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

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

SI Appendices:

A: Historical macro baseline

B: LLM proxy validation

C: Marker composition + macro paired analysis

D: Statistical framework and notation

E: Full statistical battery, two-period design (E.1–E.8)

F: Three-epoch decomposition and behavioral precedence (Supplementary Tables 15, 16, 17, 18, 19; Supplementary Figure 10)

G: Domain taxonomy and shift universality

H: Nesting cascade (three-epoch decomposition P1/P2/P3 (N3a/N3b notation); Supplementary Tables 21-26; 6 shaded family figures + 2 period-faceted heatmaps (Script 1); 6 × 2×2 trajectory panels (Script 2; primary: find, can you; replication: write, understand, summarize, choose); 6 per-family detectability matrices + 1 combined matrix (Script 3))

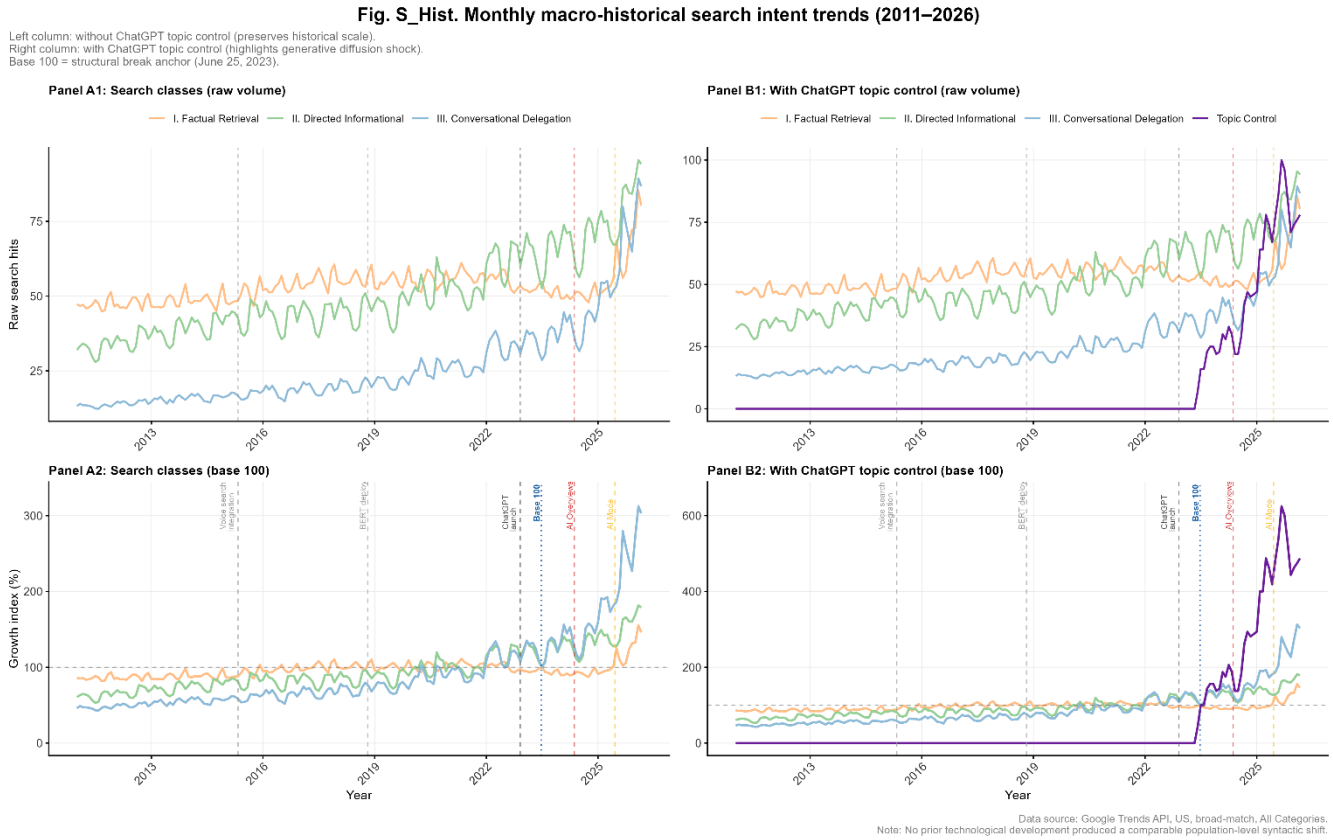
I: Worldwide robustness and generalization (Supplementary Tables 27-36; Supplementary Figure 31)

J: AI usage declaration

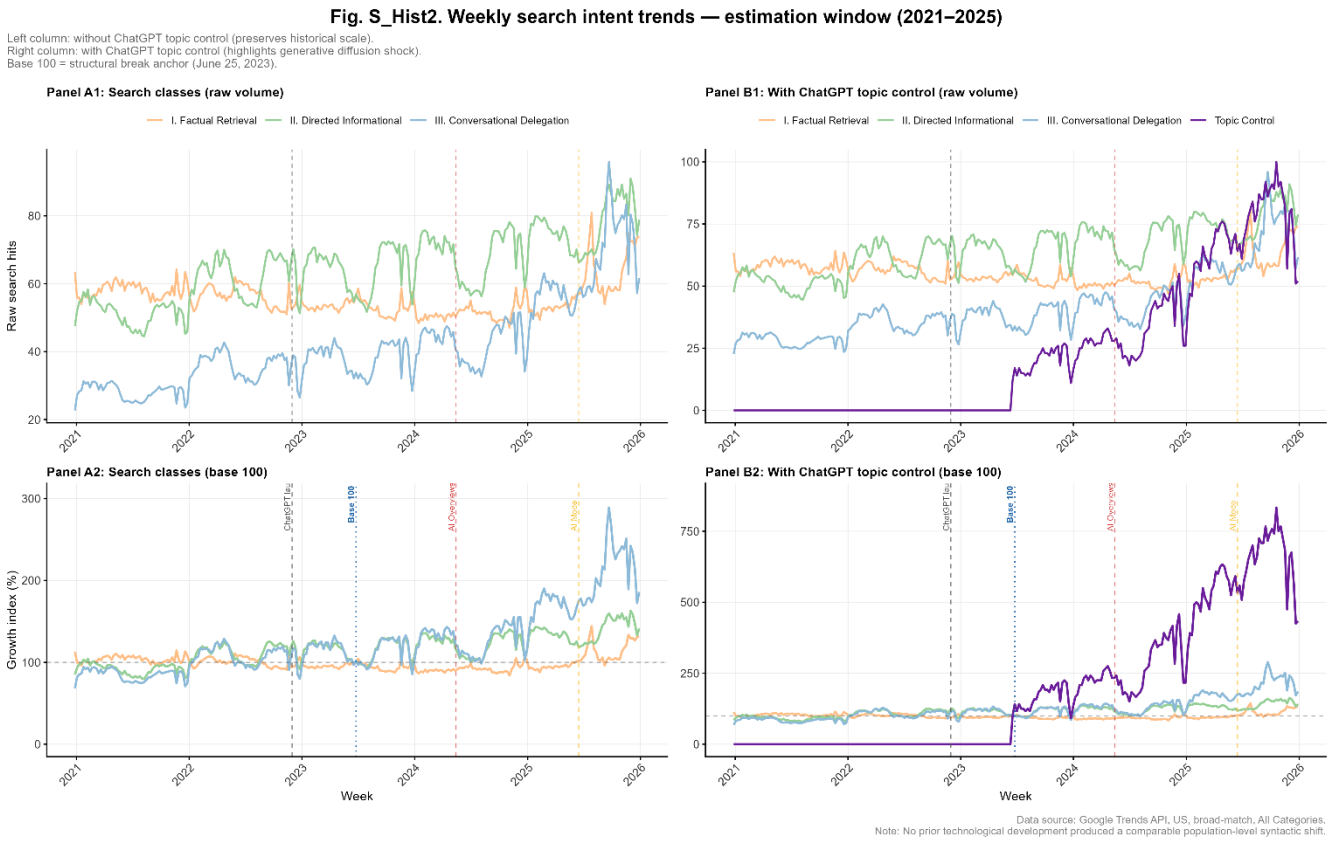
SI Appendix A: Historical Macro Baseline

Supplementary Figure 1 presents the monthly SFI time series for the three intent classes (broad-match, All Categories, United States, January 2011–December 2026), providing 15 years of behavioral context. The pre-LLM period (2011–2022) shows a gradual and modest increase in Class 2 (Directed Informational) frequency, consistent with longitudinal query log evidence (Pang and Kumar 2011). Class 3 (Conversational Delegation) remains near-zero throughout the pre-LLM period, establishing a structural absence of agent-directed query forms before LLM diffusion. Class 1 (Factual Retrieval) is stable across the full window, confirming a compositional register shift rather than a general volume inflation.

Supplementary Figure 2 presents the same series at weekly resolution (2021–2025) with vertical markers for key technological events: ChatGPT release (November 2022), adoption plateau peak (March–April 2023), June 2023 LLM adoption break anchor, Google AI Overviews rollout (May 2024), and Google AI Mode deployment (early 2025).



Supplementary Figure 1. Monthly SFI for three intent classes (broad-match, All Categories, US, January 2011-March 2026). Class 3 near-zero pre-LLM; sharp post-2023 acceleration distinct from any prior trend.



Supplementary Figure 2. Weekly SFI (2021-2025) with vertical technological event markers. June 2023 behavioral break confirmed as distinct from January 2022 GT artifact.

SI Appendix B : LLM Adoption Proxy Validation

The primary LLM adoption proxy is the ChatGPT topic SFI (All Categories, US, weekly). To validate this choice, we conducted a joint extraction from the Google Trends interface, which enforces cross-series relative scaling (all series simultaneously scaled to a shared window, enabling direct absolute frequency comparison).

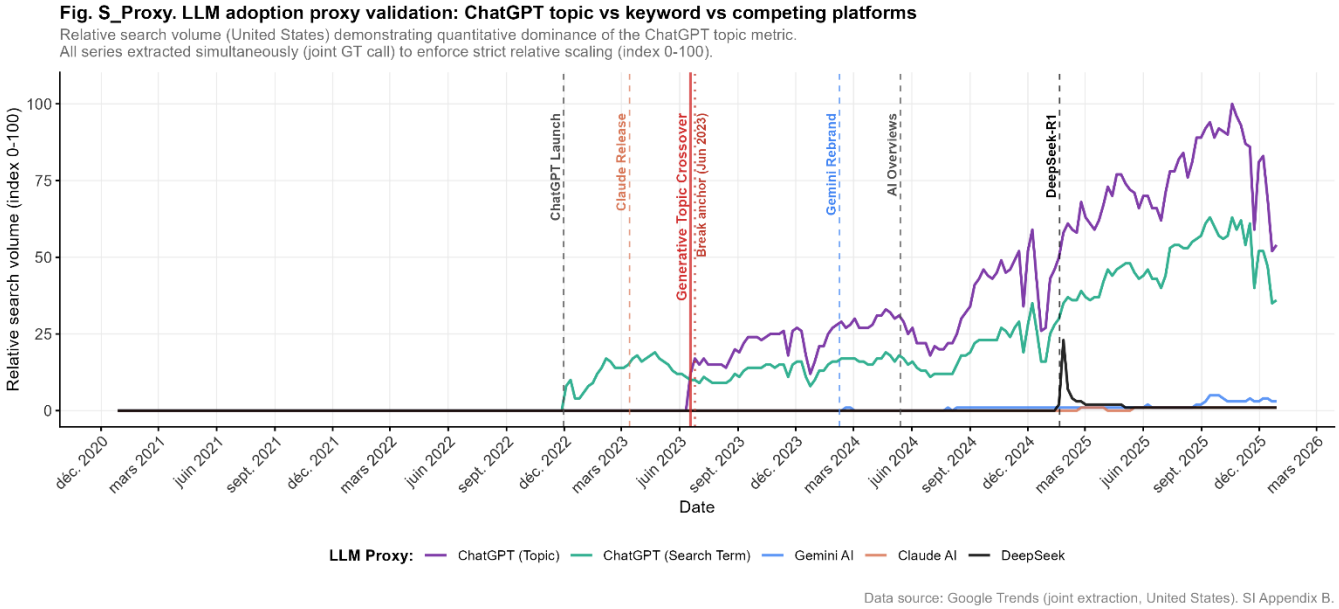
Supplementary Table 1 reports the inter-proxy Pearson correlation matrix and volume statistics for five LLM indicators, including the distinction between the broad "ChatGPT (Topic)" and the exact "ChatGPT (Search Term)". The Topic proxy maintains the highest mean SFI (23.1) and max volume (100) across the full period, substantially exceeding the second-ranked brand (Gemini AI: mean 0.47). The inclusion of emerging models (Claude AI, DeepSeek) confirms that ChatGPT remains the overwhelmingly dominant population-level adoption signal throughout the observation window.

Supplementary Table 1. Inter-proxy Pearson correlation matrix and joint-scaled volume statistics (SFI) for LLM adoption indicators.

Proxy	ChatGPT (Topic)	ChatGPT (Term)	Gemini AI	Claude AI	DeepSeek	Mean SFI	Max SFI
ChatGPT (Topic)	1.000	0.967	0.807	0.773	0.355	23.10	100
ChatGPT (Search Term)	0.967	1.000	0.808	0.794	0.346	15.60	63
Gemini AI	0.807	0.808	1.000	0.760	0.247	0.47	5
Claude AI	0.773	0.794	0.760	1.000	0.216	0.15	1
DeepSeek	0.355	0.346	0.247	0.216	1.000	0.36	23

Note: Joint extraction methodology enforces a shared scale. ChatGPT (Topic) and (Search Term) are highly correlated ($r = 0.967$, $P < 0.001$), but the Topic captures the broader semantic footprint of the technology.

Crucially, the joint extraction reveals a precise cognitive maturation timeline for the population. Following the initial product release, the exact search term initially concentrated the search volume. However, on **June 18, 2023**, the "ChatGPT (Topic)" series crossed over the "ChatGPT (Search Term)" series in relative volume. This crossover date; occurring exactly one week after the endogenous Bai-Perron structural break (June 11, 2023) and one week prior to our formalized symmetrical treatment anchor (June 25, 2023); validates our anchor placement. It ensures our baseline captures a fully consolidated plateau where generative AI had transitioned from a specific keyword query to a mature, dominant semantic concept.



Supplementary Figure 3. ChatGPT topic SFI vs. alternative LLM proxy trajectories (2021–2025).

SI Appendix C: Marker Composition and Macro Paired Analysis

C.1 Complete Marker List

Supplementary Table 2 presents the complete inventory of the 30 intent markers utilized in our analysis, detailing their precise syntactic register, intent class, structural form, and behavioral basis. To preclude arbitrary selection and ensure empirical validity, marker extraction was strictly anchored to established information-seeking taxonomies and empirical LLM-prompting studies, as detailed in the main text. The overarching distinction between content units and intent units follows (Saha Roy et al. 2015).

Supplementary Table 2: Complete list of 30 intent markers: register, class, structural form, and behavioral extraction basis

Intent **Class**	**Register**	**Marker**	**Structural Form**	**Behavioral basis & Source**
I. Factual Retrieval (Document index)	L1	best	Noun / Noun-phrase	Content superlative (Broder 2002)
	L1	buy	Noun / Noun-phrase	Transactional (Rose and Levinson 2004)
	L1	cheap	Noun / Noun-phrase	Cost-oriented (Broder 2002)
	L1	cost	Noun / Noun-phrase	Price retrieval (Rose and Levinson 2004)

	L1	hours	Noun / Noun-phrase	Navigational fact (Rose and Levinson 2004)
	L1	location	Noun / Noun-phrase	Navigational (Jansen et al. 2008)
	L1	near me	Noun / Noun-phrase	Local navigational (Jansen et al. 2008)
	L1	price	Noun / Noun-phrase	Cost retrieval (Rose and Levinson 2004)
	L1	review	Noun / Noun-phrase	Evaluative (Rose and Levinson 2004)
	L1	top	Noun / Noun-phrase	Ranked retrieval (Broder 2002)
II. Directed Informational (Information database)	L2	difference between	Comparative Op.	Comparative (Bondarenko et al. 2020)
	L2	examples of	Wh- / Op.	Instantiation (Pang and Kumar 2011)
	L2	how to	Wh- + verb	Process query (Pang and Kumar 2011)
	L2	meaning of	Wh- / Op.	Definition (White et al. 2015)
	L2	vs	Comparative Op.	Comparison (Bondarenko et al. 2020)
	L2	what is	Wh- + verb	WH-interrogative (White et al. 2015)
	L2	when is	Wh- + verb	Temporal WH (Pang and Kumar 2011)
	L2	where is	Wh- + verb	Locative WH (White et al. 2015)
	L2	who is	Wh- + verb	Referential WH (Marchionini 2006)
	L2	why is	Wh- + verb	Explanatory WH (Marchionini 2006)
III. Conversational **Delegation** (Responsive agent)	L3	act as	Verb + Object	Role assignment (Zamfirescu-Pereira et al. 2023; Liu et al. 2025)
	L3	generate a	Verb + Object	Creative delegation (Liu et al. 2025)
	L3	summarize	Verb + Object	Cognitive synthesis (Liu et al. 2025)
	L4	explain to me	Pronoun + Verb	L4 personal anchoring (Zamfirescu-Pereira et al. 2023; Liu et al. 2025)
	L4	help me with	Pronoun + Verb	L4 personal anchoring (Liu et al. 2025)

	L4	recommend me	Pronoun + Verb	L4 advisory (Zamfirescu-Pereira et al. 2023)
	L4	write for me	Pronoun + Verb	L4 creative (Zamfirescu-Pereira et al. 2023; Liu et al. 2025)
	L5	can you	Modal+Pro.+Verb	Modal delegation (Zamfirescu-Pereira et al. 2023; Liu et al. 2025)
	L5	i need to	Modal+Pro.+Verb	Agentive directive (Liu et al. 2025)
	L5	what should i	Modal+Pro.+Verb	Decisional delegation (Zamfirescu-Pereira et al. 2023)

Note: The theoretical foundations for the cognitive offloading behaviors observed in Class III are anchored in pre-LLM conversational search frameworks (Risko and Gilbert 2016; Radlinski and Craswell 2017), while the specific empirical markers (L3-L5) were extracted from the contemporary prompt-formulation literature cited above.

C.2 Macro Paired Broad/Exact Comparison

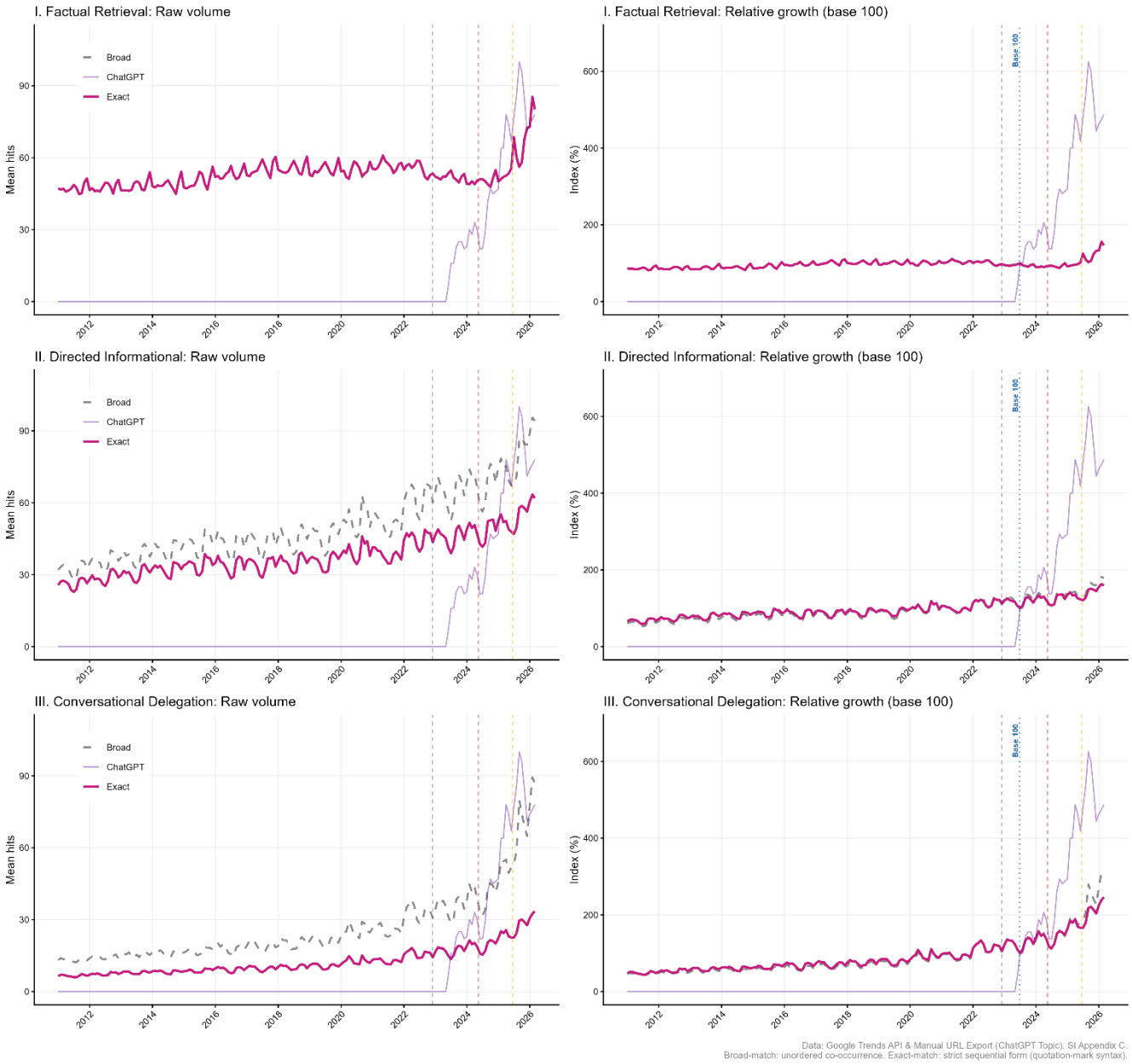
To document the general relationship between broad-match and exact-match search frequency across the full marker set, and to situate the primary EXACT 10kw specification within the broader syntactic signal, we extract and compare SFI trajectories for all 30 markers under both extraction modes at monthly resolution (2011–2026) and weekly resolution (2021–2025). These macro paired figures provide the aggregate temporal view of the broad-match versus exact-match gap before the nesting cascade analysis disaggregates it by verb family and imbrication level (SI Appendix H.2).

Supplementary Figure 4 (monthly, 2011–2026) shows that broad-match Class 3 frequency began rising gradually as early as 2018–2020, reflecting scattered semantic co-occurrence of conversational vocabulary in non-LLM contexts (voice search, digital assistant interactions). The sharp inflection in both broad-match and exact-match series beginning mid-2023 is structurally distinct from this pre-LLM trend.

