syntactic form emerging following generative AI diffusion at the cross-linguistic global scale (**SI Appendix I.2, Supplementary Table 27**). The directional precedence likewise replicates: in the worldwide error-correction model, LLM adoption remains weakly exogenous and conversational query frequency adjusts to it (long-run adjustment $\varphi = -0.117$, $P = 0.0002$; reverse direction non-significant), confirming that the lead from diffusion to syntax is not US-specific (**SI Appendix I.3, Supplementary Table _29**). Finally, the panel ceiling coefficient is not only preserved but amplified worldwide ($\beta_{AIO} = -5.334$, SE=1.272, $P=2.8 \times 10^{-5}$), consistent with a user-led behavioral spillover that preceded platform interface adaptations globally (**SI Appendix F, Supplementary Table 18; SI Appendix I.7**).

DISCUSSION

These findings document a durable, population-scale shift in how users formulate requests to the dominant web search interface. The temporal sequencing, the cross-geographic replication, and the eleven-month lag between the behavioral break and the platform's deployment of AI Overviews together support a user-led reading: the conversational shift is a behavioral spillover from external exposure to large language models, not an adaptation to a platform interface change. The historical pattern in which users adapt their formulations to the machine's constraints has, in this episode, run in the opposite direction: the population's external habituation to generative AI preceded, and was then accommodated by, the legacy platform. Three results that might be read separately, the continued rise in conversational delegation after the platform's deployment, the diminishing marginal response to LLM adoption captured by the negative ceiling coefficient, and the late emergence of long, ordered high-imbrication forms. They are best understood as a single mechanism observed through two distinct quantities. The level of conversational delegation, measured against time, tracks how far the behavior has moved. Its responsiveness to LLM adoption, measured against the adoption proxy, tracks how tightly the behavior is still being pulled upward by new exposure. These two quantities answer different questions, and after the platform's deployment they move apart: the level keeps rising while the responsiveness flattens. Stated precisely, the deployment of AI Overviews precedes a level shift in mean conversational frequency, but not a regime shift; the elasticity to LLM diffusion dries

up even as the correlation remains strong. This divergence has a natural reading. Independent evidence on generative-AI usage indicates that growth in usage volume increasingly outpaced growth in the user base, with the active-user base expanding only modestly from mid-2023 onward even as usage volume rose several-fold (Y. Liu et al. 2025; Y. Liu and Wang 2026). A plausible mechanism, consistent with this evidence among other explanations, is that the marginal source of promptification shifted over time: early on, the behavior was fed by new users carrying conversational habits into the search bar, which a responsiveness measure tied to adoption captures directly; later, as the inflow of new users slowed, the behavior was sustained and deepened by the existing population. A measure tied to the adoption proxy registers this slowdown as a ceiling, whereas a measure tied to time continues to rise, because the deepening of use by existing adopters does not require fresh adoption to proceed. The negative ceiling coefficient and the continued rise in level are therefore not in tension; they are two views of the same transition.

This account also makes sense of the nesting cascade and assigns the platform interface its proper role. An interface that elicits natural language lowers the friction of an already latent behavior; it does not create the behavior (Papenmeier et al. 2021). If late-period growth is driven by existing users deepening their use rather than by new users joining, the progressive standardization of long, ordered, personalized conversational expressions is the expected fine-grained signature of that deepening: the same population, increasingly fluent in the conversational register, settles into more elaborate and more sequentially exact forms over time. Two features of the measurement are consistent with this reading. First, because Google Trends records the query that initiates a search and not any subsequent conversational turn, what is captured even after the platform deployed generative search remains largely the entry-level register that users bring to the bar. Second, the absence of a regime change at the deployment boundary, against the clear, datable break the same model detects for factual retrieval, indicates that the platform's interface did not initiate the behavior it later accommodated.