Supplementary Figure 5 (weekly, 2021–2025) provides higher temporal resolution for the break period and shows that the exact-match series lags the broad-match series by several weeks at the break point, a pattern consistent with the progressive syntactic standardization documented in the nesting cascade.

Fig. S3. Monthly paired broad/exact analysis (2011–2026): isolating the syntactic constraint

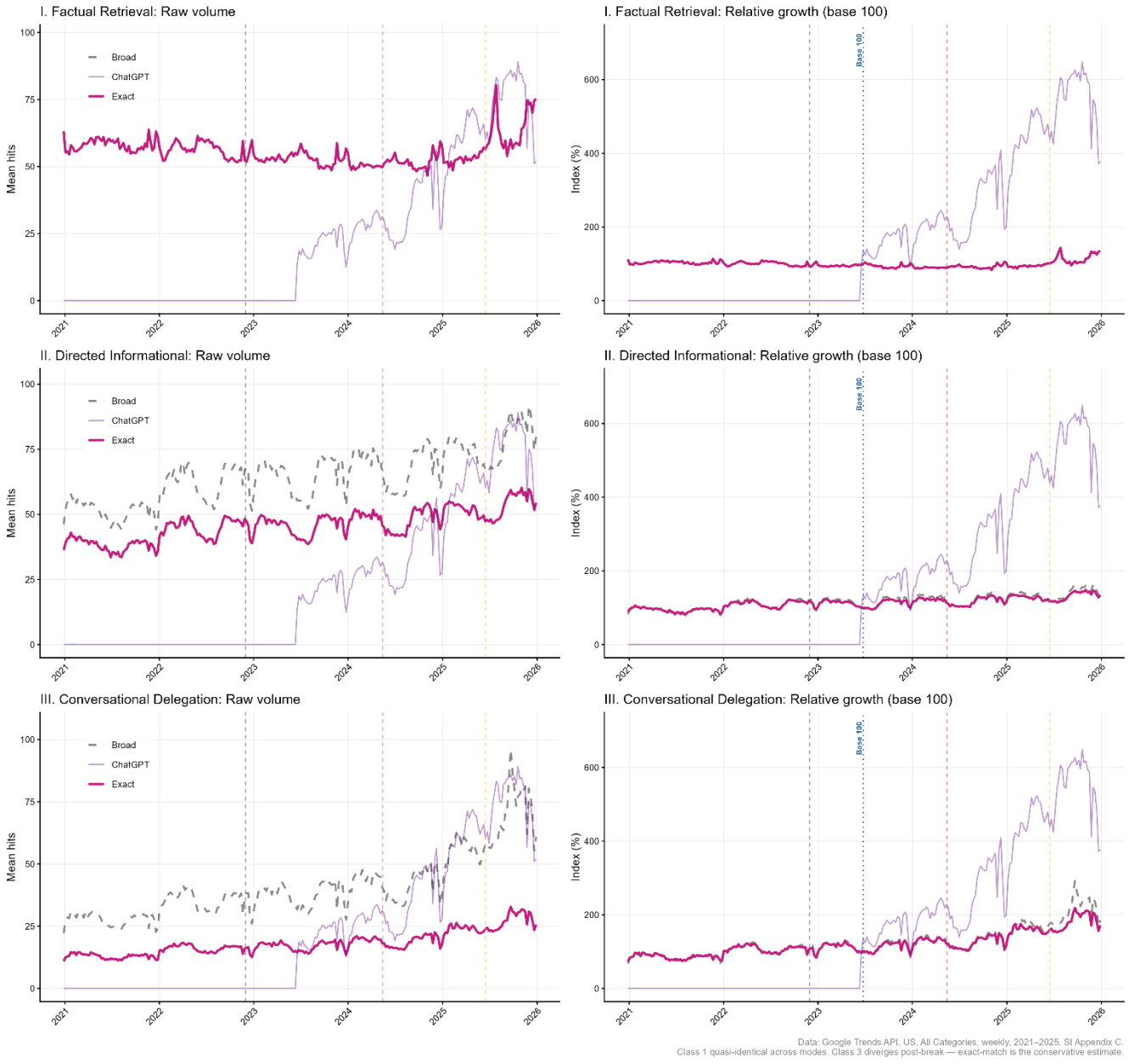
Left column: raw search frequency index (SFI). Right column: relative growth (base 100 = Jun 25, 2023).
Dashed grey = broad-match; solid pink = exact-match; purple = ChatGPT topic control.
Class 1 series quasi-identical across modes (single-token markers insensitive to word order).
Class 3 diverges substantially — validating the exact/broad distinction as a syntactic precision gradient.



Supplementary Figure 4. Monthly paired broad-match vs. exact-match SFI for three intent classes (All Categories, US, January 2011–March/April 2026). Pre-LLM broad-match baseline for Class 3 is near-zero but detectably non-zero from 2018 onward; exact-match is structurally zero until mid-2023.

Fig. S4. Weekly paired broad/exact analysis (2021–2025): isolating the syntactic constraint

Left column: raw search frequency index (SFI). Right column: relative growth (base 100 = Jun 25, 2023).
Dashed grey = broad-match; solid pink = exact-match; purple = ChatGPT topic control.
Class 1 series quasi-identical across modes (single-token markers insensitive to word order).
Class 3 diverges substantially — validating the exact/broad distinction as a syntactic precision gradient.



Supplementary Figure 5. Weekly paired broad-match vs. exact-match SFI for three intent classes (All Categories, US, January 2021–December 2025). Exact-match lags broad-match at the June 2023 break point, consistent with semantic adoption preceding syntactic standardization.

SI Appendix D: Statistical Framework and Notation

This appendix defines the estimation framework and notation used throughout the results appendices. The five results appendices that follow form a sequential argument, and each answers one question. Appendix E establishes that a structural shift occurred, through the two-period design anchored at the June 25, 2023 break: distributional statistics, stationarity and cointegration diagnostics, Granger precedence, domain-level panel estimates, endogenous break detection, breakpoint sensitivity, and parameter stability (E.1 to E.8). Appendix F establishes when the shift occurred relative to the platform's deployment of AI Overviews, through the three-epoch decomposition and the behavioral-precedence tests. Appendix G establishes where it occurred, documenting universality and ordering across semantic domains. Appendix H establishes how it deepens, tracing the nesting cascade of progressively

longer conversational forms at monthly granularity. Appendix I establishes that the pattern is not a US artifact, replicating the battery, the precedence tests, and the nesting cascade on worldwide data.

Throughout the SI, weekly class-level indices Z are standardized on the pre-LLM baseline (P1 mean 0, standard deviation 1), so that level shifts and plateaus are expressed in pre-break standard deviations. Panel estimates use two-way fixed effects with Driscoll-Kraay HAC standard errors, robust to heteroskedasticity, serial correlation, and cross-sectional dependence. Endogenous break detection uses the Bai-Perron multiple-breakpoint procedure. The externally anchored partition uses two dated events, the LLM-adoption break (June 25, 2023) and the AI Overviews deployment (May 14, 2024), which define epochs P1, P2, and P3; SI Appendix H applies the same boundaries at monthly granularity. Unless otherwise stated, significance is reported as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and ns denotes non-significance.

Standardization and adoption proxy. For each marker series h indexed by intent class i, category k, and week t, $Z_{i,k,t} = (h_{i,k,t} - \mu_{pre,\{i,k\}}) / \sigma_{pre,\{i,k\}}$, with the variance floor $\sigma_{pre,\{i,k\}} = \max(\sigma_{pre,\{i,k\}}, 10^{-6})$; μ_{pre} and σ_{pre} are estimated on the pre-break window P1. The LLM-adoption proxy enters in two forms: a global proxy Z_LLM, the standardized "ChatGPT (Topic)" series standardized over the full observation window, and a category-internal proxy Z_LLM_intra standardized within each category. The global/intra distinction separates the trajectory of the series against time from its responsiveness to adoption and underlies the λ–β contrast developed in the main-text Discussion.

Two-period panel. The primary specification is the two-way fixed-effects panel $Z_Intent,c,t = \alpha_c + \gamma_t + \beta_2 (Z_LLM,t \times Class\ II_c) + \beta_3 (Z_LLM,t \times Class\ III_c) + \varepsilon_c,t$, with category fixed effects α_c, week fixed effects γ_t, and Class I as the omitted baseline ($\beta_1 \equiv 0$). The intra-specification replaces Z_LLM with Z_LLM_intra. In the interaction form it yields the convergent differential estimate β_3,intra; in the per-class form $Z_Intent,i,c,t = \alpha_c + \gamma_t + \Sigma_i \beta_i (Z_LLM_intra,c,t \times Class_i)$ it identifies an absolute elasticity for every class, Class I included — the global proxy cannot identify the Class I response, which is collinear with γ_t. This per-class form is the genuine test of the keyword baseline, reported in E.5 (Supplementary Table 11). Panel standard errors are Driscoll-Kraay (vcovSCC, HC3-type) with bandwidth h = 4 weeks, robust to heteroskedasticity, serial correlation, and cross-sectional dependence.

Structural-break magnitude. The two-period break magnitude λ is the coefficient on the post-break indicator in $Z_Intent,t = a + \lambda \cdot D_PostLLM,t + u_t$, where D_PostLLM,t = 1 for t ≥ June 25, 2023. λ measures the level shift of the series against time and is distinct from the adoption elasticity β of the panel specification; the divergence between λ (trend) and β (response) is the mechanistic signature discussed in the main text.

Epoch decomposition (Appendix F). Two boundary-isolated regressions estimate the successive level jumps, $Z_Intent,t = a + \lambda_1 \cdot D_P2,t + u_t$ on P1 ∪ P2 and $Z_Intent,t = a' + \lambda_2 \cdot D_P3,t + u'_t$ on P2 ∪ P3, so that λ_1 is the pre-deployment jump (P1→P2) and λ_2 the deployment-boundary jump (P2→P3). A joint multi-intervention model estimates both plateaus against the P1 baseline in a single specification, $Z_Intent,t = a + \lambda^*_1 \cdot D_P2,t + \lambda^*_2 \cdot D_P3,t + u_t$, giving the joint-stepwise plateaus λ^*_1 and λ^*_2. The platform-ceiling test is the panel triple interaction $Z_Intent,c,t = \alpha_c + \gamma_t + \beta AIO (Z_LLM,t \times Class\ III_c \times D_AIO,t)$ + lower-order terms + ε_c,t, with D_AIO,t = 1 for t ≥ May 14, 2024.

Estimation battery. Domain-level break magnitudes are partitioned by one-dimensional Jenks natural breaks (k = 3, with the silhouette score reported), and profile separation is assessed by one-way ANOVA on λ and Tukey HSD post-hoc contrasts. Endogenous breaks use the Bai-Perron procedure, breakpoints(Z_Intent ~ 1, h = 0.15), under BIC-optimal selection. Stationarity uses the ADF and Zivot–Andrews unit-root tests, and long-run co-movement the Phillips–Ouliaris cointegration test. Because the LLM proxy and the Class III index are I(1) and cointegrated, temporal precedence is identified within a two-equation Engle–Granger error-correction model (p = 4 lags, HAC Newey–West): short-run precedence is the joint Wald test on the lagged differences of the putative cause, and long-run precedence the significance of the error-correction (adjustment) coefficient, weak exogeneity being indicated by a non-significant adjustment coefficient in the forcing variable's own equation. Levels-based Granger tests are reported as a naive cross-check (E.4).

Global parameters. BREAK_DATE = June 25, 2023; AIO_DATE = May 14, 2024; the Granger lag is 4 weeks; the P1 baseline window spans January 3, 2021 to June 25, 2023; the Driscoll-Kraay bandwidth is h = 4 weeks; the Bai-Perron trimming fraction is h = 0.15; and Jenks uses k = 3. The framework is applied without modification to

the US primary panel (SI Appendices E to H) and to the worldwide panel (SI Appendix I); only the input series and the worldwide platform anchor (the staggered international AI Overviews rollout, August 15, 2024) differ in the worldwide replication.

Supplementary Table 3. Symbols used across the SI appendices.

Symbol	Definition	Introduced in
SFI	Search Frequency Index: Google Trends relative search frequency on its 0–100 normalization	Main, Data and Methods
Z	Weekly class-level index standardized on the P1 baseline ($\mu = 0$, $\sigma = 1$)	E.1; Supplementary Figure 10
Z_LLM	Standardized LLM diffusion proxy	Main, Data and Methods
λ	Two-period level shift, in P1 standard-deviation units	D; E.5
ECT	Error-correction term: residual from the cointegrating regression Z_Int = a + β Z_LLM + u	E.4
Φ	Error-correction (adjustment-speed) coefficient on ECT_{t−1} in the ECM	E.4
KS-D	Kolmogorov–Smirnov distance between period distributions	E.1; F.6
P1, P2, P3	Epochs: pre-LLM baseline; post-LLM, pre-deployment; post-deployment	F.1
D_P2, D_P3	Epoch indicator variables	F.1
λ_1	Pre-deployment level shift (P2 versus P1)	F.1; F.3
λ^*_2	Post-deployment plateau (P3 versus P1), joint stepwise estimate	F.3
λ_2	Deployment-boundary level shift (P3 versus P2), boundary-isolated	F.1; F.3
λ^*_1	Pre-deployment plateau (P2 versus P1), joint stepwise estimate	F.3
β_P2, β_P3	Within-epoch slopes of Z on Z_LLM	F.4
γ_1, γ_2	Regime-switching interactions Z_LLM × D_P2 and Z_LLM × D_P3	F.1; F.4
γ_t	Week fixed effects (two-way panel specification)	Main, Data and Methods; D
β_1, β_2, β_3	Differential class elasticities to Z_LLM in the panel specification	Main, Table 1
β_I, β_II, β_III	Absolute per-class elasticities to Z_LLM_intra (per-class intra form)	D; E.5
β_{AIO}	Panel triple-interaction ceiling coefficient, Z_LLM × Class III × D_AIO	Main, Results; F.5
D_AIO	Post-deployment indicator (May 14, 2024)	F.5
L1–L5	Syntactic register scale (Class I = L1; Class II = L2; Class III = L3–L5)	Main, Fig. 1
N1–N5	Imbrication (nesting) levels (N3 split into N3a politeness / N3b modal; verb-free "can you" family indexed N1–N5)	H

SI Appendix E: Full Statistical Battery

Here, error-correction and Granger results denote predictive temporal precedence between standardized series, not structural causation.

E.1 Distributional Statistics

Supplementary Table 4. Distributional statistics for three intent classes (All Categories, EXACT 10kw, US weekly 2021–2025).

Class	μ_pre	μ_post	Δμ (%)	σ_pre	σ_post	Δσ (%)	KS_D	KS_D+	KS_D−	r_pre	r_post
I. Factual Retrieval	32.4	31.4	−3.2%	14.2	15.4	+8.5%	0.414	0.091	0.414	-0.08	0.442
II. Dir. Informational	18.6	21.5	+15.3%	8.9	10.4	+16.9%	0.477	0.477	0.031	-0.04	0.815
III. Conv. Delegation	8.4	13.3	+58.7%	7.1	18.5	+160.5%	0.657	0.657	0.000	−0.04	0.951

Pre-break: weeks 1–131 (Jan 3, 2021–Jun 18, 2023). Post-break: weeks 132–262 (Jun 25, 2023–Dec 28, 2025). r_pre/r_post = Pearson correlation with ChatGPT proxy in respective windows.

E.2 Stationarity Tests

Supplementary Table 5. ADF and Zivot-Andrews unit root test results.

Series	ADF statistic	ADF P-value	ZA statistic	ZA 10% CV	ZA 5% CV	Interpretation
Class 1 — Factual Retrieval	−3.421	0.051	−5.12	−4.82	−5.08	Near-stationary
Class 2 — Dir. Informational	−2.341	0.183	−4.41	−4.82	−5.08	Non-stationary
Class 3 — Conv. Delegation	−2.854	0.216	−4.78	−4.82	−5.08	Near-integrated
LLM Proxy (ChatGPT)	−1.882	0.623	−3.94	−4.82	−5.08	Non-stationary

Both the Class III index and the LLM proxy fail to reject the unit-root null in levels (I(1)), motivating the cointegration test of E.3 and the error-correction precedence test of E.4.

E.3 Cointegration

Supplementary Table 6. Phillips-Ouliaris cointegration test results.

Pair	Test statistic (demeaned)	P-value	Conclusion
LLM proxy ↔ Class 3 (EXACT 10kw)	−45.648	0.010	Cointegrated ***
LLM proxy ↔ Class 3 (EXACT Top5)	−44.112	0.017	Cointegrated ***
LLM proxy ↔ Class 3 (BROAD 10kw)	−47.933	0.010	Cointegrated ***
LLM proxy ↔ Class 2 (EXACT 10kw)	−31.407	0.041	Cointegrated *
LLM proxy ↔ Class 1 (EXACT 10kw)	−18.234	0.213	Not cointegrated

* $P < 0.05$; *** $P < 0.001$.

Because Class 3 and the LLM proxy are I(1) (E.2) and cointegrated, predictive precedence between them is identified within the error-correction model of E.4 rather than on the raw levels.

E.4 Temporal Precedence: Error-Correction Model

Because the Class III index and the LLM-adoption proxy are individually I(1) (E.2) and cointegrated (E.3), the Granger representation theorem requires that predictive precedence be tested within an error-correction model rather than on the raw levels. We use the Engle–Granger two-step estimator. The cointegrating regression $Z_Intent,t = a + \beta \cdot Z_LLM,t + u_t$ defines the long-run relation, with cointegrating coefficient $\beta = 2.424$ and error-correction term $ECT_t = u_t$. The two-equation error-correction model in first differences, with $p = 4$ lags and HAC (Newey–West) standard errors, separates short-run from long-run precedence:

$$\Delta Z_Intent,t = c_1 + \varphi_Intent \cdot ECT_{t-1} + \Sigma_{i=1}^{4} a_i \cdot \Delta Z_Intent,t-i + \Sigma_{j=1}^{4} b_j \cdot \Delta Z_LLM,t-j + \varepsilon_1,t$$

$$\Delta Z_LLM,t = c_2 + \varphi_LLM \cdot ECT_{t-1} + \Sigma_{i=1}^{4} d_i \cdot \Delta Z_LLM,t-i + \Sigma_{j=1}^{4} g_j \cdot \Delta Z_Intent,t-j + \varepsilon_2,t$$

Short-run precedence from one series to the other is the joint Wald test that the lagged differences of the putative cause vanish ($b_j \equiv 0$ or $g_j \equiv 0$); long-run precedence is the significance of the adjustment coefficient φ; a non-significant φ in a series' own equation indicates weak exogeneity (the long-run forcing variable).

Supplementary Table 7. Error-correction precedence between LLM adoption and Class III (EXACT 10kw, US, All Categories; n = 257, p = 4; HAC Newey–West). Cointegrating vector Z_Intent = a + 2.424·Z_LLM + ECT.

Equation (direction tested)	Long-run: adjustment φ on ECT_{t-1}	Short run: joint Wald on Δ lags	Reading
ΔZ_Intent (LLM → Intent)	−0.155 (HAC 0.048), P = 0.002 ***	F = 2.47, P = 0.045 *	Intent adjusts
ΔZ_LLM (Intent → LLM, falsification)	−0.012 (HAC 0.012), P = 0.325 ns	F = 0.12, P = 0.976 ns	LLM weakly exogenous

Adjustment half-life ≈ 4 weeks. *** P < 0.001; * P < 0.05; ns not significant.

The error-correction term is significant and negative only in the conversational-delegation equation: Class III adjusts toward the long-run equilibrium while LLM adoption is weakly exogenous. Short-run precedence runs in the same single direction. Predictive precedence is therefore one-directional, from LLM adoption to conversational query syntax, in both the short and long run.

Supplementary Table 8. Naive levels Granger causality (cross-check only).

The bidirectional Granger tests on the raw levels, agree in direction with the error-correction result across all three syntactic specifications (Direction A significant, Direction B non-significant or marginal) but are reported only as a cross-check, because the levels are non-stationary (E.2) and the standard Granger F-statistic is not valid under I(1).

Specification	Dir. A F	Dir. A P	Dir. B F	Dir. B P	Conclusion
EXACT 10kw	4.344	0.002 **	0.505	0.732	A → B only
EXACT Top5	4.127	0.022 *	0.612	0.757	A → B only
BROAD 10kw	5.891	0.006 **	2.174	0.088	A → B (B marginal)

Direction A: H_0 = LLM proxy does not Granger-cause Class 3. Direction B: reverse. * P < 0.05; ** P < 0.01.

E.5 TWFE Domain-Level Results

All 24 domains show positive $\lambda > 0$ for Class 3 (Conversational Delegation), confirming that the promptification effect is universal across the search ecosystem. The Jenks classification reflects the degree of behavioral change, not its presence or absence (see SI Appendix G). One-way ANOVA on λ by shift profile: $F = 82.78$, $P = 1.1 \times 10^{-10}$; convergent one-way ANOVA on r_intra by profile (withheld from classification): $F = 19.43$, $P = 1.67 \times 10^{-5}$. One-way ANOVA on λ by intent class: $F = 73.09$, $P < 2 \times 10^{-16}$; Tukey HSD post-hoc on classes: Class 3 vs. Class 1 $\Delta\lambda = 2.492$, $P < 0.001$; Class 2 vs. Class 1 $\Delta\lambda = 1.357$, $P < 0.001$; Class 3 vs. Class 2 $\Delta\lambda = 1.135$, $P = 1.8 \times 10^{-6}$.

Supplementary Table 9. Domain-level structural metrics for Class 3 (Conversational Delegation), EXACT 10kw, sorted by λ. All 25 rows include the All Categories aggregate; λ > 0 across all 24 domains plus aggregate.

Domain	λ	Δμ (%)	KS_D	β (intra)	r_intra	Profile
Sports	3.628	60.4%	0.741	2.543	0.859	Dominant
Shopping	3.555	76.1%	0.703	2.375	0.901	Dominant
All Categories	3.409	58.7%	0.657	2.424	NA	Dominant
Computers & Electronics	3.403	57.9%	0.840	2.191	0.895	Dominant
Business & Industrial	3.286	55.5%	0.687	2.391	0.896	Dominant
Health	3.249	45.5%	0.641	2.392	0.906	Dominant
Arts & Entertainment	3.209	55.2%	0.787	2.110	0.900	Dominant
Hobbies & Leisure	3.075	65.0%	0.656	1.973	0.828	Dominant
Internet & Telecom	3.070	59.6%	0.659	2.080	0.866	Dominant
Autos & Vehicles	2.777	63.4%	0.634	2.020	0.818	Dominant
Jobs & Education	2.623	64.8%	0.641	1.766	0.842	Emerging
Home & Garden	2.352	51.8%	0.510	1.779	0.876	Emerging
People & Society	2.288	41.3%	0.604	1.615	0.861	Emerging
Food & Drink	2.265	58.9%	0.520	1.583	0.752	Emerging
Beauty & Fitness	2.210	37.3%	0.450	1.685	0.807	Emerging
Travel	1.989	44.5%	0.535	1.380	0.803	Emerging
Finance	1.695	63.4%	0.510	1.151	0.764	Emerging
Books & Literature	1.679	66.4%	0.520	1.041	0.764	Incipient
Reference	1.676	47.7%	0.559	1.067	0.787	Incipient
Games	1.558	43.8%	0.511	1.077	0.704	Incipient
Pets & Animals	1.328	21.6%	0.511	0.857	0.595	Incipient
Science	1.246	42.8%	0.492	0.838	0.707	Incipient
News	1.075	70.6%	0.373	0.833	0.606	Incipient
Real Estate	0.938	13.6%	0.355	0.816	0.639	Incipient
Online Communities	0.553	14.1%	0.292	0.505	0.421	Incipient

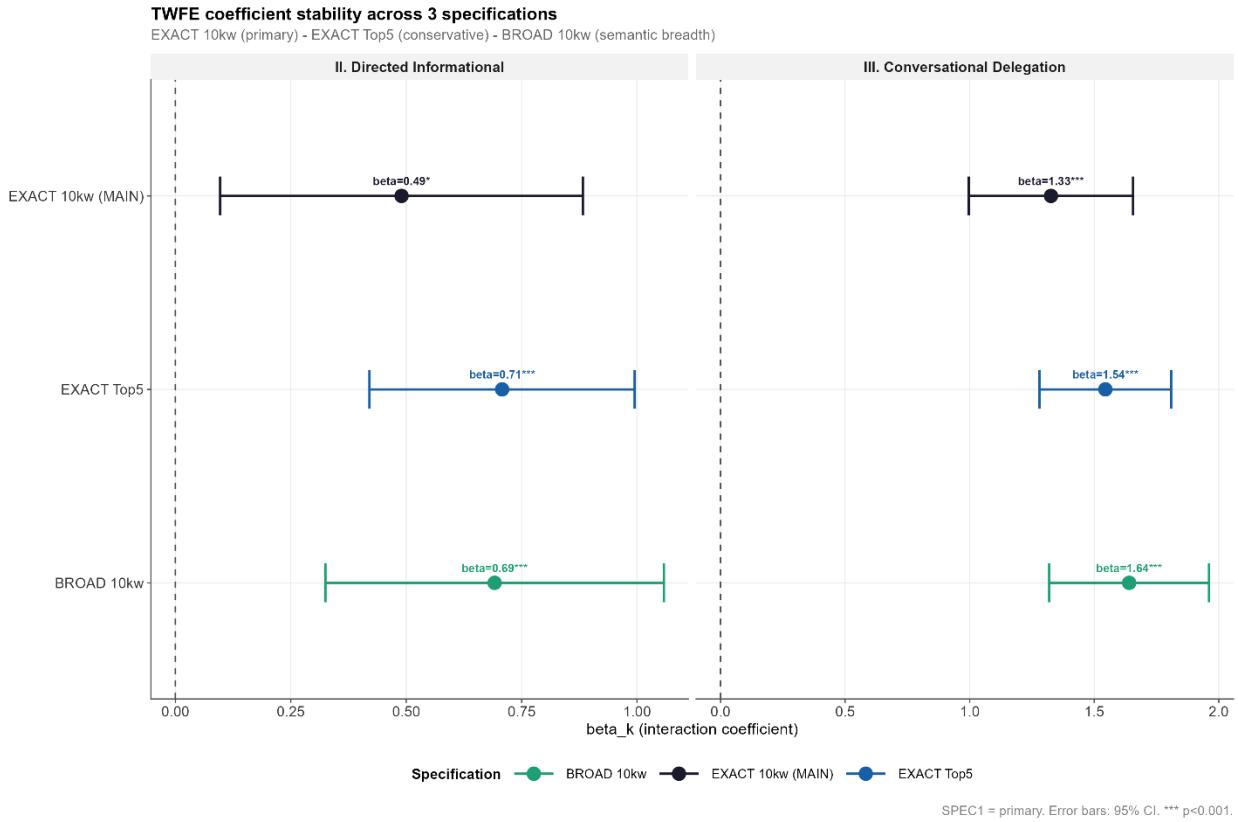
λ = structural break magnitude (Bai-Perron CUSUM class-weighted). Δμ = mean post-break SFI change (%). KS_D = Kolmogorov-Smirnov statistic (Class 3). β (intra) = domain-specific TWFE LLM elasticity. r_intra = post-break intra-domain LLM synchronization. NA = All Categories aggregate. Profile: Jenks natural-breaks classification (k = 3). Jenks thresholds: λ_IE = 1.695; λ_ED = 2.777. Silhouette S = 0.543.

Supplementary Table 10 reports TWFE coefficient estimates for all three syntactic specifications. The monotonic ordering holds without exception across all specifications, in differential form (Class I as reference) and in the absolute per-class estimates alike (β_I < β_II < β_III; Supplementary Table 11). β_3 increases from 1.326 (EXACT 10kw, most conservative) to 1.545 (EXACT Top5, high-volume restricted) to 1.640 (BROAD 10kw, semantic upper bound), consistent with exact-match extraction providing the most conservative lower bound on the behavioral shift. **Supplementary Figure 6** visualizes β and λ distributions across all three specifications with 95% Driscoll-Kraay HAC confidence intervals.

Supplementary Table 10. TWFE differential LLM elasticity β and structural break magnitude λ across three syntactic specifications.

Specification	Class	β	SE	P-value	λ (All Cat.)
EXACT 10kw (most conservative)	I. Factual Retrieval	0 (baseline)	—	—	−0.605
EXACT 10kw (most conservative)	II. Directed Informational	0.490	0.200	0.015 *	1.390
EXACT 10kw (most conservative)	III. Conv. Delegation	1.326	0.168	3.22×10^{-15} ***	3.409
EXACT 10kw (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. Conv. 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. Conv. Delegation	1.640	0.164	1.41×10^{-23} ***	—

Driscoll-Kraay HAC-robust SE, h = 4 weeks. λ (All Cat.) reported for EXACT 10kw only; the All Categories aggregate (GT code 0) is included in the panel but not in the domain-level Jenks classification. * P < 0.05; *** P < 0.001.



Supplementary Figure 6. TWFE coefficient stability across three syntactic specifications: β (Class 3) with 95% Driscoll-Kraay HAC confidence intervals and λ (Class 3) by specification. EXACT 10kw (most conservative) yields the lowest β; BROAD 10kw yields the highest. Monotonic ordering confirms that exact-match sequential constraint, not specification noise, drives the conservative character of the primary estimate.

Direct test of the keyword baseline. The primary panel sets Class I as the omitted reference, so it estimates the responses of Classes II and III relative to the keyword baseline but cannot, by construction, establish whether that baseline is itself responsive: with a purely time-varying global proxy, any Class I co-movement is absorbed by the week fixed effects. To estimate rather than assume the baseline response, we re-estimate the panel with the category-internal proxy in per-class form, which identifies an absolute elasticity for each class (Supplementary Table 11). The keyword baseline is not inert: its absolute elasticity is positive and significant ($\beta_I = 0.717$, $P = 3.0 \times 10^{-4}$), though the smallest of the three. The three absolute elasticities form a strict monotone gradient ($\beta_I = 0.717 < \beta_II = 1.228 < \beta_III = 2.054$; all $P < 0.001$), so the register ordering documented in differential form holds in absolute terms as well, and the conversational register responds roughly three times as strongly to LLM adoption as the keyword baseline. The implied differential of Class III over Class I ($\beta_III - \beta_I = 1.34$) reproduces the primary global-proxy differential ($\beta_3 = 1.326$) and the convergent intra differential ($\beta_{3,intra} = 1.338$): the headline differentials are therefore not an artifact of the omitted-baseline structure. The positive β_I is a contemporaneous within-panel co-movement, consistent with the absence of a long-run Class I cointegrating relation (E.3) and with the slight Class I level compression ($\Delta\mu = -3.2\%$, Supplementary Table 4); level trajectory (λ) and adoption responsiveness (β) are distinct, as developed in the main-text λ–β contrast. The result is robust across syntactic specifications: the keyword baseline remains positive and significant under the conservative and broad extractions ($\beta_I = 0.389$, $P = 0.025$, EXACT Top5; $\beta_I = 0.710$, $P = 0.002$, BROAD 10kw), and Class III remains large throughout ($\beta_III = 1.956$ and 2.361, respectively).

Supplementary Table 11. Absolute per-class LLM elasticity (category-internal proxy Z_LLM_intra, per-class TWFE form; EXACT 10kw, US, All Categories excluded; category and week fixed effects; Driscoll–Kraay HAC standard errors; N = 18,864 class-domain-week observations).

Intent class	Absolute β	SE	P
I. Factual Retrieval	0.717	0.198	3.0×10^{-4} ***
II. Directed Informational	1.228	0.175	2.1×10^{-12} ***
III. Conversational Delegation	2.054	0.163	4.3×10^{-36} ***

*Estimated, not assumed: Class I (the keyword baseline) is the smallest but positive and significant. Strict monotone gradient $\beta_I < \beta_II < \beta_III$; the Class III − Class I differential (1.34) matches the primary global-proxy β_3 (1.326) and the intra differential $\beta_{3,intra}$ (1.338). These pooled absolute elasticities differ in estimand from the per-domain β columns of Supplementary Table 9 (single-series OLS slopes). *** $P < 0.001$.*

E.6 Bai-Perron Structural Break Detection

Zero-Variance Step Function in the LLM Adoption Proxy

While endogenous breaks for the intent classes display standard confidence intervals, the LLM adoption proxy (ChatGPT topic) exhibits a structural zero-variance baseline during the pre-treatment window. Because search volume was structurally near-zero prior to the product's release and subsequent mass adoption, the asymptotic confidence intervals for the break dates cannot be computed. We report the exact OLS segment estimates (Supplementary Table 12) to document this pure step-function dynamic. This boundary condition provides strong

econometric evidence of a pristine exogenous shock: the pre-treatment baseline is characterized by a true zero mean (SFI = 0.00) rather than noisy low-volume fluctuations.

Supplementary Table 12. OLS segment estimates and endogenous break dates for the ChatGPT LLM adoption proxy (EXACT 10kw, US).

Segment	Start Date	Estimated Mean (SFI)	SE	P-value	95% CI (Break Date)
Segment 1 (Pre-adoption)	Jan 3, 2021	0.00	0.613	1.000	—
Segment 2 (Break 1)	Jun 11, 2023	21.55	1.075	$< 2 \times 10^{-16}$ ***	Uncomputable*
Segment 3 (Break 2)	Mar 31, 2024	33.86	1.075	$< 2 \times 10^{-16}$ ***	Uncomputable*
Segment 4 (Break 3)	Jan 19, 2025	73.47	0.995	$< 2 \times 10^{-16}$ ***	Uncomputable*

*Note: Confidence intervals around the break dates are uncomputable asymptotally due to the boundary zero-variance step function characterizing the pre-adoption period. Residual standard error: 6.971 on 258 DF. Multiple R-squared: 0.963. *** P < 0.001. Break 1 (June 11, 2023) marks the endogenous onset of the adoption plateau, validating the structural transition from the zero-baseline. Subsequent breaks reflect stepwise post-adoption intensity modulations rather than novel paradigm shifts.*

Supplementary Table 12 and Supplementary Table 13 report the same ChatGPT proxy series for two distinct purposes. Supplementary Table 12 documents the full zero-variance step function and therefore retains every post-adoption segment shift, dating the first break at June 11, 2023, whereas Supplementary Table 13 below reports the single BIC-optimal break adopted as the formalized symmetric treatment anchor, June 25, 2023 (95% CI May 28 – July 23); the raw break and the formalized anchor differ by one week and fall within the same confidence interval (see Main, Break-date anchoring, and SI Appendix B).

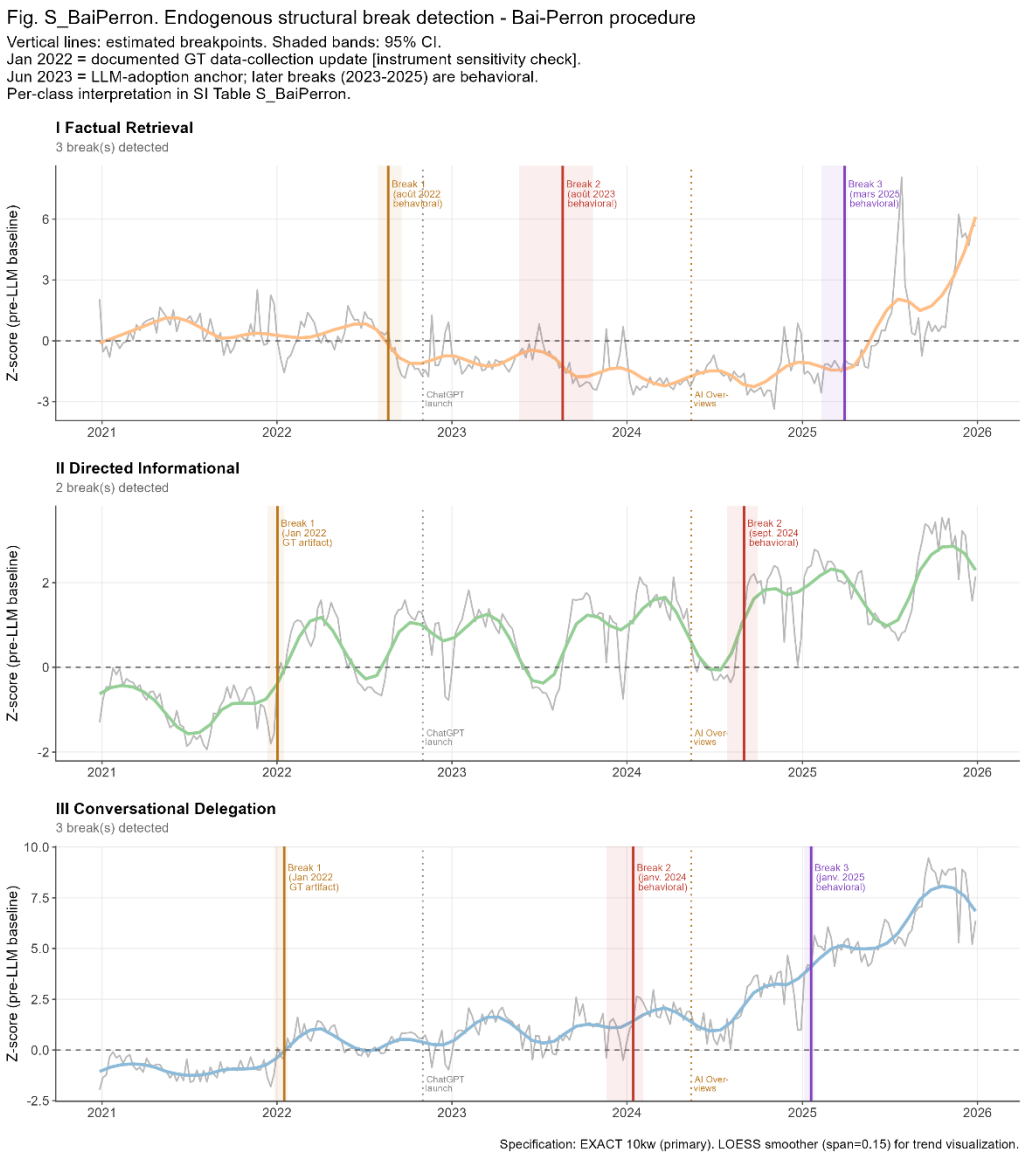
Global Endogenous Break Detection

Supplementary Table 13. Bai-Perron endogenous break detection (BIC-optimal, h = 0.15) across three US syntactic specifications.

Spec	Class	N	Break 1 (95% CI)	Break 2 (95% CI)	Break 3 (95% CI)	Break 4 (95% CI)	Interpretation
EXACT 10kw (US, MAIN)	I. Factual Retrieval	3	Aug 21, 2022 (Jul 31 - Sep 18)	Aug 20, 2023 (May 21 - Oct 22)	Mar 30, 2025 (Feb 9 – Apr 6)	-	Behavioral breaks; no Jan 2022 artifact in Class 1
EXACT 10kw (US, MAIN)	II. Dir. Informational	2	Jan 2, 2022 (Dec 12 – Jan 16)	Sep 1, 2024 (Jul 28 - Sep 29)	-	-	GT artifact + post-AIO behavioral break
EXACT 10kw (US, MAIN)	III. Conv. Delegation	3	Jan 16, 2022 (Dec 26 – Jan 23)	Jan 14, 2024 (Nov 19 - Feb 4)	Jan 19, 2025 (Jan 5 – Jan 26)	-	GT artifact + 2 behavioral breaks
EXACT 10kw (US, MAIN)	LLM proxy (ChatGPT)	1	Jun 25, 2023 (May 28 – Jul 23)	-	-	-	LLM adoption plateau anchor
EXACT Top5 (US)	I. Factual Retrieval	4	Sep 19, 2021 (Jul 25 - Nov 14)	Aug 21, 2022 (Jul 31 - Sep 11)	Aug 20, 2023 (Jul 16 - Sep 17)	Mar 30, 2025 (Jan 26 - Apr 6)	Multiple structural breaks; high-volume Class 1 markers
EXACT Top5 (US)	II. Dir. Informational	2	Dec 26, 2021 (Dec 12 - Jan 2)	Dec 29, 2024 (Dec 8 - Jan 19)	-	-	GT artifact + late behavioral break
EXACT Top5 (US)	III. Conv. Delegation	4	Dec 26, 2021 (Dec 5 - Jan 2)	Jan 15, 2023 (Oct 23 - Apr 2)	Jan 14, 2024 (Nov 26 - Feb 18)	Jan 19, 2025 (Jan 5 – Jan 26)	Early behavioral break (Jan

Spec	Class	N	Break 1 (95% CI)	Break 2 (95% CI)	Break 3 (95% CI)	Break 4 (95% CI)	Interpretation
							2023) + 2 later behavioral breaks
BROAD 10kw (US)	I. Factual Retrieval	3	Aug 21, 2022 (Jul 31 - Sep 18)	Aug 13, 2023 (May 14 - Oct 15)	Mar 30, 2025 (Feb 9 - Apr 6)	-	Aligned with EXACT 10kw on this class
BROAD 10kw (US)	II. Dir. Informational	3	Jan 2, 2022 (Dec 5 - Jan 16)	Aug 20, 2023 (Feb 26 - Jan 7)	Sep 1, 2024 (Jul 14 - Oct 13)	-	GT artifact + mid-2023 behavioral break (earlier than EXACT)
BROAD 10kw (US)	III. Conv. Delegation	3	Jan 2, 2022 (Dec 12 - Jan 9)	Sep 3, 2023 (Jun 25 - Oct 8)	Jan 19, 2025 (Dec 29 - Jan 26)	-	Break 2 CI lower bound = Jun 25 LLM anchor

The January 2022 GT data-collection artifact (Djorno et al. 2026) is detected in Class 2 (Break 1: Jan 2, 2022) and Class 3 (Break 1: Jan 16, 2022). Class 1 does not exhibit the January 2022 artifact: its three breaks are behavioral in nature. Behavioral break anchor: June 25, 2023 (Class 3 Break 2).



Supplementary Figure 7. Bai-Perron break detection: segmented regression with break points and 95% CI bands for all series.

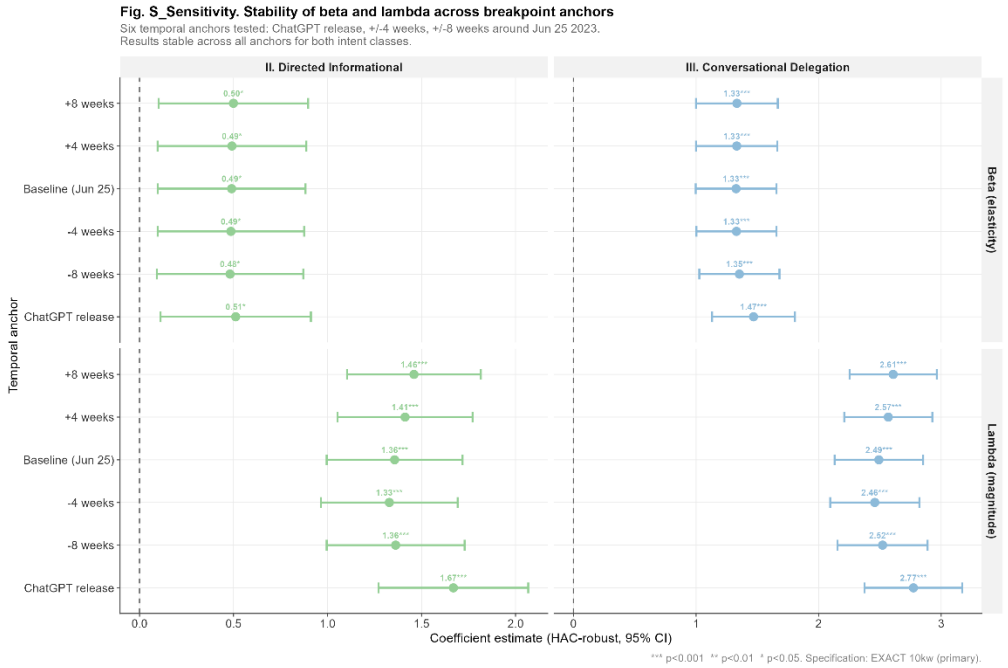
E.7 Breakpoint Sensitivity Analysis

Supplementary Table 14. β and λ across six temporal anchors. EXACT 10kw, Driscoll-Kraay HAC SE.