Beyond the syntactic record itself, the pattern documented here carries implications for the wider information ecosystem, which we flag as directions rather than findings. Complex conversational queries are increasingly resolved directly on the results page through generative summaries; emerging evidence indicates that such summaries reallocate attention away from publisher pages (Zheng et al. 2025; Agarwal and Sen 2026; Grossman et al. 2026; Huang et al. 2026; Khosravi and Yoganarasimhan 2026; Rolon-Merette et al. 2026), that content producers are beginning to optimize for citation by generative engines rather than for classical ranking (Ma et al. 2025), and that LLM adoption itself reshapes broader online behavior (Padilla et al. 2025). In this environment, the deployment of AI Overviews reads naturally as a strategic accommodation of a demand shift that the present data show was already under way; the platform's own framing of the rollout, "let Google do the searching for you" (Reid 2024), is consistent with retaining increasingly conversational query traffic on the results page, a deployment whose competitive and regulatory dimensions are now actively debated (Ng and Wessel 2026; Singh and Scott Morton 2026), and whose underlying content supply is itself becoming contested, as major web-infrastructure providers move to block AI crawlers by default absent compensation (Prince 2025). One year into the deployment, the platform itself reported that people were asking "more complex, longer and multimodal questions" (Reid 2025). The platform dates this change to its own launch; the decomposition reported here locates its onset roughly eleven months earlier. The two records are compatible: the post-deployment rise the platform observes is real, and appears in our data as a continued level shift; what the deployment did not do is increase the behavior's responsiveness to LLM adoption. The reading consistent with both is that of an interface that accommodated, and may have helped consolidate and standardize, a register already in place rather than one it originated. The trajectory of accommodation has since continued past our observation window: in May 2026 the platform announced a redesign of the search box itself, dynamically expanding to accommodate longer, more descriptive input (Reid 2026), an interface evolution that follows the syntactic shift documented here rather than preceding it.

The same dynamics define the limits of the present method. As formulations become more personalized and idiosyncratic, a growing share of them falls below the volume-reporting threshold of public indices, taking the finest-grained syntax of human curiosity progressively out of macroscopic view; the late, threshold-crossing emergence documented in the nesting cascade is the visible edge of this process. Population-level syntactic indexation of the kind used here may therefore be most informative precisely in the window this paper covers. Downstream questions, including how conversational delegation alters depth of learning and knowledge formation (Spatharioti et al. 2023; Bastani et al. 2025; Melumad and Yun 2025) and how trust in AI-mediated information is calibrated (Venkit et al. 2024), will require individual-level and experimental designs rather than further observational indexation.

Several limitations qualify these claims. The design is observational and population-scale; it establishes temporal precedence and panel co-movement rather than a counterfactual causal effect. The shift persists across our observation window; only future studies can establish whether it is permanent. The mechanism linking the slowing of new-user growth to the deepening of use among existing users is supported by external usage evidence but is not measured directly in the search data, and is therefore offered as one coherent explanation among others. The marker set is in English and the worldwide aggregate is dominated by Latin-alphabet ecosystems, so generalization to specific non-English national search systems remains an open empirical question (SI Appendix I). Finally, the Search Frequency Index reflects relative rather than absolute query frequencies, which limits interpretation of absolute levels while leaving structural-break identification and elasticity estimation intact

REFERENCES

Agarwal, Saharsh, and Ananya Sen. 2026. Google AI Overviews and Publisher Traffic: Evidence from a Field Experiment. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.6513059.

Aghaei, Sareh, Kevin Angele, Elwin Huaman, Geni Bushati, Mathias Schiestl, and Anna Fensel. 2022. Interactive Search on the Web: The Story So Far. *Information* 13. Multidisciplinary Digital Publishing Institute: 324. https://doi.org/10.3390/info13070324.

Angrisani, Marco, Maria Casanova, Nathanael J Fast, Jimmy Narang, and Juliana Schroeder. 2026. Gaps in large language model awareness, usage, and perceptions in the United States: Evidence from a nationally representative longitudinal survey. *PNAS Nexus* 5: pgag007. https://doi.org/10.1093/pnasnexus/pgag007.

Athukorala, Kumaripaba, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. 2016. Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology* 67: 2635–2651. https://doi.org/10.1002/asi.23617.

Bai, Jushan, and Pierre Perron. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18: 1–22. https://doi.org/10.1002/jae.659.

Barr, Cory, Rosie Jones, and Moira Regelson. 2008. The linguistic structure of English web-search queries. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 1021–1030.

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2025. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences* 122. Proceedings of the National Academy of Sciences: e2422633122. https://doi.org/10.1073/pnas.2422633122.

Belkin, N.J. 1980. Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science* 5: 133–143.

Belkin, N.J., R.N. Oddy, and H.M. Brooks. 1982. Ask for information retrieval: part I. background and theory. *Journal of Documentation* 38: 61–71. https://doi.org/10.1108/eb026722.

Bick, Alexander, Adam Blandin, and David J. Deming. 2026. The Rapid Adoption of Generative AI. *Management Science*. INFORMS. https://doi.org/10.1287/mnsc.2025.02523.

Bilal, Dania, and Jacek Gwizdka. 2018. Children's query types and reformulations in Google search. *Information Processing & Management* 54: 1022–1041. https://doi.org/10.1016/j.ipm.2018.06.008.

Bondarenko, Alexander, Pavel Braslavski, Michael Völske, Rami Aly, Maik Fröbe, Alexander Panchenko, Chris Biemann, Benno Stein, and Matthias Hagen. 2020. Comparative Web Search Questions. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, 52–60. WSDM '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3336191.3371848.

Broder, Andrei. 2002. A taxonomy of web search. *SIGIR Forum* 36: 3–10. https://doi.org/10.1145/792550.792552.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. Generative AI at Work. *The Quarterly Journal of Economics* 140: 889–942. https://doi.org/10.1093/qje/qjae044.

Candrian, Cindy, and Anne Scherer. 2022. Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior* 134: 107308. https://doi.org/10.1016/j.chb.2022.107308.

Caramancion, Kevin Matthe. 2024. Large Language Models vs. Search Engines: Evaluating User Preferences Across Varied Information Retrieval Scenarios. arXiv. https://doi.org/10.48550/arXiv.2401.05761.

Chatterji, Aaron, Thomas Cunningham, David J. Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. How People Use ChatGPT. Working Paper. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w34255.

Choi, Hyunyoung, and Hal Varian. 2012. Predicting the Present with Google Trends. *Economic Record* 88: 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x.

Djorno, Candice, Mauricio Santillana, and Shihao Yang. 2026. Restoring the forecasting power of Google Trends with statistical preprocessing. *International Journal of Forecasting* 42: 1104–1122. https://doi.org/10.1016/j.ijforecast.2026.03.001.

Driscoll, John, and A.C. Kraay. 1998. Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Rev. Econ. Stat.* 80: 549–560.

Engle, Robert, and Clive Granger. 1987. Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55. Econometric Society: 251–76.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457. Nature Publishing Group: 1012–1014. https://doi.org/10.1038/nature07634.

Google. 2026. FAQ about Google Trends data - Trends Help.

Granger, C. W. J. 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37. [Wiley, Econometric Society]: 424–438. https://doi.org/10.2307/1912791.

Grossman, Riley, Songjiang Liu, Michael K. Chen, Mike Smith, Cristian Borcea, and Yi Chen. 2026. How Generative AI Disrupts Search: An Empirical Study of Google Search, Gemini, and AI Overviews. arXiv. https://doi.org/10.48550/arXiv.2604.27790.

Gwizdka, Jacek. 2010. Distribution of cognitive load in Web search. *Journal of the American Society for Information Science and Technology* 61: 2167–2187. https://doi.org/10.1002/asi.21385.

Hersh, William. 2024. Search still matters: information retrieval in the era of generative AI. *Journal of the American Medical Informatics Association* 31: 2159–2161. https://doi.org/10.1093/jamia/ocae014.

Hohenstein, Jess, Rene F. Kizilcec, Dominic DiFranzo, Zhila Aghajari, Hannah Mieczkowski, Karen Levy, Mor Naaman, Jeffrey Hancock, and Malte F. Jung. 2023. Artificial intelligence in communication impacts language and social relationships. *Scientific Reports* 13. Nature Publishing Group: 5487. https://doi.org/10.1038/s41598-023-30938-9.

Huang, Michelle, Agam Goyal, Koustuv Saha, and Eshwar Chandrasekharan. 2026. Answer Bubbles: Information Exposure in AI-Mediated Search. *arXiv.org*. March 17.

Jansen, Bernard J., Danielle L. Booth, and Amanda Spink. 2008. Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management* 44: 1251–1266. https://doi.org/10.1016/j.ipm.2007.07.015.

Jun, Seung-Pyo, Jaeho Yeom, and Jong-Ku Son. 2014. A study of the method using search traffic to analyze new technology adoption. *Technological Forecasting and Social Change* 81: 82–95. https://doi.org/10.1016/j.techfore.2013.02.007.

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. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–7. CHI EA '25. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3706599.3720123.

Kammerer, Yvonne, and Maja Bohnacker. 2012. Children's web search with Google: the effectiveness of natural language queries. In *Proceedings of the 11th International Conference on Interaction Design and Children*, 184–187. IDC '12. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2307096.2307121.