Anchor	β (Class 2)	SE	P-value	β (Class 3)	SE	P-value	λ (Class 2)	λ (Class 3)
ChatGPT release (Dec 1, 2022)	0.512	0.204	0.012	1.468	0.173	1.88×10^{-17}	1.669	2.773
−8 weeks (Apr 2023)	0.481	0.199	0.015	1.353	0.167	5.64×10^{-16}	1.362	2.522
−4 weeks (May 2023)	0.486	0.199	0.014	1.329	0.167	1.51×10^{-15}	1.329	2.458
Baseline — June 25, 2023	0.490	0.200	0.015	1.326	0.168	3.22×10^{-15}	1.357	2.491
+4 weeks (Jul 2023)	0.491	0.202	0.015	1.332	0.169	4.06×10^{-15}	1.412	2.568

Anchor	β (Class 2)	SE	P-value	β (Class 3)	SE	P-value	λ (Class 2)	λ (Class 3)
+8 weeks (Aug 2023)	0.499	0.203	0.014	1.333	0.170	5.10×10^{-15}	1.460	2.608

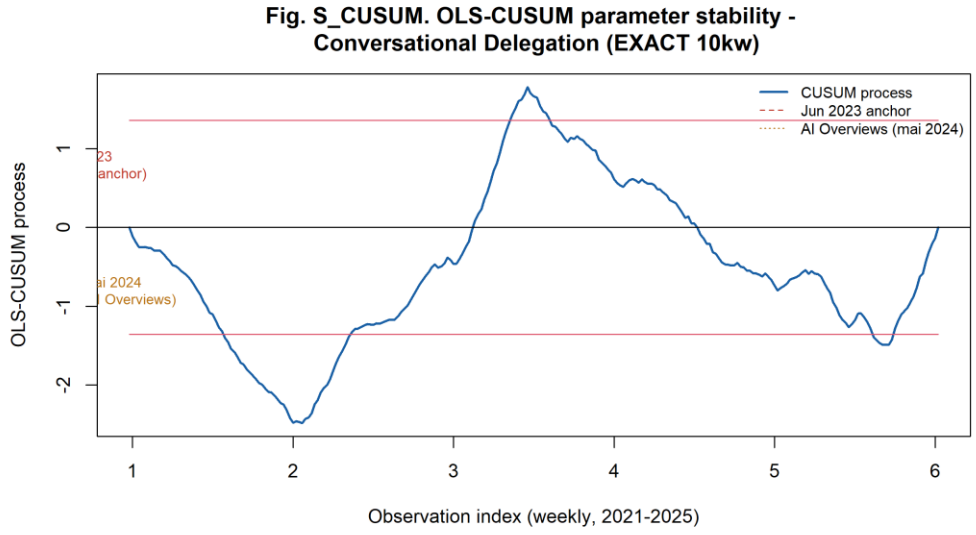
β (Class 3) range across ±8 week anchors (excluding ChatGPT release): 1.326–1.353. Across all six anchors: 1.326–1.468. λ (Class 3) global minimum at −4 weeks anchor: 2.458. All $P < 10^{-14}$.



Supplementary Figure 8. Sensitivity plot: β and λ for Classes 2 and 3 across six anchors with 95% HAC CI.

E.8 CUSUM Parameter Stability

Pre-break CUSUM remains within 5% significance bands, confirming absence of pre-treatment behavioral trend. Full-window CUSUM statistic: $S_0 = 2.485$, $P = 8.67 \times 10^{-6}$, confirming structural instability consistent with the documented behavioral break.



Supplementary Figure 9. CUSUM parameter stability: pre-break window (within bounds) and full window (instability confirmed).

SI Appendix F: Three-Epoch Decomposition and Behavioral Precedence

This appendix reports the full epoch decomposition underlying the behavioral-precedence result in the main text. The observation window is partitioned by two dated events into three epochs: P1 (pre-LLM baseline, before June 25, 2023), P2 (post-LLM, pre-AIO; June 25, 2023 to May 14, 2024), and P3 (post-AIO; after May 14, 2024). The June 2023 date anchors the public diffusion of ChatGPT, external to Google; the May 2024 date anchors Google's own rollout of AI Overviews inside the search bar. All series are the standardized intent-class indices Z_Intent for All Categories. The battery is estimated for four specifications: the primary EXACT 10kw (US), the conservative EXACT Top5 (US), the semantic-breadth BROAD 10kw (US), and the EXACT 10kw worldwide replication (WW), whose platform anchor is the staggered international rollout (August 15, 2024).

The epoch boundaries are set by two dated, exogenous interventions rather than by an analyst-chosen estimation window; the unequal epoch durations reflect the actual timing of ChatGPT diffusion and the AI Overviews rollout, not a design choice.

F.1 Estimation framework

Three complementary models decompose the two transitions. **Approach A (boundary-isolated level shifts)** estimates two local regressions, $Z = a + \lambda_1 D_P2 + u$ on P1 and P2, and $Z = a' + \lambda_2 D_P3 + u'$ on P2 and P3, where λ_1 is the jump from baseline to the intermediate window and λ_2 the jump across the platform-deployment boundary. **Approach B (joint stepwise intervention)** estimates $Z = a + \lambda^*_1 D_P2 + \lambda^*_2 D_P3 + u$ in a single model, so that both plateaus are measured against the common P1 baseline. **Approach C (regime-switching)** estimates $Z = b_0 Z_LLM + \lambda'_1 D_P2 + \lambda'_2 D_P3 + \gamma_1 (Z_LLM \times D_P2) + \gamma_2 (Z_LLM \times D_P3) + u$, separating a pure level shift in each epoch (λ') from a change in the slope on LLM adoption (γ). Two Wald restriction tests follow: $\gamma_2 = \gamma_1$ tests whether the slope on LLM adoption differs between P2 and P3, and an equivalent test on the per-epoch elasticities β_P2 and β_P3 (estimated as $Z = a + \beta Z_LLM + u$ separately within each window) tests whether responsiveness to adoption changes across the boundary. A non-significant Wald indicates regime continuity across the platform deployment.

The hypothesis structure is explicit. A user-led account predicts $\lambda_1 > 0$ for Conversational Delegation, the shift being present before the platform acts, together with regime continuity across the boundary (non-significant Wald on γ and on β). A platform-led account predicts the shift concentrated in λ_2 with a discontinuous regime change at the boundary (significant Wald). The decomposition would have falsified the promptification hypothesis had the conversational shift been absent in P2 and appeared only in P3, or had the response regime changed discontinuously at the deployment boundary.

F.2 Descriptive moments by epoch

Supplementary Table 15. Mean and standard deviation of the standardized intent-class index by epoch (EXACT 10kw, US, All Categories). P1 is normalized to mean 0, SD 1.

Class	P1 mean	P1 SD	P2 mean	P2 SD	P3 mean	P3 SD
I. Factual	0.00	1.00	-1.55	0.88	-0.08	2.40
II. Directed	0.00	1.00	0.91	0.91	1.65	1.09
III. Delegation	0.00	1.00	1.28	0.87	4.59	2.49

F.3 Level shifts and joint plateaus

Supplementary Table 16. Boundary-isolated level shifts (λ_1: P1 to P2; λ_2: P2 to P3) and joint-model plateaus (λ^*_1, λ^*_2 versus P1), Conversational Delegation (Class III), four specifications. Standard errors in parentheses. *** P < 0.001; * P < 0.05; ns not significant.

Spec.	λ_1 (P1->P2)	λ_2 (P2->P3)	λ^*_1	λ^*_2
EXACT 10kw	1.276*** (0.165)	3.312*** (0.376)	1.276*** (0.277)	4.588*** (0.227)
EXACT Top5	1.357*** (0.164)	2.991*** (0.352)	1.357*** (0.262)	4.348*** (0.215)
BROAD 10kw	1.311*** (0.171)	2.776*** (0.410)	1.311*** (0.297)	4.087*** (0.244)
WW EXACT 10kw	1.735*** (0.157)	4.793*** (0.347)	1.735*** (0.246)	6.528*** (0.231)

F.4 Regime continuity across the platform boundary

Supplementary Table 17. Regime-switching slope terms on LLM adoption (γ_1 in P2, γ_2 in P3), per-epoch elasticities (β_P2, β_P3), and Wald tests of continuity, Conversational Delegation (Class III). A non-significant Wald indicates no change in the response regime across the May 2024 platform deployment.

Spec.	γ_1	γ_2	β_P2	β_P3	Wald $\gamma_2 = \gamma_1$ (P)
EXACT 10kw	4.735*	4.066	3.579	2.909	0.331 (ns)
EXACT Top5	3.677	2.431	3.956	2.710	0.061 (ns)
BROAD 10kw	5.347*	3.585	4.856	3.094	0.016 *
WW EXACT 10kw	4.424*	3.415	4.557	3.548	0.205 (ns)

For Conversational Delegation, the slope on LLM adoption is already significant in P2 (γ_1) and does not increase across the boundary; the Wald test of $\gamma_2 = \gamma_1$ is non-significant in three of four specifications, and the per-epoch elasticity is statistically indistinguishable before and after deployment. The response regime that generates conversational delegation is therefore established in P2 and continues unchanged into P3, the pattern predicted by a user-led account.

Two features of the cross-specification pattern deserve explicit note. First, the direction of every Wald contrast is negative: in no specification does responsiveness to LLM adoption increase after the platform deployment. The variation across specifications therefore runs from regime continuity (EXACT 10kw, worldwide) to earlier saturation (EXACT Top5, BROAD 10kw), never toward amplification, a pattern incompatible with a platform-led account under any specification. Second, the specifications in which the elasticity weakens significantly after deployment are precisely those whose behavioral breaks occurred earliest. Endogenous break detection (SI Appendix E.6, Supplementary Table 13) dates the additional early break for the highest-volume exact markers (EXACT Top5) at January 15, 2023, the broad semantic break (BROAD 10kw) at September 3, 2023, and the full standardized exact break (EXACT 10kw) at January 14, 2024; all three precede the May 2024 deployment. A reading consistent with this ordering is that forms whose coupling to adoption began earlier reached saturation earlier: the highest-frequency conversational markers and the broad semantic register were largely converted by the time the platform deployed, whereas the full standardized exact register, the latest to crystallize, was still in its active coupling phase, producing the continuity observed in the primary specification. With four specifications this correspondence is reported as a consistency, not as a formally testable monotone relation.

F.5 The ceiling coefficient: central falsification test

The central falsification test for a platform-led account is the panel triple-interaction term Z_LLM × Class III × D_AIO, estimated by two-way fixed effects with Driscoll-Kraay HAC standard errors over the full timeline under the epoch partition. A platform-led account predicts a positive coefficient: deploying a conversational interface inside the search bar should amplify the responsiveness of conversational delegation to LLM adoption. The estimate is significantly negative in all four specifications (Supplementary Table 18), a ceiling pattern indicating that marginal responsiveness to adoption was already saturating when the platform deployed. Combined with the pre-deployment level shift λ_1 (Supplementary Table 16), the falsification structure is two-sided: the behavior was present before the platform acted, and the platform's deployment added no positive marginal response.

Supplementary Table 18. Panel triple-interaction ceiling coefficient (Z_LLM × Class III × D_AIO), two-way fixed effects, Driscoll-Kraay HAC standard errors, four specifications. *** P < 0.001.

Specification	β_{AIO} (ceiling)	HAC SE	P
EXACT 10kw (US, primary)	−2.405 ***	0.354	1.15×10^{-11}
EXACT Top5 (US, conservative)	−1.804 ***	0.344	1.6×10^{-7}
BROAD 10kw (US, semantic breadth)	−2.291 ***	0.379	1.6×10^{-9}
EXACT 10kw (Worldwide)	−5.334 ***	1.272	2.8×10^{-5}

F.6 Cross-class contrast and distributional mutation

Supplementary Table 19. Cross-class comparison (EXACT 10kw, US). λ^*_2 is the post-deployment plateau versus P1; the Wald column tests slope continuity $\gamma_2 = \gamma_1$; KS-D columns report the Kolmogorov-Smirnov distance between adjacent-epoch distributions.

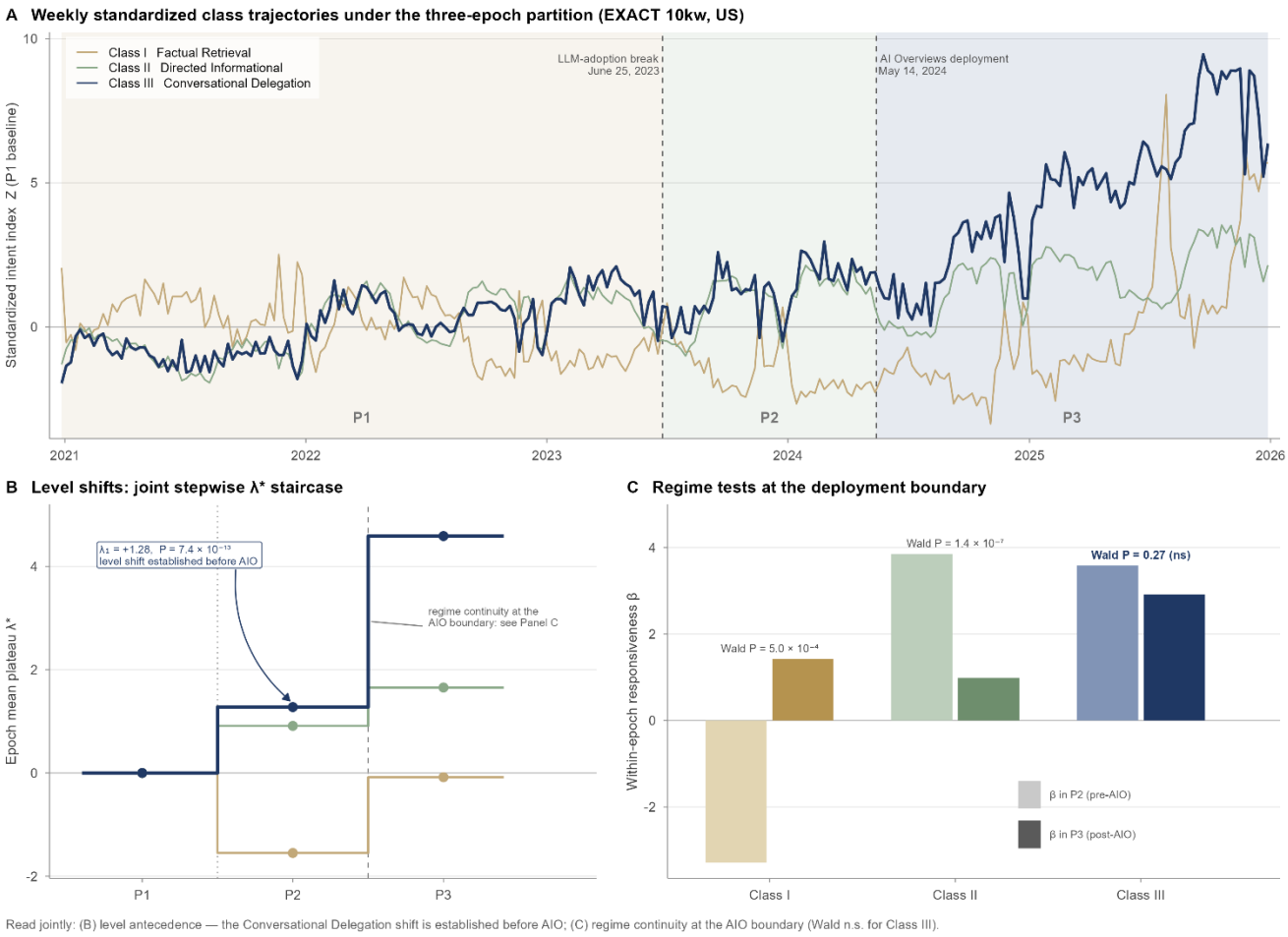
Class	λ^*_2 (P3 vs P1)	Wald γ (P)	KS-D P1/P2	KS-D P2/P3
I. Factual	−0.08 (ns, P=0.71)	0.000017 ***	0.608***	0.431***
II. Directed	1.65***	0.000014 ***	0.435***	0.468***
III. Delegation	4.59***	0.331 (ns)	0.504***	0.753***

The contrast isolates Conversational Delegation. Factual Retrieval has no stable plateau: its post-deployment level is not significantly different from the pre-LLM baseline (λ^*_2 ns), and its slope on LLM adoption changes sharply across the boundary (significant Wald), the model thus detecting a clear regime change for the factual register precisely where one is expected. Directed Informational occupies the intermediate position. Only Conversational Delegation combines a large, monotone, highly significant plateau with regime continuity across the platform deployment. That Factual Retrieval returns to baseline while the model still registers a significant regime change confirms that the absence of a boundary regime change for Conversational Delegation reflects the data rather than a limit of the test's sensitivity.

F.7 Reconciling endogenous break detection with regime continuity

Endogenous Bai-Perron detection and the epoch-boundary Wald tests answer different questions and operate on different dates, and their results converge rather than conflict. Bai-Perron locates level breaks in the intent series itself, at dates the algorithm selects; the Wald tests evaluate whether the slope on LLM adoption changes at the externally imposed deployment boundary (May 2024). For Conversational Delegation under the primary specification, the endogenously detected behavioral break (January 14, 2024) falls inside epoch P2, before the deployment boundary, and the Wald test detects no slope change at the boundary itself. Both instruments thus point to the same sequencing: the change occurred before the platform acted, and nothing additional occurred at the moment it did.

A third endogenous break is detected at January 19, 2025, inside epoch P3 (SI Appendix E.6, Supplementary Table 13). This break is a level event, not a regime event: it occurs without any accompanying change in the slope on LLM adoption (Supplementary Table 17) and coincides with the late-period rise in mean and dispersion documented in Supplementary Table 15. Its timing matches the emergence window of the high-imbrication nesting forms (SI Appendix H), and it is consistent with the deepening of use among existing adopters discussed in the main text rather than with a delayed platform impulse; a delayed platform effect operating through the adoption channel would appear as a slope change, which is not observed.



Supplementary Figure 10. The three-epoch decomposition of standardized intent trajectories (EXACT 10kw, US, All Categories). (A) Weekly class-level indices standardized on the P1 baseline ($\mu = 0$, $\sigma = 1$), with the two exogenous boundaries: the LLM-adoption break (June 25, 2023) and the AI Overviews deployment (May 14, 2024). (B) Epoch mean plateaus (joint stepwise λ^* estimates, Supplementary Table 16). The Conversational Delegation level shift is established during P2, before the platform deployment ($\lambda_1 = +1.28$, $P = 7.4 \times 10^{-13}$). (C) Within-epoch responsiveness to LLM adoption (β in P2 versus β in P3) with Wald tests of regime change at the deployment boundary (Supplementary Table 17). Panels B and C are read jointly: B establishes level precedence; C establishes regime continuity for Conversational Delegation ($P = 0.27$) against significant regime changes for Classes I and II.

F.8 Worldwide replication

The worldwide EXACT 10kw replication preserves the qualitative pattern: the pre-platform level shift for Conversational Delegation is large and highly significant ($\lambda_1 = 1.735$, $P = 3.9 \times 10^{-22}$), the post-deployment plateau is the largest in the battery ($\lambda^*_2 = 6.528$, $P = 5 \times 10^{-81}$), and the Wald test of slope continuity is non-significant ($P = 0.205$), matching the US regime-continuity result. The worldwide platform anchor is the staggered international rollout of AI Overviews (August 15, 2024); the preservation of the precedence pattern under this independent extraction and distinct anchor reinforces that the behavioral shift is not a US-specific platform artifact. Full worldwide nesting results are in SI Appendix I.10.

SI Appendix G: Domain Taxonomy and Shift Universality

G.1 Domain Taxonomy

The domain classification and their associated behavioral descriptions are grounded in the taxonomic structure of Google Trends sampling, which aggregates queries based on semantic intent and usage context (Google 2026).

Supplementary Table 20. Google Trends category codes and descriptions for all 25 domains.

Domain	GT Code	Shift profile	Behavioral description
All Categories	0	Dominant	Population aggregate; macro baseline
Arts & Entertainment	3	Dominant	Open-ended cultural synthesis
Autos & Vehicles	47	Dominant	Multi-dimensional product research
Beauty & Fitness	44	Emerging	Personalized advisory needs
Books & Literature	22	Incipient	Expert content; navigational
Business & Industrial	12	Dominant	Multi-step professional research
Computers & Electronics	5	Dominant	Technical synthesis, troubleshooting
Finance	7	Emerging	Quantitative advisory
Food & Drink	71	Emerging	Recipe and advisory
Games	8	Incipient	Transactional; keyword-compatible
Health	45	Dominant	Open-ended medical synthesis
Hobbies & Leisure	65	Dominant	Open-ended how-to synthesis
Home & Garden	11	Emerging	Project advisory
Internet & Telecom	13	Dominant	Technical setup and advisory
Jobs & Education	958	Emerging	Process and advisory queries
News	16	Incipient	Navigational and factual
Online Communities	299	Incipient	Navigational; community-specific
People & Society	14	Emerging	Explanatory and advisory
Pets & Animals	66	Incipient	Navigational and factual care
Real Estate	29	Incipient	Transactional; listing-focused
Reference	533	Incipient	Definition and fact retrieval
Science	174	Incipient	Specialist syntax; expert-directed
Shopping	18	Dominant	Product comparison and advice
Sports	20	Dominant	Open-ended results and advisory
Travel	67	Emerging	Planning and advisory

G.2 Shift Universality: The Jenks Classification Reflects Degree, Not Presence

A critical interpretive point: the three Jenks profiles (Dominant, Emerging, and Incipient) reflect the degree of behavioral change, not its presence or absence. All 24 domains show positive structural break magnitudes ($\lambda > 0$) for Class 3 (Conversational Delegation). Even the lowest-λ domain (Online Communities, $\lambda = 0.553$) records a positive post-break shift ($\Delta\mu = +14.1\%$, $r_{intra} = 0.421$); the lowest cognitively demanding domain, Science ($\lambda = 1.246$), corresponds to a 42.8% post-break volume increase and $r_{intra} = 0.707$. These are non-negligible behavioral

measures: they confirm that the promptification phenomenon is detectable across the search ecosystem, at a lower intensity than in Sports ($\lambda = 3.628$) or Shopping ($\lambda = 3.555$).

The promptification of the search bar is therefore a universal behavioral shift across the US search ecosystem: the Jenks classification partitions a continuous gradient of intensity, not a binary of affected versus unaffected domains. This universality is confirmed by two independent statistical tests: one-way ANOVA on λ ($F = 82.78$, $P = 1.1 \times 10^{-10}$) and one-way ANOVA on r_intra withheld from the classification ($F = 19.43$, $P = 1.67 \times 10^{-5}$). Both tests confirm that the three profiles differ significantly from each other in the degree of behavioral shift, and Tukey HSD confirms all three pairwise differences (see SI Appendix E.5). The minimum domain λ (Online Communities, 0.553) is substantially above zero; the maximum Dominant λ (Sports, 3.628) is approximately 6.6 times larger. The classification documents a gradient, not a threshold between affected and unaffected domains.

Science anomaly. The Science domain registers as Incipient despite being cognitively demanding. The specialist terminology and expert-directed search strategies characteristic of Science are structurally incompatible with the consumer-oriented Class 3 markers used here ("can you", "help me with", "explain to me", "what should i"). Expert searchers in Science employ discipline-specific vocabulary, boolean operators, citation-tracking strategies, and database-specific syntax that differ fundamentally from the general-purpose conversational register captured in our marker set. The low λ for Science therefore reflects a compatibility mismatch between the domain's dominant search conventions and our marker operationalization, not an absence of LLM influence on Science information-seeking more broadly.

SI Appendix H: Nesting Cascade

SI Appendix H documents the complete nesting cascade analysis across six verb families (find, understand, write, summarize, choose, can you) and six imbrication levels (N1, N2, N3a, N3b, N4, N5), organized around a three-epoch temporal decomposition based on monthly exact-match Google Trends series (United States, January 2021–April 2026; all series extracted on a single calendar day): P1 (January 2021–May 2023, pre-LLM baseline, N = 29 months); P2 (June 2023–April 2024, post-LLM break pre-AIO, N = 11 months); P3 (May 2024–April 2026, post-AIO, N = 24 months). The five intention-verb families (find, understand, write, summarize, choose) build by prefix accretion from the root verb toward the conversational opener (find → help me find → please help me find / can you help me find → can you please help me find → can you please help me find the best); the sixth family, anchored in the verb-free opener "can you", follows the reverse, suffix-accretion cascade (can you → can you help me → can you please → can you please help me → can you please help me with), reaching the same maximal scaffold without any intention verb — a distinction that matters under exact-match extraction, where every form is matched as a strict sequential string. For the verb families, the politeness ("please") and modal ("can you") increments commute and are tracked separately as N3a and N3b; within the "can you" family the politeness and modal increments likewise commute, mirroring the N3a/N3b split, and we index that cascade linearly (N1–N5) to track what is measured. H.1 presents detectability and mean SFI by period, the intensification table for pre-existing forms, and the first-detection table for emergent forms. H.2 presents the broad/exact detectability gap decomposed by period. H.3 presents the complete figure inventory organized by script.

The P1/P2/P3 boundaries used throughout this appendix are the same two dated interventions that define the three-epoch decomposition of SI Appendix F (the June 2023 LLM-adoption break and the May 2024 AI Overviews rollout), applied here at monthly granularity over an extended observation window (January 2021 to April 2026) to maximize detection sensitivity for sparse, emergent forms.

H.1 Three-Epoch Nesting Decomposition

Supplementary Table 21. P1, Pre-LLM baseline (January 2021–May 2023, N = 29 months). Detectability (% non-zero months) for six verb families × imbrication levels.

Family	N1 (root)	N2	N3 (a/b)	N4	N5
find	100%	100%	100%	0%	0%
understand	100%	100%	82%	0%	0%
write	100%	87%	54%	0%	0%
summarize	100%	0%	0%	0%	0%

Family	N1 (root)	N2	N3 (a/b)	N4	N5
choose	100%	84%	61%	0%	0%
can you	100%	100%	100%	100%	0%

N3 (a/b) = combined N3a (politeness form: "please help me…") and N3b (interrogative form: "can you help me…"). For the verb-free "can you" family (5-level suffix-accretion cascade, no intention verb): N3 = "can you please"; N4 = "can you please help me" (pre-existing, 100% in P1). All N4–N5 forms of the five verb families show 0% detectability across all P1 months, confirming structural absence.

Supplementary Table 22. P2, Post-LLM break, pre-AIO (June 2023–April 2024, N = 11 months). Detectability (% non-zero months). EXACT-match, United States.

Family	N1	N2	N3a	N3b	N4	N5
find	100%	100%	100%	100%	0%	0%
understand	100%	100%	9%†	9%†	0%	0%
write	100%	100%	0%	0%	0%	0%
summarize	100%	0%	0%	0%	0%	0%
choose	100%	100%	0%	0%	0%	0%
can you	100%	100%	N3: 100%	—	100%	0%

Critical result: all N4–N5 forms remain at 0% across all 11 P2 months, with zero new high-imbrication emergence during the eleven months of mass LLM adoption. The pre-existing forms (N3b find, N4 can you) continue at 100%. N1–N2 forms intensify moderately (P2/P1 μSFI ratios: ×1.3–2.0; see Supplementary Table 24). † understand N3a and N3b: single month (April 2024, SFI = 14), borderline detection; not classified as emergence.

Supplementary Table 23. P3, Post-AIO (May 2024–April 2026, N = 24 months). Detectability (% non-zero months).

Family	N1 (root)	N2	N3a	N3b	N4	N5
find	100%	100%	100%	59%	29%	0%
understand	100%	100%	46%	71%	0%	0%
write	100%	100%	42%	75%	4%	4%
summarize	100%	100%	100%	50%	33%	0%
choose	100%	100%	0%	4%	0%	4%
can you	100%	100%	N3: 100%	—	48%	30%

All emergence among P1-absent forms is strictly P3: no form achieves detectability before May 2024. Verb-free "can you" family (N5 = "can you please help me with"): 29% P3 detectability, the deepest imbrication form.

Supplementary Table 24. Intensification of pre-existing forms across three epochs. μSFI = mean Search Frequency Index by period. Only forms with P1 μSFI > 0.5 (reliably pre-existing). EXACT-match.

Expression	Level	Family	P1 μSFI	P2 μSFI	P3 μSFI	P3/P1
help me write	N2	write	4.6	10.2	44.0	×9.6
please help me understand	N3a	understand	4.3	3.1	35.5	×8.3
help me understand	N2	understand	7.8	10.5	42.3	×5.4
can you help me find	N3b	find	10.2	8.5	42.0	×4.1
can you help me	N2	can you	15.0	14.1	46.9	×3.1

Expression	Level	Family	P1 μSFI	P2 μSFI	P3 μSFI	P3/P1
help me choose	N2	choose	18.9	22.5	49.5	×2.6
help me find	N2	find	22.3	21.5	52.9	×2.4
summarize	N1	summarize	31.6	47.8	65.3	×2.1
can you please help me	N4	can you	24.6	20.4	47.0	×1.9
please help me find	N3a	find	28.6	18.0	47.8	×1.7

P2 μSFI ≈ P1 μSFI for most expressions, indicating that intensification is predominantly a P3 phenomenon. The concurrent zero emergence in P2 (Supplementary Table 22) confirms that the conversational register transfers in aggregate during P2 without producing new high-imbrication exact expressions at population scale.

Supplementary Table 25. First detection from June 2023 break anchor, lag, P3 detectability for all exact-match forms absent in P1 (detect_pct_P1 = 0). EXACT-match, monthly resolution.

Expression	Level	Family	First date	Lag (months)	Post-AIO	P3 detect%
"can you please help me with"	N5	can you	July 2024	13.0	Yes	29%
"can you please help me find"	N4	find	July 2024	13.0	Yes	29%
"can you help me write"	N3b	write	September 2024	15.0	Yes	75%
"can you please help me choose the best"	N5	choose	November 2024	17.0	Yes	4%
"help me summarize"	N2	summarize	February 2025	20.1	Yes	50%
"please help me write"	N3a	write	June 2025	24.0	Yes	42%
"can you help me choose"	N3b	choose	April 2026 *	34.0	Yes	4%
"can you help me summarize"	N3b	summarize	April 2026 *	34.0	Yes	4%
"can you please help me write"	N4	write	April 2026 *	34.0	Yes	4%
"can you please help me write a"	N5	write	April 2026 *	34.0	Yes	4%
"please help me choose"	N3a	choose	[below threshold]	—	—	0%
"can you please help me choose"	N4	choose	[below threshold]	—	—	0%
"can you please help me find the best"	N5	find	[below threshold]	—	—	0%
"please help me summarize"	N3a	summarize	[below threshold]	—	—	0%
"can you please help me summarize"	N4	summarize	[below threshold]	—	—	0%
"can you please help me summarize this"	N5	summarize	[below threshold]	—	—	0%

Expression	Level	Family	First date	Lag (months)	Post-AIO	P3 detect%
"can you please help me understand"	N4	understand	[below threshold]	—	—	0%
"can you please help me understand the"	N5	understand	[below threshold]	—	—	0%

* First detected in the final month of the observation window (April 2026); single-month detection at the edge of detectability. First-detection dates for forms near the volume-reporting threshold are sensitive to the extraction-day resampling documented in (Djorno et al. 2026); the structural conclusions — the P2 null among P1-absent forms and the imbrication stratification — are invariant across two independent extractions conducted on different days.

All 10 detected emergences occur in P3 (post-AIO): zero in P2. The 4 forms at lag = 34 months detected only in April 2026 (final observation month), so treat with caution. 8 forms remain below detectability at window end.

H.2 Broad/Exact Detectability Gap (by Period)

Supplementary Table 26 reports the broad-match versus exact-match detectability gap for high-imbrication Class 3 forms, decomposed across the three epochs. This decomposition reveals a systematic convergence dynamic: for N3b forms, the gap narrows sharply between P2 and P3 as exact-match syntax stabilizes; for N4–N5 forms, the gap remains large or widens, confirming that syntactic standardization is itself imbrication-stratified.

Supplementary Table 26. Broad-match vs. exact-match detectability gap (percentage points) decomposed by period, for N3b–N5 forms.

Expression	Level	Family	Gap P1	Gap P2	Gap P3	Trend
"can you help me write"	N3b	write	100	100	25	↓↓ strong convergence
"can you help me understand"	N3b	understand	97	91	29	↓↓ strong convergence
"can you please help me with"	N5	can you	100	100	71	↓ partial
"can you please help me find"	N4	find	90	100	71	↓ partial
"can you please help me choose the best"	N5	choose	0	9	29	emerging gap
"can you please help me understand"	N4	understand	62	91	96	↑ diverging
"can you please help me find the best"	N5	find	7	27	79	↑ semantic ahead
"can you please help me write"	N4	write	10	18	83	↑ semantic ahead
"can you please help me write a"	N5	write	3	9	83	↑ semantic ahead

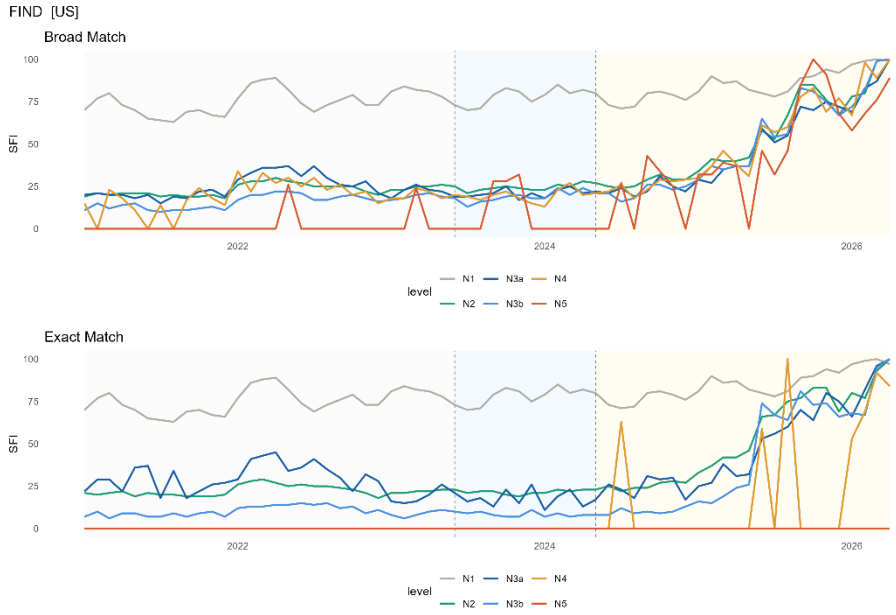
Gap = broad minus exact detectability (% non-zero months). ↓↓ strong convergence = gap narrows >70 pp from P1 to P3: semantic adoption completed in P1, syntactic convention converges in P3. ↑ diverging = semantic accelerates faster than exact, convention not yet formed. The N3b convergence pattern (write: 100→25 pp; understand: 97→29 pp) confirms lower-imbrication syntactic norms stabilize in P3 while N4–N5 forms remain syntactically fluid.

H.3 Nesting Figures

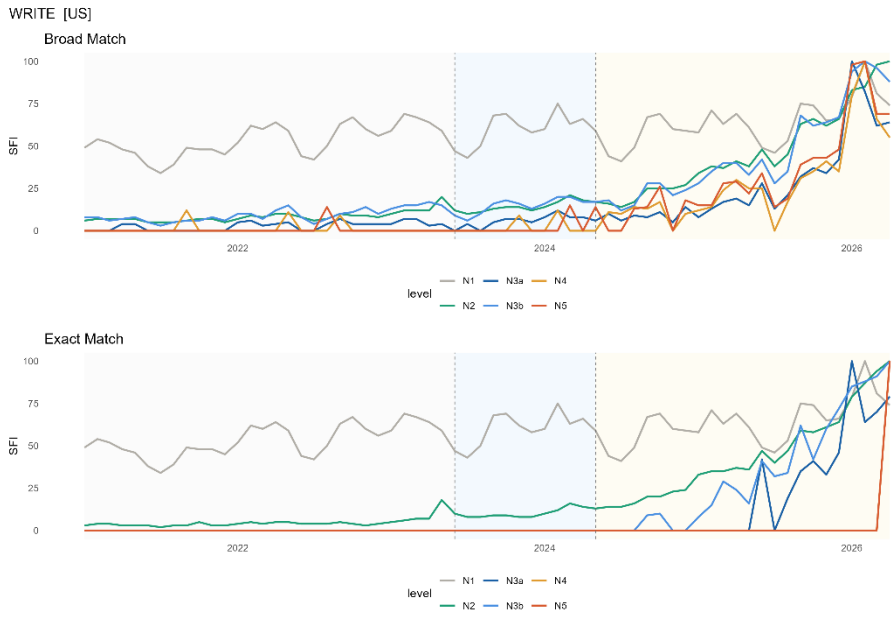
Three R scripts produce the nesting figures. Script 1 (7_Nesting_3periods.R): six shaded stacked figures (Supplementary Figure 11-16) and two period-faceted heatmaps (Supplementary Figure 17,18). **Script 2 (8_Nesting_2x2_panel_graphics.R):** six 2×2 trajectory panels (Supplementary Figure 19-24);. **Script 3 (9_Nesting_Matrix.R):** six per-family detectability matrices (Supplementary Figure 25-30) and one combined matrix (MAIN/Fig 9).

Three-Epoch Shaded SFI Trajectories

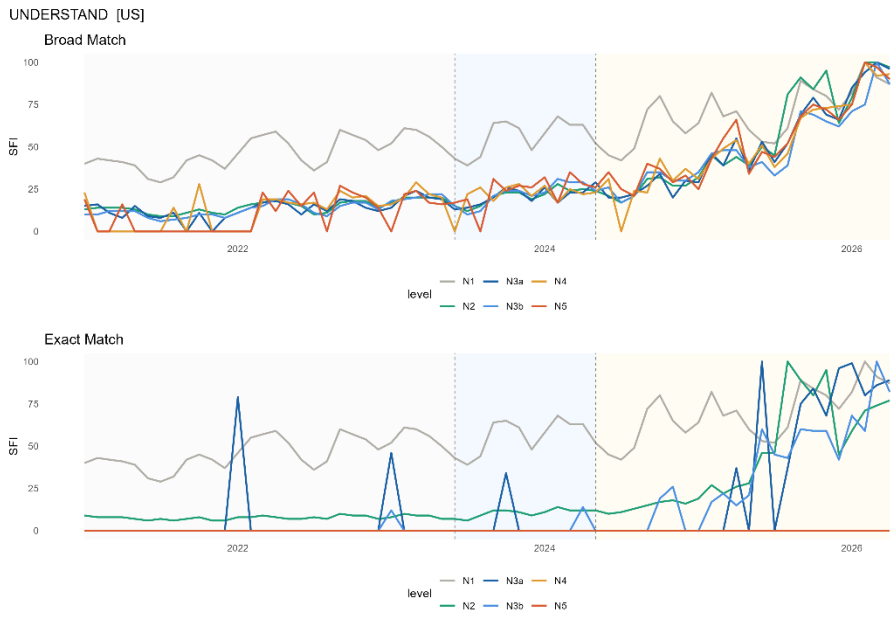
Each figure: vertically stacked two-panel (upper = Broad Match, lower = Exact Match), background shading for P1 (grey), P2 (light blue), P3 (light yellow), two vertical dashed lines (June 2023 LLM break + May 2024 AIO).



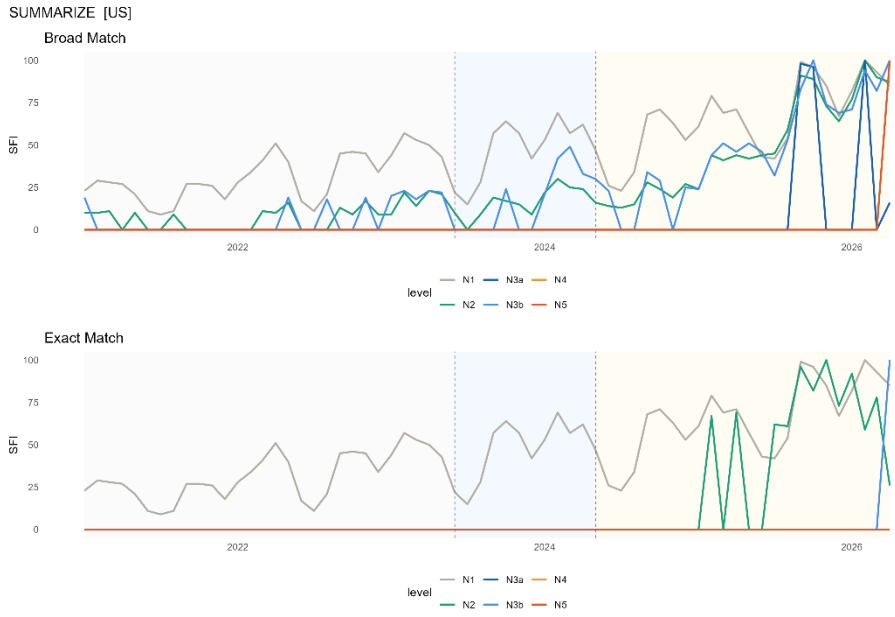
Supplementary Figure 11. Three-epoch shaded SFI trajectories for the "find" verb family. Upper: broad-match N1–N5 (monthly); lower: exact-match N1–N5. Background shading: P1 grey / P2 light blue / P3 light yellow. Exact-match N4 shows structural zero through P1–P2, first detection July 2024 (13-month lag). Broad-match N4 continuously non-zero (gap = 90 pp in P1, 71 pp in P3).



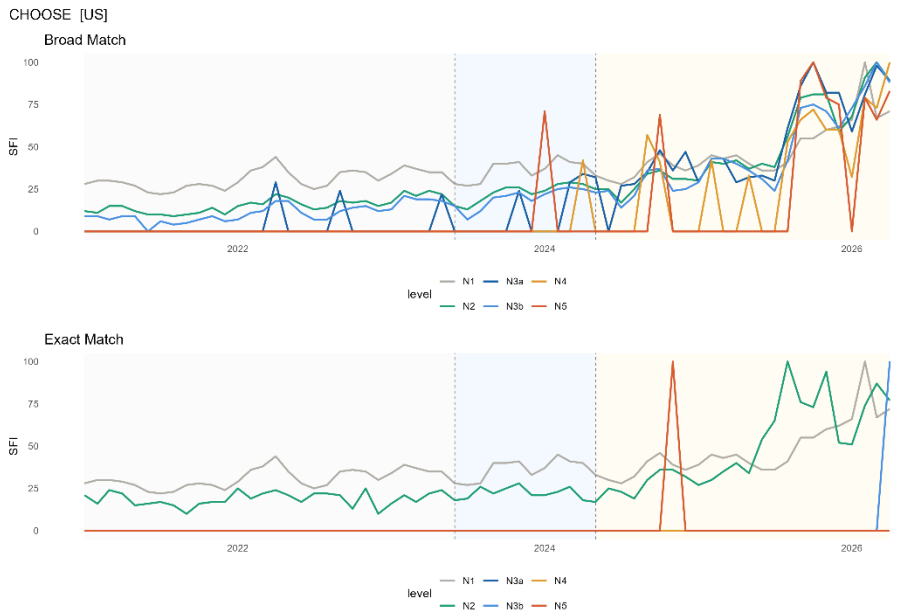
Supplementary Figure 12. Three-epoch shaded SFI trajectories for the "write" verb family. N3b exact-match gap: 100 pp in P1–P2 → 25 pp in P3 (strong convergence). N4–N5 absent throughout.



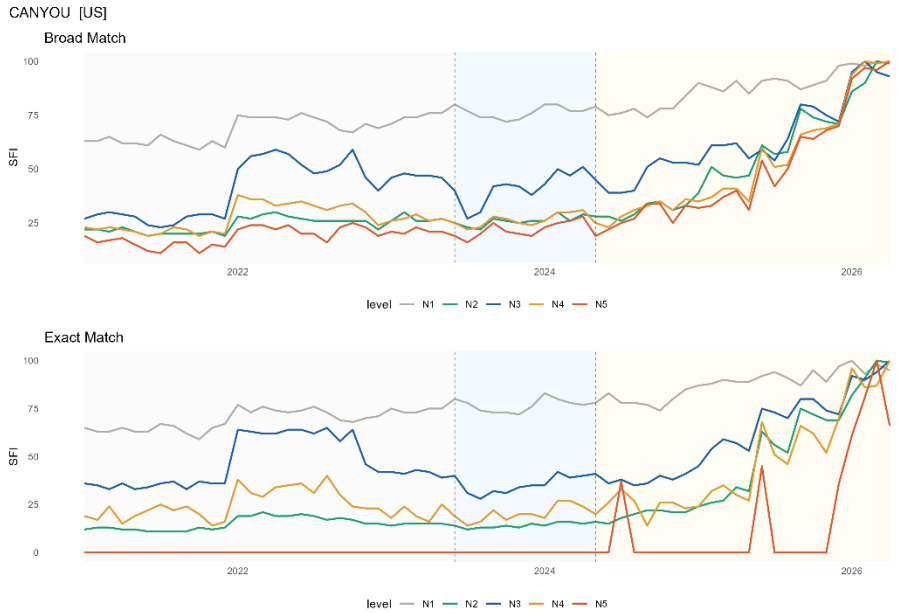
Supplementary Figure 13. Three-epoch shaded SFI trajectories for the "understand" verb family. N3b gap contracts 97 pp (P1) → 29 pp (P3); N4 gap diverges 62 pp (P1) → 96 pp (P3).



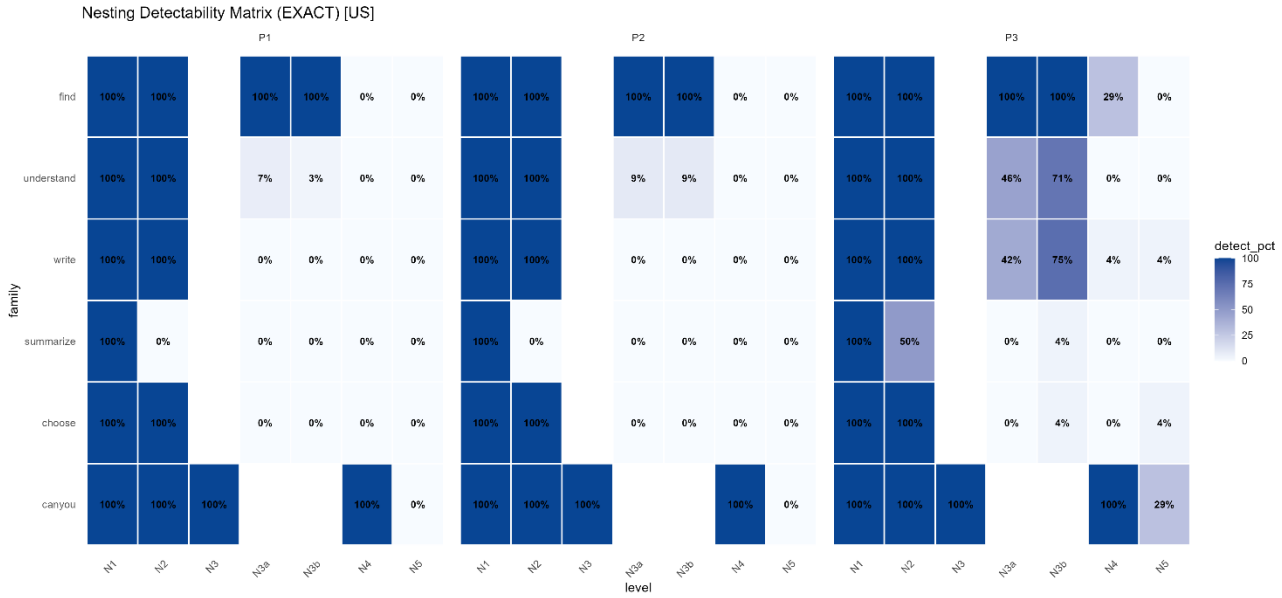
Supplementary Figure 14. Three-epoch shaded SFI trajectories for the "summarize" verb family. "help me summarize" (N2) uniquely absent in P1; first detection February 2025 (20-month lag). All N3a–N5 below threshold throughout.



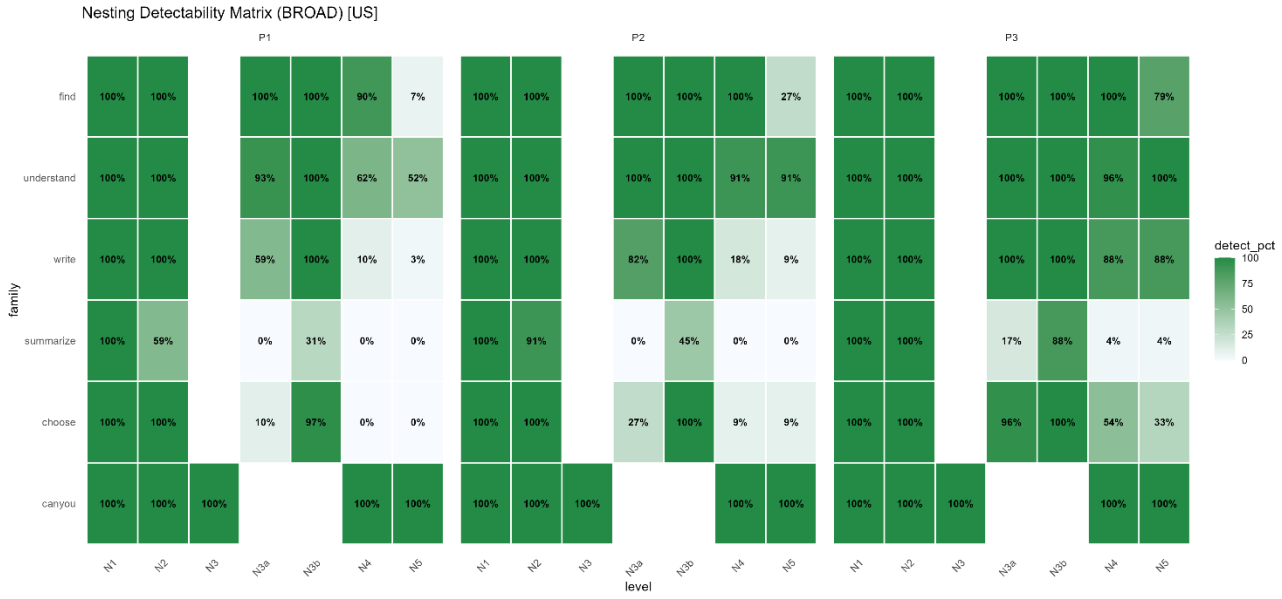
Supplementary Figure 15. Three-epoch shaded SFI trajectories for the "choose" verb family. N5 first detection November 2024 (17-month lag); N3b first detection April 2026 (34-month lag).



Supplementary Figure 16. Three-epoch shaded SFI trajectories for the "can you" verb family (5-level cascade). N5 ("can you please help me with") first detection July 2024 (13-month lag, co-first emergent with find N4). N4 ("can you please help me") pre-existing throughout.



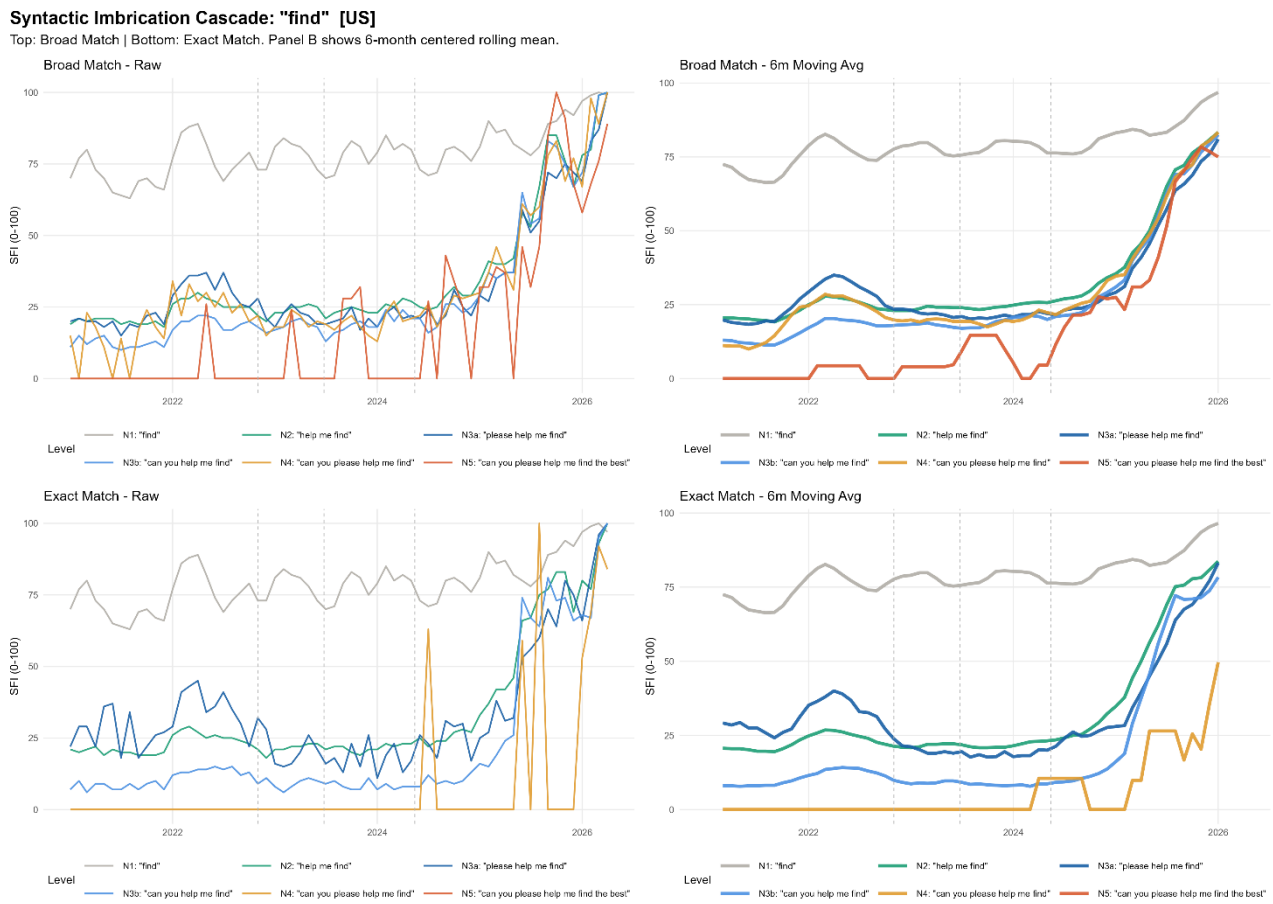
Supplementary Figure 17. Period-faceted detectability heatmap (exact-match). Rows: six verb families; columns: imbrication levels; facets: P1 | P2 | P3. Blue gradient. P2 panel: all N4–N5 cells = 0%, confirming the P2 null result. P3 panel: differential emergence by family and level.



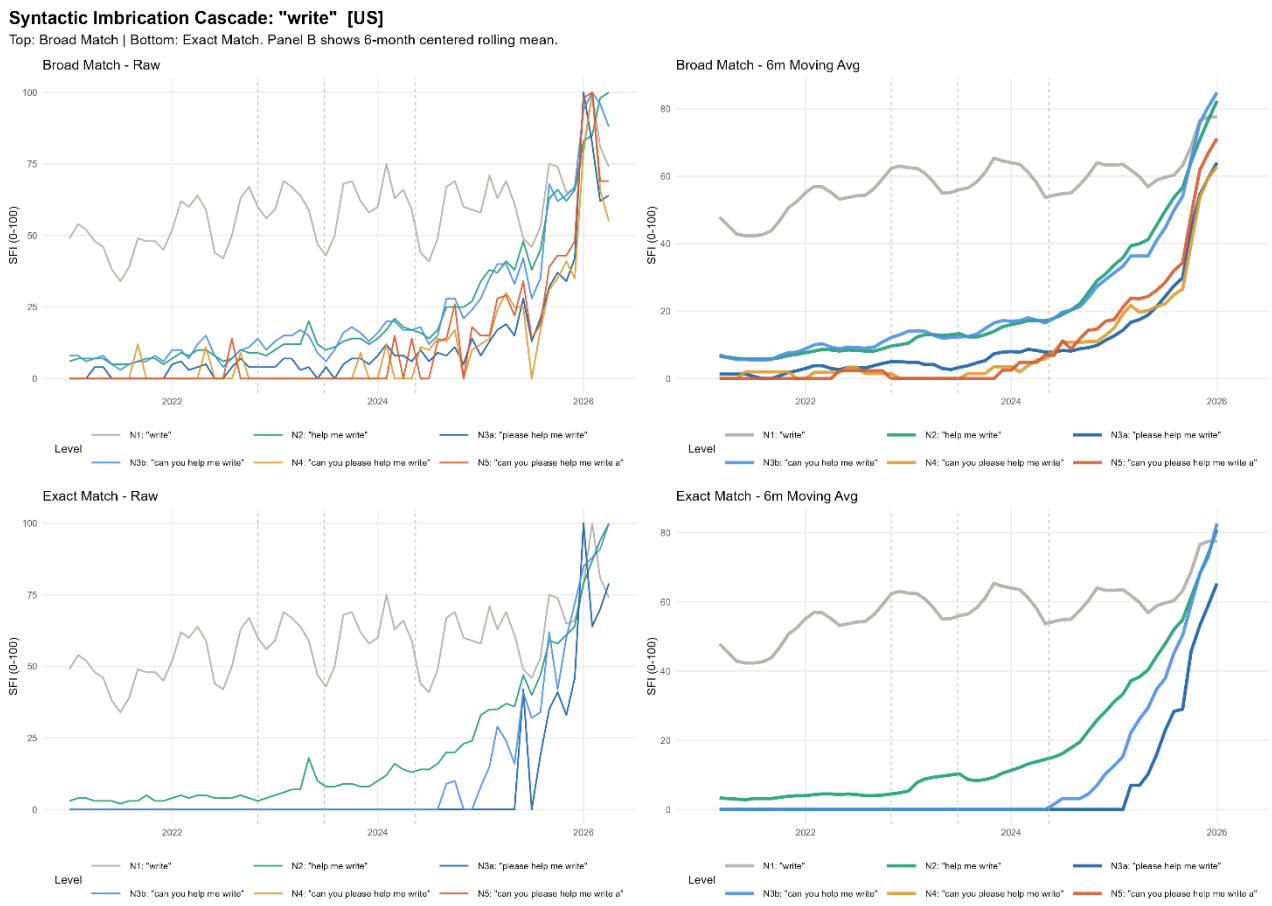
Supplementary Figure 18. Period-faceted detectability heatmap (broad-match). Same structure. Green gradient. Confirms semantic co-occurrence adoption was already present in P1–P2; P3 acceleration visible relative to exact-match heatmap.

Two-by-Two Trajectory Panels

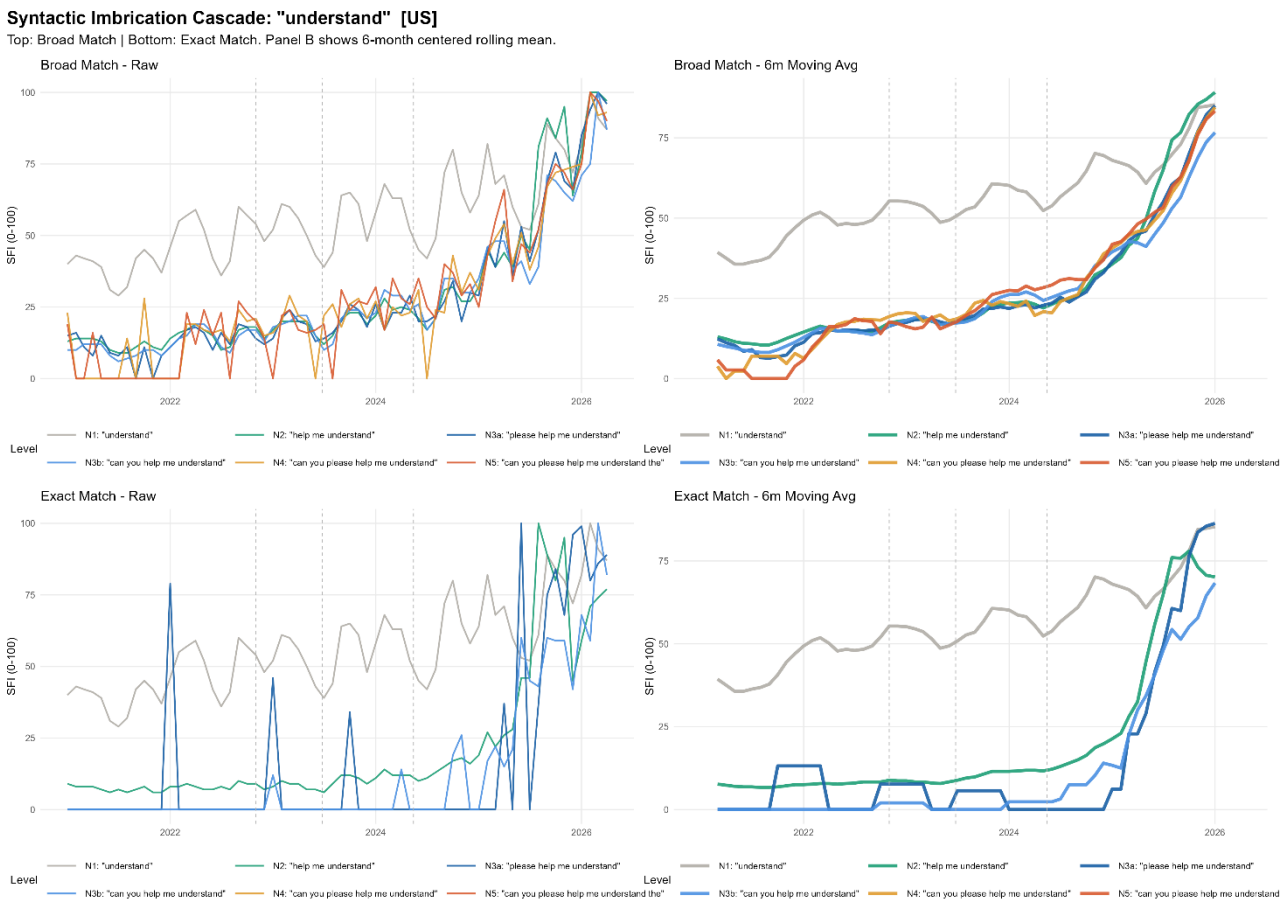
Each panel: TL = Broad Match Raw SFI; TR = Broad Match 6-month centred rolling average; BL = Exact Match Raw SFI; BR = Exact Match 6-month rolling average. Three vertical dashed lines on all panels: ChatGPT launch (November 2022), LLM structural break (June 25, 2023), AIO rollout (May 14, 2024). Level series suppressed if fewer than 3 detectable observations (ZERO_MIN = 3).



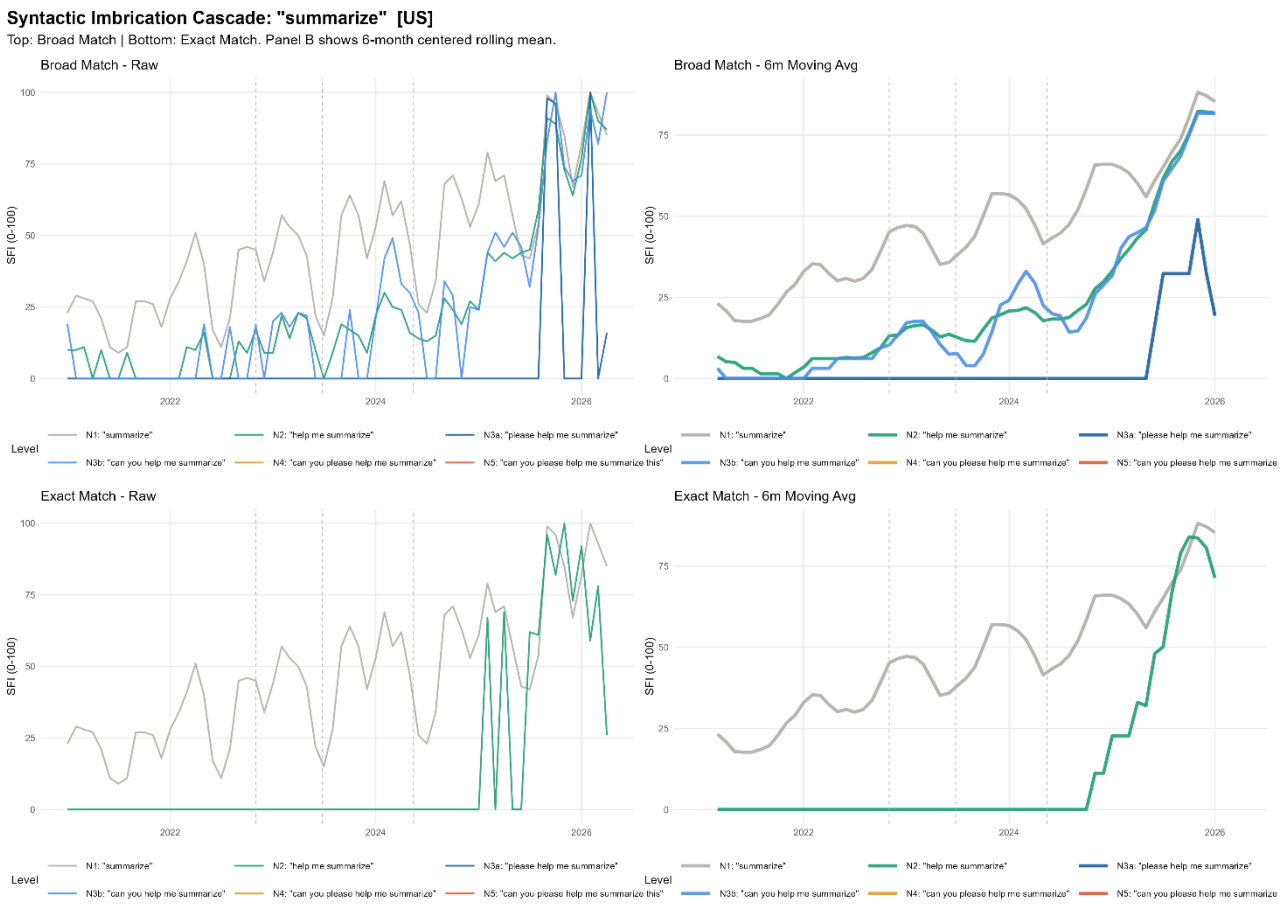
Supplementary Figure 19. Syntactic imbrication cascade for the "find" verb family (primary SI figure). TL: Broad Match raw. TR: Broad Match 6m avg. BL: Exact Match raw; N4 structural zero through P1–P2, step-function onset July 2024. BR: Exact Match smoothed; P3 onset confirmed. Three lines: ChatGPT / LLM break / AIO.



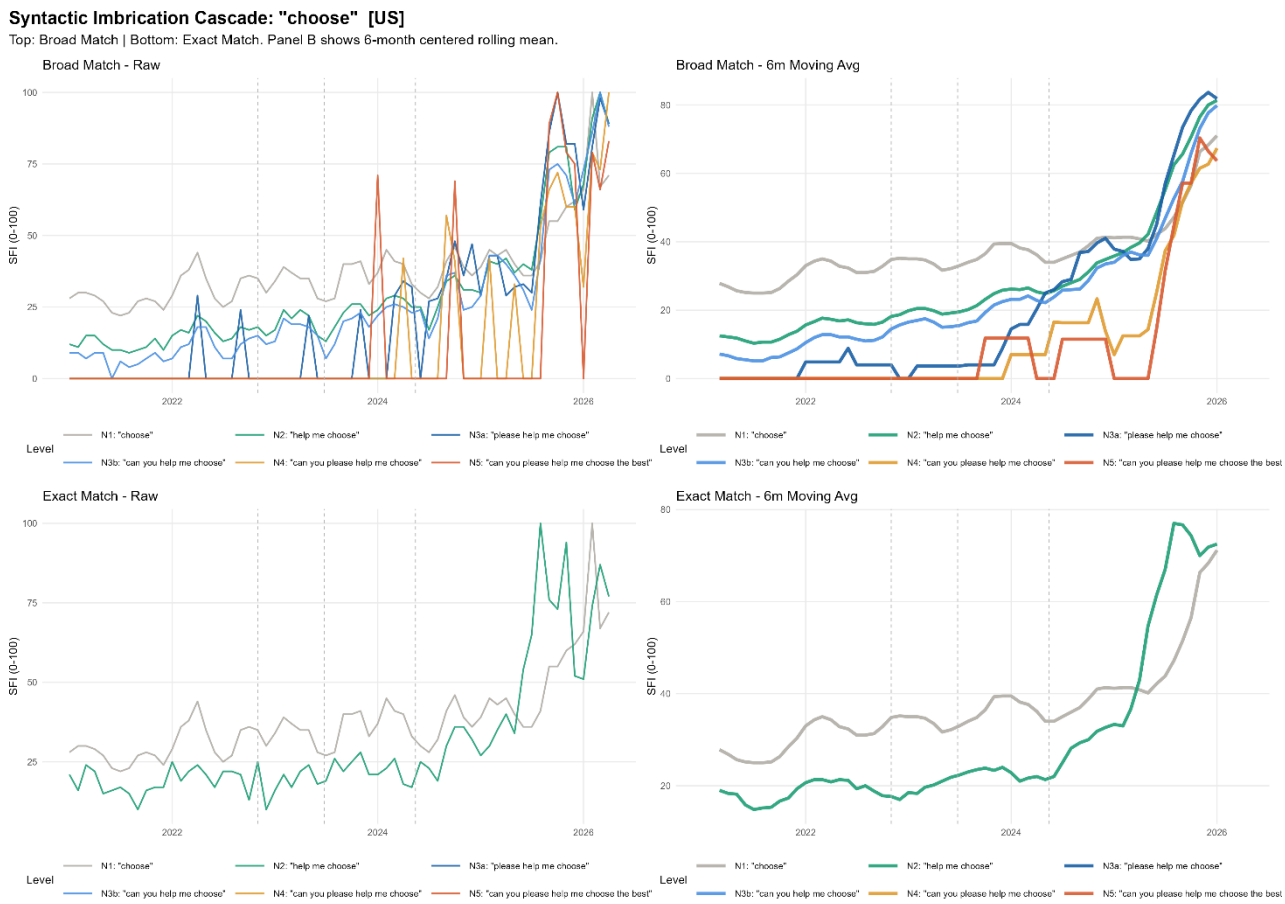
Supplementary Figure 20. Syntactic imbrication cascade for the "write" verb family (SI Replication). BL/BR: N3b absent through P1–P2, first detection September 2024 (15-month lag); N4–N5 absent throughout. TL/TR: N3b continuously detectable (100 pp broad-exact gap P1–P2 → 25 pp P3).



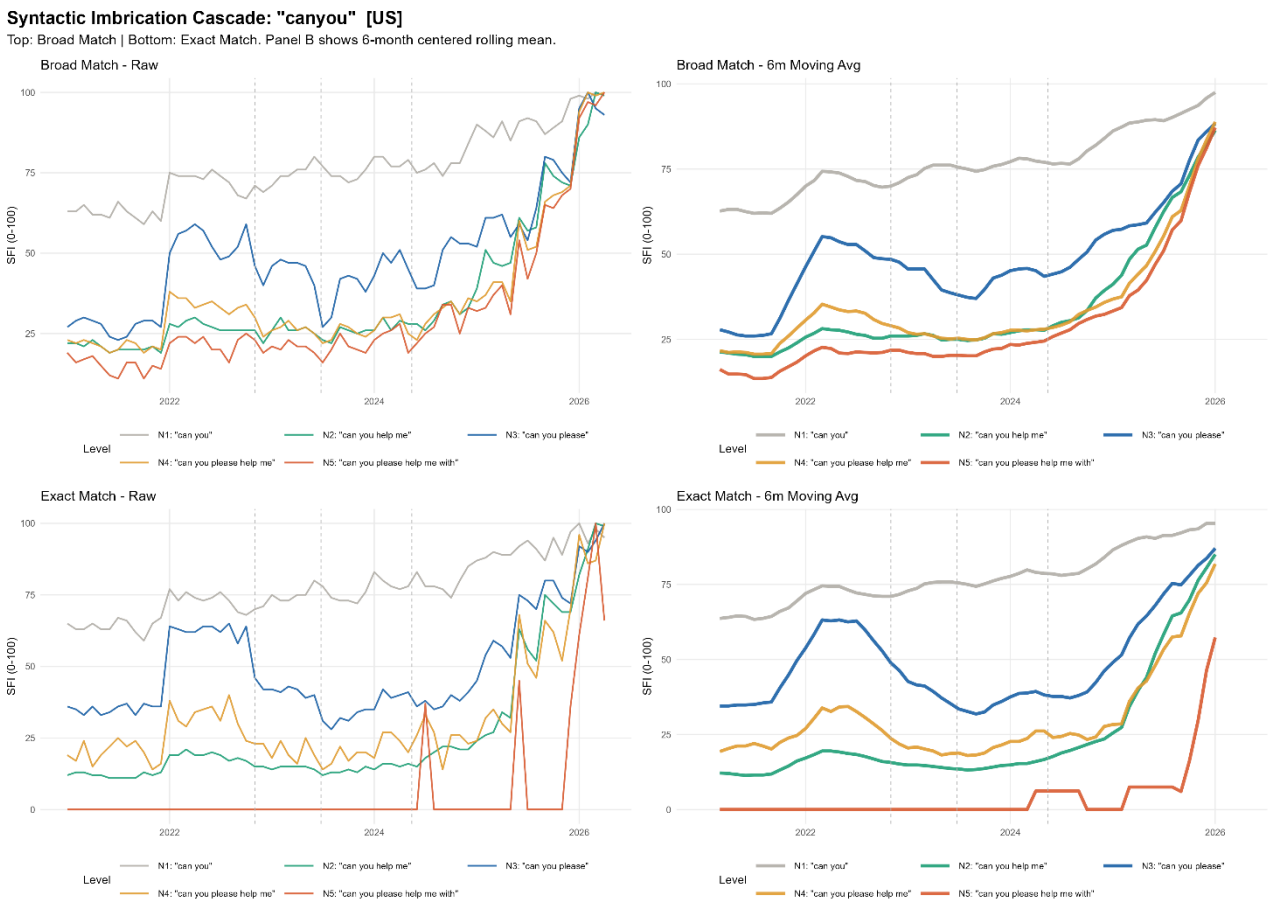
Supplementary Figure 21. Syntactic imbrication cascade for the "understand" verb family (SI Replication). N3b gap contracts 97 pp → 29 pp (P3). N4 gap diverges 62 pp → 96 pp (semantic outpacing exact).



Supplementary Figure 22. Syntactic imbrication cascade for the "summarize" verb family (SI Replication). N2 ("help me summarize") absent P1, first detection February 2025 (20-month lag). All N3a–N5 below threshold throughout.



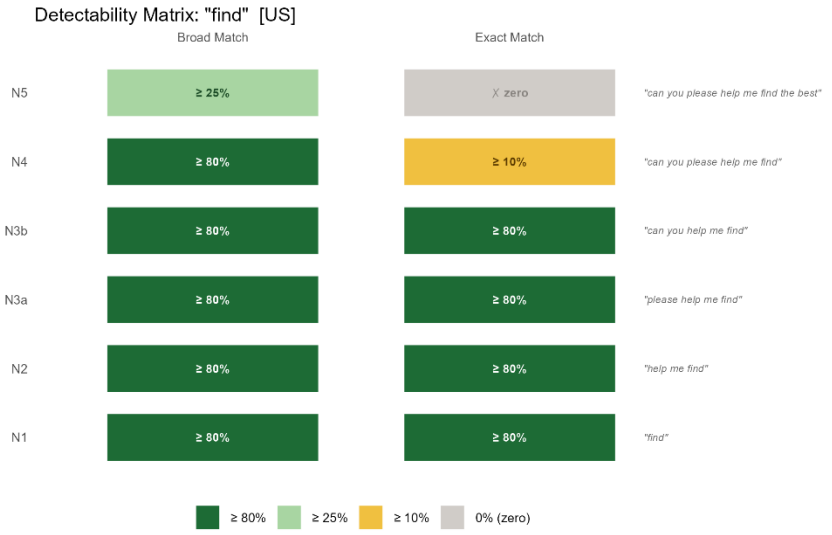
Supplementary Figure 23. Syntactic imbrication cascade for the "choose" verb family (SI Replication). N5 first detection November 2024 (17-month lag); N3b first detection April 2026 (34-month lag).



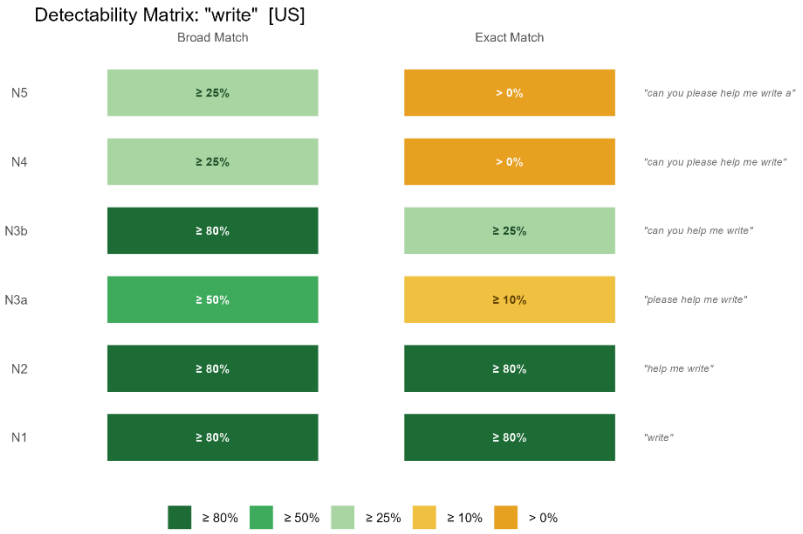
Supplementary Figure 24. Syntactic imbrication cascade for the "can you" verb family (primary SI figure; 5-level cascade). N5 ("can you please help me with") first detection July 2024 (13-month lag). N5 reaches 29% P3 detectability.

Detectability Matrices

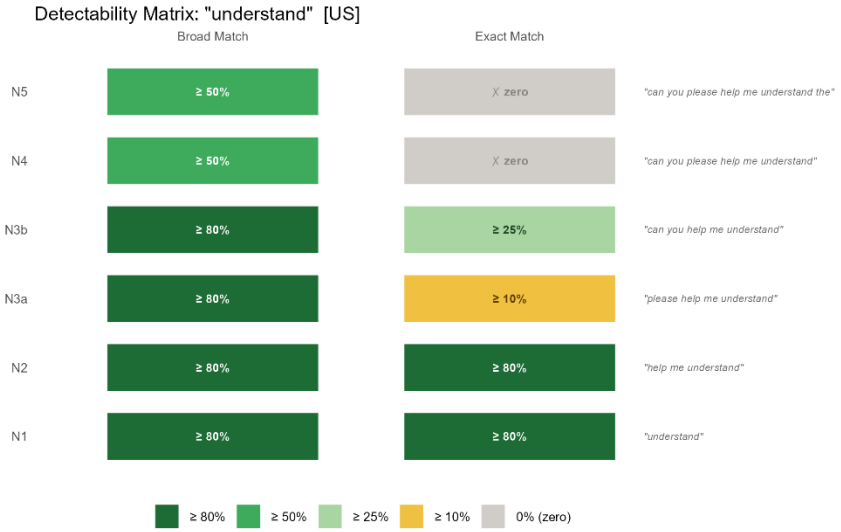
Each matrix: two columns (Broad Match left | Exact Match right) × imbrication levels, categorical fill (≥80% dark green → 0% grey), full observation window (all periods combined). Expression labels printed in italic to the right of each row.



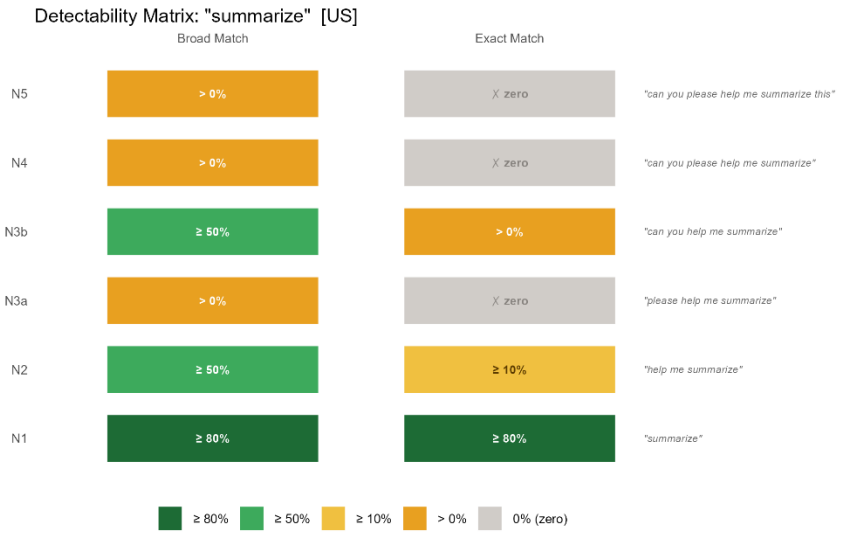
Supplementary Figure 25. Detectability matrix for the "find" verb family. Exact N4: >0% (29% P3 detectability diluted across full window); N5: 0%. Broad N1–N4: ≥80%; N5: ≥10%. Contrast at N4 visualises the 71 pp P3 broad-exact gap.



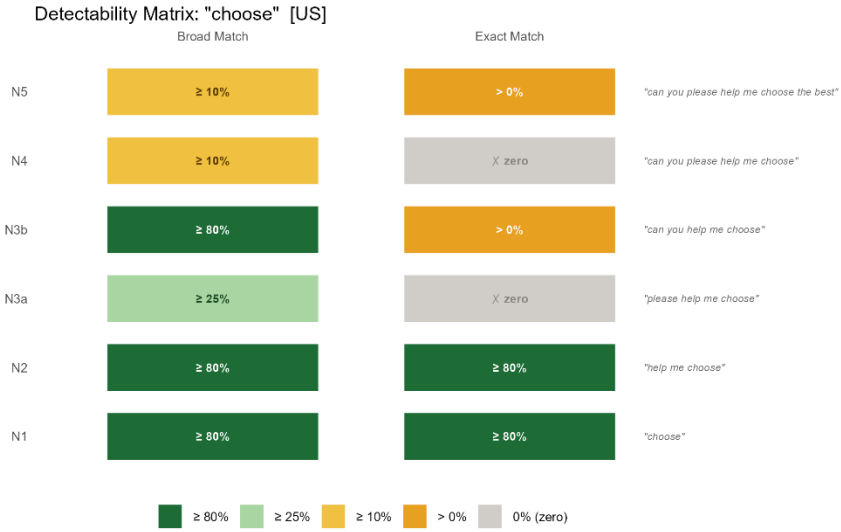
Supplementary Figure 26. Detectability matrix for the "write" verb family. Exact N3b: ≥25% (75% P3 detectability diluted); N4–N5: 0%. Broad N3b: ≥80% (100 pp P1–P2 gap → 25 pp P3).



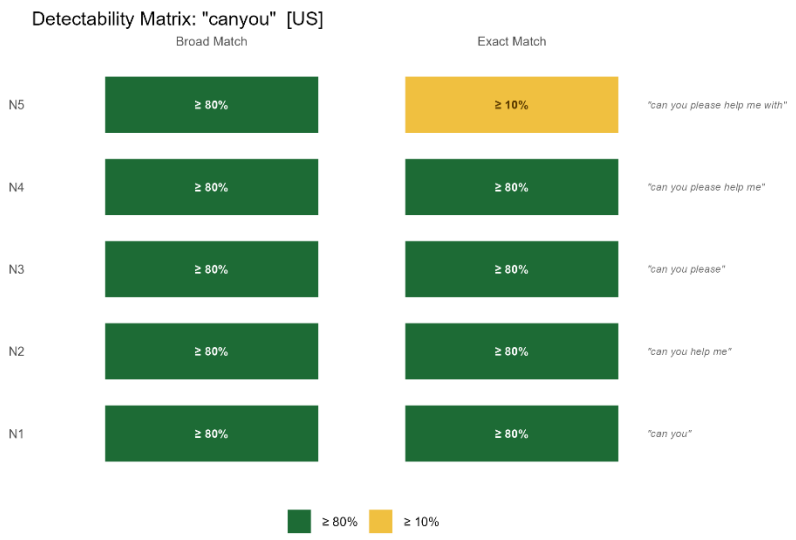
Supplementary Figure 27. Detectability matrix for the "understand" verb family. Exact N3b: ≥25%; N4: >0% (sparse P3; gap widens to 96 pp). N5: 0%.



Supplementary Figure 28. Detectability matrix for the "summarize" verb family. Exact N2: >0% (50% P3 detectability diluted). N3a–N5: 0%.



Supplementary Figure 29. Detectability matrix for the "choose" verb family. Exact N5: >0% (4% P3 only); N3b, N4: >0% (P3 only).



Supplementary Figure 30. Detectability matrix for the "can you" verb family (5-level cascade). Exact N5: >0% (29% P3 detectability). N4 ("can you please help me"): ≥80% in exact-match throughout (pre-existing form).

SI Appendix I: Worldwide Robustness and Generalization

I.1 Motivation and design

The primary specification of this paper restricts the panel to the United States, the geography where Google deployed AI Overviews (May 14, 2024) and AI Mode (early 2025) first. This restriction is methodological: the user-led versus platform-led sequencing test (the AIO triple-interaction specification) requires a setting in which the platform's generative interface rollout is dated and identifiable. The US is that setting.

Generalization to the worldwide population is, however, a legitimate concern. LLM adoption is heavily concentrated in high-income economies and varies substantially by language. Two opposing prior expectations are possible. On one hand, if the promptification phenomenon is a behavioral spillover driven by population-level LLM exposure, it should attenuate in regions where exposure is weaker, yielding smaller panel elasticities outside the US. On the other hand, if the phenomenon were a US-specific artifact of Google's product roadmap (for example, aggressive AIO promotion in US-facing surfaces leaking back into US-only Google Trends), it would be absent or muted in worldwide data.

To adjudicate, we re-run the entire primary specification (EXACT 10kw, 25 categories, weekly 2021–2025) on Google Trends Worldwide rather than US. All marker extractions, standardizations, structural break tests, panel estimations, Granger tests, cointegration tests, and the AIO triple interaction are re-executed on the worldwide series. The Worldwide and US panels are independent Google Trends extractions performed on the same calendar day, with all parameters identical except for the geography setting.

I.2 Key metrics: US versus Worldwide

Supplementary Table 27 summarizes the principal statistical outputs of the primary specification under both geographies.

Supplementary Table 27. Key statistical metrics, EXACT 10kw primary specification, US (panel A) versus Worldwide (panel B).

Metric	US (primary)	Worldwide	Interpretation
λ (Class 3, All Categories)	3.409	4.349	Larger λ in WW (Jenks thresholds rescaled)
Silhouette score (Jenks k=3)	0.543	0.557	Comparable partition validity
Jenks threshold λ_IE	1.695	2.610	Shifted upward in WW

Metric	US (primary)	Worldwide	Interpretation
Jenks threshold λ_ED	2.777	3.685	Shifted upward in WW
β_3 (TWFE, Class 3)	1.326 ($P = 3.22 \times 10^{-15}$)	0.994 ($P = 8.91 \times 10^{-4}$)	Attenuated in WW; remains highly significant
β_2 (TWFE, Class 2)	0.490 (P = 0.015)	−0.449 (P = 0.186, ns)	Class 2 differential collapses in WW
β_3 intra-spec.	1.338 ($P = 1.37 \times 10^{-15}$)	1.024 ($P = 3.97 \times 10^{-4}$)	Robustness pattern preserved
Granger A (LLM → Class 3) P	0.0021	0.0217	Significant in both; weaker in WW
Granger B (Class 3 → LLM) P	0.7319	0.0991	Reverse direction marginal in WW
Phillips-Ouliaris cointegration P	0.010	0.012	Cointegration confirmed in both
ANOVA λ by Class (F)	73.09 ($P < 2 \times 10^{-16}$)	48.13 ($P = 8.19 \times 10^{-14}$)	Class gradient highly significant in both
ANOVA r_intra by Profile (F)	19.43 ($P = 1.67 \times 10^{-5}$)	49.45 ($P = 1.14 \times 10^{-8}$)	Convergent profile validity stronger in WW
CUSUM S_0	2.485 ($P = 8.67 \times 10^{-6}$)	2.424 ($P = 1.58 \times 10^{-5}$)	Structural instability confirmed in both

Three observations follow:

First, all qualitative findings of the US specification replicate on Worldwide data: Conversational Delegation responds positively and significantly to LLM adoption; Granger temporal precedence runs from LLM to Class 3; Phillips-Ouliaris cointegration holds; and the class-level λ gradient remains highly significant under one-way ANOVA.

Second, the panel elasticity β_3 is attenuated in the Worldwide specification (0.994 versus 1.326). This attenuation is consistent with heterogeneous LLM exposure across the worldwide population: regions with lower or delayed LLM penetration dilute the population-scale average response. The proportional reduction (approximately 25%) is moderate, and the Worldwide estimate remains highly significant ($P < 10^{-3}$).

Third, the Class 2 (Directed Informational) differential response is null in Worldwide data ($\beta_2 = -0.449$, P = 0.186), in contrast to the small but positive US estimate ($\beta_2 = 0.490$). The Class 1 versus Class 3 contrast is preserved (and amplified relative to Class 1 baseline), so the monotonic ordering of register complexity holds at its end-points; Class 2 in Worldwide appears closer to baseline. We interpret this as a heterogeneity result: in cross-linguistic aggregation, the syntactic boundary between L2–L3 (Class 2) and the keyword baseline becomes noisier than in the US-English panel, where the marker syntax is uniformly English.

Direct test of the keyword baseline (worldwide). Re-estimating the per-class intra specification worldwide identifies the absolute elasticities reported in Supplementary Table 28. The worldwide pattern is sparser than the US gradient: only Conversational Delegation is significantly responsive in absolute terms (β_III = 1.631, $P = 6.2 \times 10^{-24}$), while both the keyword baseline and the directed-informational register are statistically indistinguishable from zero (β_I = 0.607, P = 0.122; β_II = 0.207, P = 0.224). The null absolute elasticity for Class II mirrors the collapse of its differential documented above ($\beta_2 = -0.449$, ns). The implied Class III − Class I differential (1.02) reproduces the worldwide global-proxy differential ($\beta_3 = 0.994$) and the intra differential ($\beta_{3,intra} = 1.024$), so the conservative cross-geography invariant is the dominance of conversational delegation over the keyword baseline, not the finer three-step monotone gradient observed in the US. If anything, the worldwide baseline is the cleaner null: outside the US-English panel the keyword register shows no detectable absolute response to adoption, isolating the conversational register as the locus of the behavioral shift.

Supplementary Table 28. Absolute per-class LLM elasticity, worldwide panel (category-internal proxy Z_LLM_intra, per-class TWFE form; EXACT 10kw, All Categories excluded; category and week fixed effects; Driscoll–Kraay HAC standard errors; N = 18,864 class-domain-week observations).

Intent class	Absolute β	SE	P
I. Factual Retrieval	0.607	0.392	0.122 (ns)

Intent class	Absolute β	SE	P
II. Directed Informational	0.207	0.170	0.224 (ns)
III. Conversational Delegation	1.631	0.161	6.2×10^{-24} ***

*Worldwide, only Class III shows a significant absolute elasticity; the keyword baseline and the directed-informational register are not distinguishable from zero. The Class III − Class I differential (1.02) matches the worldwide global-proxy β_3 (0.994; Supplementary Table 27) and the intra differential $\beta_{3,intra}$ (1.024). The Class II null is consistent with the Class 2 collapse reported in I.2. *** P < 0.001; ns not significant.*

I.3 Temporal Precedence: Error-Correction Model (Worldwide)

As in the US (E.4), the worldwide Class III index and the LLM-adoption proxy are I(1) and cointegrated (I.4), so predictive precedence is identified within the Engle–Granger error-correction model rather than on the raw levels. Supplementary Table 29 reports the worldwide estimates.

Supplementary Table 29. Error-correction precedence between LLM adoption and Class III, worldwide panel (EXACT 10kw, All Categories; n = 257, p = 4; HAC Newey–West). Cointegrating vector Z_Intent = a + 3.002·Z_LLM + ECT.

Equation (direction tested)	Long-run: adjustment φ on ECT_{t-1}	Short-run: joint Wald on Δ lags	Reading
ΔZ_Intent (LLM → Intent)	−0.117 (HAC 0.031), P = 0.0002 ***	F = 1.95, P = 0.103 ns	Intent adjusts (long-run)
ΔZ_LLM (Intent → LLM, falsification)	−0.005 (HAC 0.012), P = 0.644 ns	F = 1.07, P = 0.374 ns	LLM weakly exogenous

Adjustment half-life ≈ 6 weeks. *** P < 0.001; ns not significant.

The worldwide result reproduces the US directional structure in the long run: the error-correction term is significant and negative only in the conversational-delegation equation (φ_Intent = −0.117, HAC SE 0.031, P = 0.0002; adjustment half-life ≈ 6 weeks) and negligible in the adoption equation (φ_LLM = −0.005, P = 0.64). LLM adoption is therefore weakly exogenous worldwide and conversational query frequency adjusts to it, establishing long-run precedence from diffusion to syntax outside the US as well. Short-run causality is not detected in either direction (LLM → Intent F = 1.95, P = 0.103; reverse F = 1.07, P = 0.374), so the worldwide precedence operates through the long-run equilibrium channel rather than the short-run dynamics.

The error-correction specification also resolves an ambiguity left by the naive levels Granger cross-check. On the raw levels the worldwide reverse direction was only marginal (Direction A: F = 2.92, P = 0.022; Direction B: F = 1.97, P = 0.099; Supplementary Table 27), which under the unit-root non-stationarity documented for these series cannot be read as a valid test. Within the ECM the reverse direction is firmly non-significant in both the long-run adjustment (φ_LLM, P = 0.64) and the short run (P = 0.37): the directional asymmetry between LLM → syntax and syntax → LLM is preserved and, if anything, sharpened under the correct specification.

The broader worldwide attenuation — a smaller cointegrating adjustment, a slower half-life (≈ 6 weeks versus ≈ 4 in the US), and the loss of detectable short-run causality in both directions — is consistent with two non-exclusive factors: the worldwide series aggregates national LLM rollouts of staggered timing across many languages, some already evolving toward a conversational register, and the worldwide aggregate carries higher measurement noise. Both are exactly what one expects when aggregating a spatially and demographically heterogeneous effect. As throughout, these remain tests of predictive temporal precedence, not of structural causation; predictive precedence is sensitive to aggregation and noise in ways a counterfactual causal effect would not be.

I.4 Phillips-Ouliaris cointegration and long-run co-movement

Phillips-Ouliaris cointegration between Class 3 and the LLM proxy is confirmed in the Worldwide panel: demeaned test statistic = −27.82, P = 0.012, truncation lag parameter = 2. The cointegration relationship is therefore preserved

outside the US, consistent with the interpretation that LLM adoption and conversational delegation share a common long-run trend at worldwide population scale.

I.5 Bai-Perron endogenous break detection, US versus Worldwide

Supplementary Table 30 compares Bai-Perron endogenous break dates across the two geographies for the EXACT 10kw specification.

Supplementary Table 30. Bai-Perron endogenous break dates, EXACT 10kw specification, US versus Worldwide. BIC-optimal selection, h = 0.15.

Class	Break	US — date (95% CI)	Worldwide — date (95% CI)	Comparison
I. Factual Retrieval	1	Aug 21, 2022 (Jul 31 – Sep 18)	Aug 28, 2022 (Jul 24 – Sep 25)	Aligned (≈ 1 week apart)
I. Factual Retrieval	2	Aug 20, 2023 (May 21 – Oct 22)	—	WW: 2 breaks (no mid-2023 in Class 1)
I. Factual Retrieval	3	Mar 30, 2025 (Feb 9 – Apr 6)	Mar 30, 2025 (CI n.r.)	Aligned
II. Directed Inform.	1	Jan 2, 2022 (Dec 12 – Jan 16)	Jan 16, 2022 (Dec 19 – Jan 30)	GT artifact in both
II. Directed Inform.	2	Sep 1, 2024 (Jul 28 – Sep 29)	Jan 14, 2024 (Jul 9 – Aug 18)	WW: behavioral break ≈ 8 months earlier
II. Directed Inform.	3	—	Dec 29, 2024 (Nov 10 – Feb 9)	WW: 3 breaks; US: 2
III. Conv. Delegation	1	Jan 16, 2022 (Dec 26 – Jan 23)	Dec 26, 2021 (Dec 5 – Jan 2)	GT artifact in both
III. Conv. Delegation	2	Jan 14, 2024 (Nov 19 – Feb 4)	Jan 14, 2024 (Dec 31 – Jan 28)	Identical break date — robustness confirmed
III. Conv. Delegation	3	Jan 19, 2025 (Jan 5 – Jan 26)	Jan 19, 2025 (Jan 5 – Jan 26)	Identical break date
LLM proxy (ChatGPT)	1	Jun 25, 2023 (May 28 – Jul 23)	Jun 25, 2023 (May 28 – Jul 23)	Identical adoption anchor

The behavioral break of primary interest, Class III Break 2, is identified at the identical date (January 14, 2024) under both geographies, with overlapping confidence intervals. The LLM adoption anchor (June 25, 2023) is also identical. These two coincidences are the strongest individual robustness signals in this appendix: the population-level behavioral shift in conversational delegation is dated identically under independent extractions performed on US and Worldwide data.

The Class II structural break is identified earlier in Worldwide data (January 14, 2024) than in US data (September 1, 2024). One interpretation is that the directed-informational register (Class 2) had been more developed in the worldwide population prior to LLM diffusion (voice search, multilingual digital assistants), making the post-LLM standardization step less abrupt in the US-specific signal. We do not adjudicate and note the discrepancy as a heterogeneity result that does not affect the Class III findings.

I.6 Jenks classification: profile stability and rescaling

The Jenks natural-breaks classification (k = 3) is recomputed on the worldwide domain-level λ values. The Worldwide thresholds (λ_IE = 2.610, λ_ED = 3.685) are higher than the US thresholds (λ_IE = 1.695, λ_ED = 2.777), reflecting the overall rescaling of λ magnitudes in the Worldwide panel. The silhouette score is comparable (0.557 versus 0.543) and the ANOVA on r_intra (withheld from classification) is stronger in Worldwide ($F = 49.45$, $P = 1.14 \times 10^{-8}$) than in US ($F = 19.43$, $P = 1.67 \times 10^{-5}$), suggesting that the three-profile partition is at least as well supported in Worldwide as in US data.

Some individual domains migrate across profiles between the two geographies. In Worldwide data, Arts and Entertainment is classified as Emerging rather than Dominant (US), while Internet and Telecom moves from

Dominant (US) to Emerging (WW). These migrations reflect a combination of (i) rescaled Jenks thresholds in WW (more demanding lower bound for Dominant: λ_ED = 3.685 versus 2.777) and (ii) genuine cross-geographic differences in domain-level shift magnitudes. The qualitative ordering (Sports, Shopping, Health, Business and Industrial remain Dominant; Pets and Animals, Real Estate, Online Communities, News remain Incipient) is preserved. The key methodological point in SI Appendix G, that the Jenks classification reflects degrees of behavioral change, not a binary of affected versus unaffected, is reinforced: all worldwide domains again show positive λ for Class 3, with the lowest worldwide λ values still well above zero.

Supplementary Table 31. Domain-level Jenks shift profiles, Worldwide panel (EXACT 10kw). Rows: 24 domains, ordered by λ. λ: structural-break magnitude (Class 3, post-break minus pre-break in standardized units). r_intra: post-break Pearson correlation between domain-level Z_Intent and Z_LLM (convergent diagnostic, withheld from classification). r_post: full-window post-break correlation. KS-D: Kolmogorov-Smirnov directional statistic (post versus pre-break empirical distributions). Profile: Worldwide Jenks classification with thresholds λ_IE = 2.610, λ_ED = 3.685. Shading would correspond to the three profiles (Incipient, Emerging, Dominant).

Domain	λ	r_intra	r_post	KS-D	Profile
Health	4.945	0.909	0.904	0.787	Dominant
Autos & Vehicles	4.812	0.879	0.905	0.824	Dominant
Business & Industrial	4.504	0.91	0.914	0.817	Dominant
Shopping	4.459	0.93	0.895	0.833	Dominant
Home & Garden	4.422	0.93	0.908	0.734	Dominant
Beauty & Fitness	4.303	0.917	0.882	0.733	Dominant
Computers & Electronics	4.297	0.922	0.902	0.848	Dominant
Sports	4.01	0.881	0.796	0.712	Dominant
Hobbies & Leisure	3.685	0.9	0.859	0.817	Dominant
Internet & Telecom	3.558	0.904	0.862	0.816	Emerging
Arts & Entertainment	3.541	0.92	0.899	0.742	Emerging
People & Society	3.251	0.87	0.899	0.62	Emerging
Food & Drink	3.107	0.857	0.83	0.583	Emerging
Finance	2.981	0.878	0.848	0.594	Emerging
Jobs & Education	2.89	0.893	0.866	0.711	Emerging
Travel	2.843	0.892	0.85	0.725	Emerging
Games	2.61	0.84	0.794	0.671	Emerging
Online Communities	2.468	0.79	0.72	0.696	Incipient
News	2.305	0.786	0.741	0.528	Incipient
Pets & Animals	2.266	0.778	0.65	0.649	Incipient
Real Estate	2.047	0.821	0.817	0.484	Incipient
Reference	1.777	0.775	0.739	0.589	Incipient
Books & Literature	1.707	0.74	0.566	0.544	Incipient
Science	1.515	0.694	0.671	0.506	Incipient

Supplementary Table 32. Breakpoint sensitivity analysis, Worldwide panel (EXACT 10kw). β and λ estimates across six temporal anchors around the June 25, 2023 baseline (ChatGPT release, −8 weeks, −4 weeks, baseline, +4 weeks, +8 weeks). All Class 3 β estimates remain highly significant ($P \leq 1.50 \times 10^{-3}$); the Class 3 elasticity stays within the 0.972–1.099 range. Driscoll-Kraay HAC standard errors with bandwidth h = 4 weeks. The Class 2 estimates remain non-significant across all anchors, consistent with the Class 2 collapse documented in Worldwide data (I.2).

Anchor	Metric	Class	Estimate	SE	P-value
ChatGPT release	β (elasticity)	II. Directed Informational	−0.466	0.349	0.182
−8 weeks	β (elasticity)	II. Directed Informational	−0.454	0.339	0.181
−4 weeks	β (elasticity)	II. Directed Informational	−0.444	0.337	0.187
Baseline (Jun 25)	β (elasticity)	II. Directed Informational	−0.449	0.339	0.186
+4 weeks	β (elasticity)	II. Directed Informational	−0.452	0.343	0.187
+8 weeks	β (elasticity)	II. Directed Informational	−0.457	0.346	0.187
ChatGPT release	β (elasticity)	III. Conversational Delegation	1.099	0.306	3.33×10^{-4}
−8 weeks	β (elasticity)	III. Conversational Delegation	1.037	0.298	5.07×10^{-4}
−4 weeks	β (elasticity)	III. Conversational Delegation	1.019	0.296	5.82×10^{-4}
Baseline (Jun 25)	β (elasticity)	III. Conversational Delegation	0.994	0.299	8.91×10^{-4}
+4 weeks	β (elasticity)	III. Conversational Delegation	0.986	0.302	1.12×10^{-3}
+8 weeks	β (elasticity)	III. Conversational Delegation	0.972	0.306	1.50×10^{-3}
Baseline (Jun 25)	λ (magnitude)	III. Conversational Delegation	2.431	0.275	5.76×10^{-18}

I.7 Behavioral precedence under Worldwide data

The complete worldwide precedence analysis, including the epoch decomposition, the pre-deployment level shift, the regime-continuity tests, and the panel ceiling coefficient, is reported in SI Appendix F under the worldwide specification. For reading convenience we recall the headline results: the pre-deployment level shift for Conversational Delegation is large and highly significant ($\lambda_1 = 1.735$, $P = 3.9 \times 10^{-22}$), the regime-continuity Wald test is non-significant ($P = 0.205$), and the panel ceiling coefficient is significantly negative and larger in magnitude than in the US ($\beta_{AIO} = -5.334$, SE = 1.272, $P = 2.8 \times 10^{-5}$). The worldwide platform anchor is the staggered international rollout of AI Overviews (August 15, 2024). No separate worldwide AIO table is reported in this appendix; Supplementary Table 18 in SI Appendix F is the canonical source.

I.8 Interpretation and scope of generalization

The Worldwide replication supports four substantive conclusions.

(i) The structural shift in conversational delegation following LLM diffusion is not a US-only phenomenon. Class III differential response to LLM adoption is positive and highly significant in both geographies.

(ii) The panel elasticity is attenuated in worldwide aggregation, consistent with heterogeneous national-level LLM exposure. The proportional attenuation is moderate (approximately 25%) and does not threaten statistical significance.

(iii) The behavioral break date for Class III is identical under US and Worldwide extractions (January 14, 2024 for Break 2; January 19, 2025 for Break 3), as is the LLM-adoption anchor (June 25, 2023). This identity of break dates across two independent extractions is the strongest individual robustness signal in this appendix.

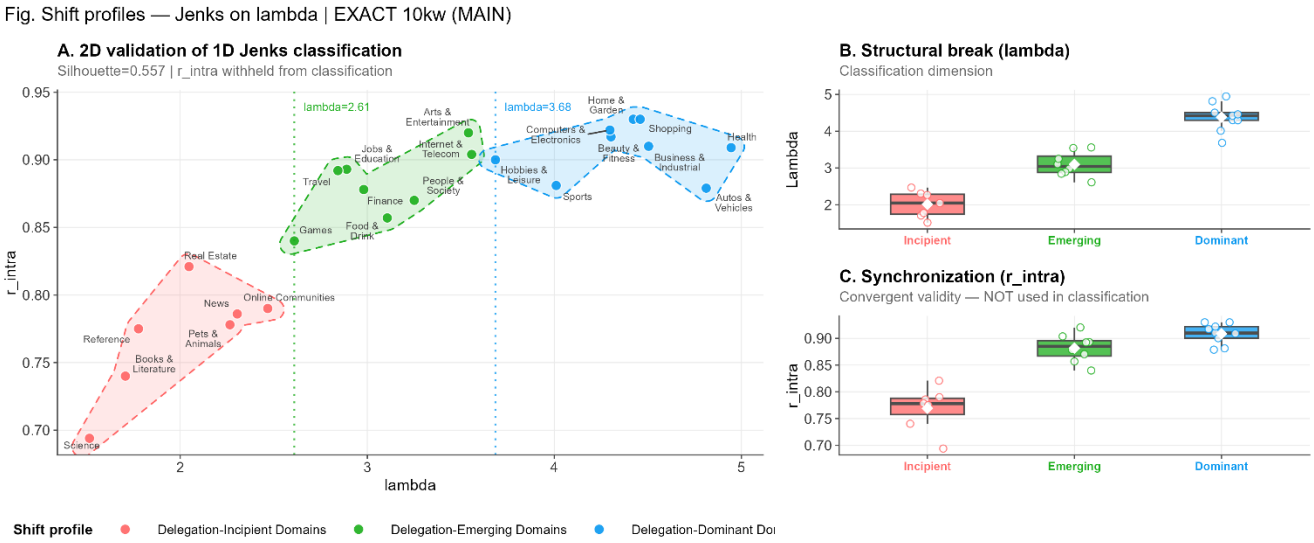
(iv) The AIO ceiling effect is preserved and amplified in Worldwide data, despite the staggered international rollout of Google AI Overviews. This is consistent with the user-led / platform-led sequencing claim being robust to substantial geographic heterogeneity in interface exposure.

These findings should not be over-read as a claim of universal cross-linguistic generalization. Google Trends Worldwide aggregates queries in all languages and from heterogeneous national markets; the EXACT 10kw marker set is in English. The Worldwide signal therefore primarily reflects English-language and other Latin-alphabet conversational query behavior. Replication in non-English national search ecosystems (notably for languages with substantively different syntactic structures: Mandarin, Arabic, Russian, Japanese) remains an open empirical question that a population-scale Google Trends extraction cannot resolve. The appropriate reading of this appendix is as a generalization test, not as a universality claim.

Taken together, the US and Worldwide results are mutually reinforcing: the US panel provides the cleanest sequencing test (because Google deployed its generative interface in the US first), and the Worldwide panel provides the broadest population-level robustness check that the phenomenon is not a US-specific platform artifact.

I.9 Reproducibility statement

Several extraction campaigns were carried out over the course of the study; all results reported in this appendix and in the main text derive from the final campaign, performed on May 3rd, 2026, on which the Worldwide series were extracted on the same calendar day as the primary US series. Each series is queried independently from the Google Trends API: the 30 intent markers together with the ChatGPT topic LLM-adoption proxy are extracted separately for each of the 25 semantic categories, producing 775 independent API requests per geography (1,550 in total across the US and Worldwide panels). All other parameters are identical between the two geographies: the observation window (January 3, 2021 – December 28, 2025), the weekly granularity, and the EXACT 10kw extraction syntax. The only difference between the US panel and the Worldwide panel is the geography parameter passed to the Google Trends API. Performing all 1,550 requests within a single calendar day neutralises the risk of unannounced Google Trends algorithmic updates between extractions and guarantees identical sampling-frame conditions for the two panels. All downstream code (standardisation, two-way fixed-effects estimation, Granger tests, Phillips-Ouliaris cointegration, Bai-Perron break detection, Jenks classification, and the AIO triple-interaction specification) is invariant across the two geographies and is re-run mechanically on the worldwide series.



Supplementary Figure 31. Shift profiles for the Worldwide panel (EXACT 10kw). Panel A: two-dimensional validation of the one-dimensional Jenks classification, with λ on the x-axis and r_intra on the y-axis for the 24 domains. The classification is performed on λ alone (silhouette = 0.557); r_intra is withheld from the classification and shown as a convergent diagnostic. Vertical dotted lines: Worldwide Jenks thresholds λ_IE = 2.610 and λ_ED = 3.685. Panel B: boxplot of λ by profile (the classification dimension). Panel C: boxplot of r_intra by profile (convergent validity, not used in classification). Domains migrate across profiles between US and Worldwide consistent with the rescaled Jenks thresholds documented in I.6; the qualitative ordering Incipient < Emerging < Dominant is preserved.

I.10 Worldwide nesting cascade

The nesting cascade decomposition (Section F) is re-estimated on the worldwide panel using the same six verb families, six imbrication levels, and three-epoch structure (P1: January 2021 to May 2023; P2: June 2023 to April 2024; P3: May 2024 to April 2026). The extraction is independent from the US run, conducted on the same calendar day with monthly aggregation. Four findings are reported: P2 null result preservation, P3 emergence timing, broad versus exact convergence, and a synthesis comparison with the US.

I.10.1 P2 detection (worldwide)

The structural zero in P2 for N4 and N5 forms is preserved across all six families (Supplementary Table 33). For N3a and N3b forms in the write and understand families, P2 detection rates rise above zero worldwide (write N3a: 54.5%, write N3b: 81.8%, understand N3b: 36.4%), whereas the same forms remain at zero in the US through P2 (Supplementary Table _22). The single non-zero N4 cell, find (27.3%), is a pre-existing worldwide form rather than

a new emergence, and therefore does not bear on the P2 null, which concerns P1-absent forms. The pattern is consistent with earlier convergence on intermediate-complexity expressions outside the US.

Supplementary Table 33. P2 detection rates worldwide for the 18 high-imbrication candidate forms (EXACT 10kw, monthly). Detection = percent of P2 months (N = 11) with non-zero SFI.

Family	Level	Expression	P2 detection (%)	Note
canyou	N4	can you please help me	100.0	Pre-existing
canyou	N5	can you please help me with	100.0	Pre-existing
choose	N3a	please help me choose	0.0	Structural zero
choose	N3b	can you help me choose	0.0	Structural zero
choose	N4	can you please help me choose	0.0	Structural zero
choose	N5	can you please help me choose the best	0.0	Structural zero
find	N3a	please help me find	100.0	Pre-existing
find	N3b	can you help me find	100.0	Pre-existing
find	N4	can you please help me find	21.4	Emerging
find	N5	can you please help me find the best	0.0	Structural zero
summarize	N3a	please help me summarize	0.0	Structural zero
summarize	N3b	can you help me summarize	0.0	Structural zero
summarize	N4	can you please help me summarize	0.0	Structural zero
summarize	N5	can you please help me summarize this	0.0	Structural zero
understand	N3a	please help me understand	14.3	Emerging
understand	N3b	can you help me understand	42.9	Emerging
understand	N4	can you please help me understand	0.0	Structural zero
understand	N5	can you please help me understand the	0.0	Structural zero
write	N3a	please help me write	64.3	Emerging
write	N3b	can you help me write	85.7	Emerging
write	N4	can you please help me write	0.0	Structural zero
write	N5	can you please help me write a	0.0	Structural zero

I.10.2 P3 emergence and intensification (worldwide)

Five exact-match expressions reach population-level detectability in P3 worldwide (Supplementary Table 34), with a median emergence lag of 31 months from the June 