Khosravi, Mehrzad, and Hema Yoganarasimhan. 2026. Impact of AI Search Summaries on Website Traffic: Evidence from Google AI Overviews and Wikipedia. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.6164926.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, et al. 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems*, 33:9459–9474. Curran Associates, Inc.

Liu, Jiqun, Jamshed Karimnazarov, and Ryen W. White. 2025. Trapped by Expectations: Functional Fixedness in LLM-Enabled Chat Search. arXiv. https://doi.org/10.48550/arXiv.2504.02074.

Liu, Kerry. 2024. The measurement errors of google trends data. *Discover Data* 2: 7. https://doi.org/10.1007/s44248-024-00013-3.

Liu, Qiaoling, Eugene Agichtein, Gideon Dror, Yoelle Maarek, and Idan Szpektor. 2012. When web search fails, searchers become askers: understanding the transition. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 801–810. SIGIR '12. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2348283.2348390.

Liu, Yan, Jingyun Huang, and He Wang. 2025. Who on Earth Is Using Generative AI? Global Trends and Shifts in 2025. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.5495020.

Liu, Yan, and He Wang. 2026. Who on earth is using Generative AI? *World Development* 199: 107260. https://doi.org/10.1016/j.worlddev.2025.107260.

Lolić, Ivana, Marina Matošec, and Petar Sorić. 2024. DIY google trends indicators in social sciences: A methodological note. *Technology in Society* 77: 102477. https://doi.org/10.1016/j.techsoc.2024.102477.

Ma, Lijia, Juan Qin, Xingchen Xu, and Yong Tan. 2025. When Content is Goliath and Algorithm is David: The Style and Semantic Effects of Generative Search Engine. arXiv. https://doi.org/10.48550/arXiv.2509.14436.

Mannuru, Nishith Reddy, Aashrith Mannuru, and Brady Lund. 2024. Large Language Models (LLMs) as a Tool to Facilitate Information Seeking Behavior. *InfoScience Trends* 1. InfoPub Press: 34–42. https://doi.org/10.61186/ist.202401.01.15.

Marchionini, Gary. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49: 41–46. https://doi.org/10.1145/1121949.1121979.

Masaoudi, Driss, Ravi Prakash Ranjan, and Hugo Trad. 2026. [dataset] Promptification of the search bar: Evidence of a Behavioral Spillover from Generative AI. Zenodo. https://doi.org/10.5281/zenodo.20812633.

McGregor, Molly, Leif Azzopardi, and Martin Halvey. 2023. A Systematic Review of Cost, Effort, and Load Research in Information Search and Retrieval, 1972–2020. *ACM Trans. Inf. Syst.* 42: 2:1-2:39. https://doi.org/10.1145/3583069.

Melumad, Shiri, and Jin Ho Yun. 2025. Experimental evidence of the effects of large language models versus web search on depth of learning. *PNAS Nexus* 4: pgaf316. https://doi.org/10.1093/pnasnexus/pgaf316.

Memon, Shahan Ali, and Jevin D. West. 2024. Search Engines Post-ChatGPT: How Generative Artificial Intelligence Could Make Search Less Reliable. arXiv. https://doi.org/10.48550/arXiv.2402.11707.

Mo, Fengran, Kelong Mao, Ziliang Zhao, Hongjin Qian, Haonan Chen, Yiruo Cheng, Xiaoxi Li, Yutao Zhu, Zhicheng Dou, and Jian-Yun Nie. 2025. A Survey of Conversational Search. *ACM Trans. Inf. Syst.* 43: 167:1-167:50. https://doi.org/10.1145/3759453.

Newey, W. K, and K. D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix* 55. Malden, MA: Blackwell: 703–708.

Ng, Robin, and Michael Wessel. 2026. *AI Overview or Overreach? Google's Strategic Deployment of Generative AI in Search*. Working paper. CRC TR 224 Discussion Paper Series. University of Bonn and University of Mannheim, Germany.

Noy, Shakked, and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381. American Association for the Advancement of Science: 187–192. https://doi.org/10.1126/science.adh2586.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, et al. 2024. GPT-4 Technical Report. arXiv. https://doi.org/10.48550/arXiv.2303.08774.

Padilla, Nicolas, H. Tai Lam, Anja Lambrecht, and Brett Hollenbeck. 2025. The Impact of LLM Adoption on Online User Behavior. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.5393256.

Pang, Bo, and Ravi Kumar. 2011. Search in the Lost Sense of "Query": Question Formulation in Web Search Queries and its Temporal Changes. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, ed. Dekang Lin, Yuji Matsumoto, and Rada Mihalcea, 135–140. Portland, Oregon, USA: Association for Computational Linguistics.

Papenmeier, Andrea, Dagmar Kern, Daniel Hienert, Alfred Sliwa, Ahmet Aker, and Norbert Fuhr. 2021. Starting Conversations with Search Engines - Interfaces that Elicit Natural Language Queries. In *Proceedings of*

the 2021 Conference on Human Information Interaction and Retrieval, 261–265. CHIIR '21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3406522.3446035.

Phillips, P. C. B., and S. Ouliaris. 1990. Asymptotic Properties of Residual Based Tests for Cointegration. *Econometrica* 58. [Wiley, Econometric Society]: 165–193. https://doi.org/10.2307/2938339.

Prince, Matthew. 2025. Content Independence Day: no AI crawl without compensation! *The Cloudflare Blog*. July 1.

Radlinski, Filip, and Nick Craswell. 2017. A Theoretical Framework for Conversational Search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, 117–126. CHIIR '17. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3020165.3020183.

Ravichander, Abhilasha, Shrusti Ghela, David Wadden, and Yejin Choi. 2025. HALoGEN: Fantastic LLM Hallucinations and Where to Find Them. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, ed. Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, 1402–1425. Vienna, Austria: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-long.71.

Reid, Elizabeth. 2024. Generative AI in Search: Let Google do the searching for you. *Google*. May 14.

Reid, Elizabeth. 2025. AI in Search: Going beyond information to intelligence. *Google*. May 20.

Reid, Elizabeth. 2026. A new era for AI Search. *Google*. May 19.

del Rio-Chanona, R Maria, Nadzeya Laurentsyeva, and Johannes Wachs. 2024. Large language models reduce public knowledge sharing on online Q&A platforms. *PNAS Nexus* 3: pgae400. https://doi.org/10.1093/pnasnexus/pgae400.

Risko, Evan F., and Sam J. Gilbert. 2016. Cognitive Offloading. *Trends in Cognitive Sciences* 20: 676–688. https://doi.org/10.1016/j.tics.2016.07.002.

Rolon-Merette, Thadde, Alexander John karran, Eugene Yuzan Guo, Mehdi Benbousta, Fabien Poivre, Sylvain Amoros, Constantinos K. Coursaris, Pierre-Majorique Léger, and Sylvain Senecal. 2026. The Zero Click Horizon? Cognitive Precursors to Behavioral Changes in AI-Assisted Web Search. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.6442189.

Rose, Daniel E., and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, 13–19. WWW '04. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/988672.988675.

Saha Roy, Rishiraj, Smith Agarwal, Niloy Ganguly, and Monojit Choudhury. 2016. Syntactic complexity of Web search queries through the lenses of language models, networks and users. *Information Processing & Management* 52: 923–948. https://doi.org/10.1016/j.ipm.2016.04.002.

Saha Roy, Rishiraj, Rahul Katare, Niloy Ganguly, Srivatsan Laxman, and Monojit Choudhury. 2015. Discovering and understanding word level user intent in Web search queries. *Journal of Web Semantics* 30. Semantic Search: 22–38. https://doi.org/10.1016/j.websem.2014.07.010.

Shah, Chirag, and Emily M. Bender. 2022. Situating Search. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval*, 221–232. CHIIR '22. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3498366.3505816.

Siddiqui, Emad, Venkatesh Terikuti, and Xuan Lu. 2026. LLM-based Semantic Search for Conversational Queries in E-commerce. arXiv. https://doi.org/10.48550/arXiv.2601.16492.

Singh, Madhavi, and Fiona M. Scott Morton. 2026. A Roadmap for a Monopolization Case Against Google: Monopsony Power and AI Overviews. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.6303280.

Skulmowski, Alexander. 2023. The Cognitive Architecture of Digital Externalization. *Educational Psychology Review* 35: 101. https://doi.org/10.1007/s10648-023-09818-1.

de Souza, Bruno Campello, Agostinho Serrano de Andrade Neto, and Antonio Roazzi. 2024. The generative AI revolution, cognitive mediation networks theory and the emergence of a new mode of mental functioning: Introducing the Sophotechnic Mediation scale. *Computers in Human Behavior: Artificial Humans* 2: 100042. https://doi.org/10.1016/j.chbah.2024.100042.

Spatharioti, Sofia Eleni, David M. Rothschild, Daniel G. Goldstein, and Jake M. Hofman. 2023. Comparing Traditional and LLM-based Search for Consumer Choice: A Randomized Experiment. arXiv. https://doi.org/10.48550/arXiv.2307.03744.

Springer, Steffen, Artur Strzelecki, Branden Holmes, Janine M. Ziermann-Canabarro, Martin Kaatz, and Michael Zieger. 2026. Setting trends with google: limits and perspectives when utilising search engine data. *Quality & Quantity* 60: 3923–3942. https://doi.org/10.1007/s11135-025-02431-0.

Venkit, Pranav Narayanan, Philippe Laban, Yilun Zhou, Yixin Mao, and Chien-Sheng Wu. 2024. Search Engines in an AI Era: The False Promise of Factual and Verifiable Source-Cited Responses. arXiv. https://doi.org/10.48550/arXiv.2410.22349.

White, Ryen W., Matthew Richardson, and Wen-tau Yih. 2015. Questions vs. Queries in Informational Search Tasks. In *Proceedings of the 24th International Conference on World Wide Web*, 135–136. WWW '15 Companion. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2740908.2742769.

Wildemuth, Barbara M., Diane Kelly, Emma Boettcher, Erin Moore, and Gergana Dimitrova. 2018. Examining the impact of domain and cognitive complexity on query formulation and reformulation. *Information Processing & Management* 54: 433–450. https://doi.org/10.1016/j.ipm.2018.01.009.

Xu, Ruiyun Rayna, Yue Feng, and Hailiang Chen. 2023. ChatGPT vs. Google: A Comparative Study of Search Performance and User Experience. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.4498671.

Ye, Dezhi, Jie Liu, Jiabin Fan, Bowen Tian, Tianhua Zhou, Xiang Chen, and Jin Ma. 2024. Enhancing Asymmetric Web Search through Question-Answer Generation and Ranking. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 6127–6136. KDD '24. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3637528.3671517.

Zamfirescu-Pereira, J.D., Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny Can't Prompt: How Non-AI Experts Try (and Fail) to Design LLM Prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–21. CHI '23. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3544548.3581388.

Zheng, Shuang, Yuting Zhu, Xin Ye, and Liang Shen. 2025. Generative Search: Evidence from a Large-Scale Field Experiment. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.5297194.

Zhu, Yutao, Huaying Yuan, Shuting Wang, Jiongnan Liu, Wenhan Liu, Chenlong Deng, Haonan Chen, Zheng Liu, Zhicheng Dou, and Ji-Rong Wen. 2025. Large Language Models for Information Retrieval: A Survey. *ACM Trans. Inf. Syst.* 44: 12:1-12:54. https://doi.org/10.1145/3748304.

Zivot, Eric, and Donald W. K. Andrews. 2002. Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics* 20. Taylor & Francis: 25–44. https://doi.org/10.1198/073500102753410372.

Zubair, Misbahu, Muhammad Abubakar Alhassan, and Farid Bello. 2025. From Queries to Prompts: Comparing User Experience in Generative AI Tools and Search Engines. *Proceedings of BCS HCI 2025*. ScienceOpen. https://doi.org/10.14236/ewic/BCSHCI2025.13.

Data availability

All data, code, and materials required to reproduce the analyses reported in this paper are publicly available in a Zenodo repository ref.(Masaoudi et al. 2026): https://doi.org/10.5281/zenodo.20812633, released under a Creative Commons Attribution 4.0 (CC BY 4.0) license. The repository contains the derived Google Trends search-frequency series (CSV) and the complete R and Python analysis and figure-generation code, together with documentation (README.md, the study-architecture data-flow diagram study_architecture.png, and the Supplementary-label correspondence file SI_label_correspondence.csv). The underlying series were extracted from the public Google Trends service (https://trends.google.com). No restrictions apply.

Author contributions

D.M. and R.P.R. conceived the study and developed the methodology. D.M. curated the data, performed the formal analysis, and prepared the visualizations. D.M. and R.P.R. conducted the investigation. R.P.R. and H.T. provided resources and supervised the research, with R.P.R. also managing project administration. D.M. wrote the original manuscript. All authors reviewed and edited the manuscript and approved the final version.

Declarations

Funding

This research did not receive funding.

Competing interests

The authors declare no competing interests.

Ethics declaration

This study did not involve human participants, human data, or animal subjects. All data used are publicly available, aggregate search-frequency indices with no personally identifiable information.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MasaoudiSupplementaryInformation.pdf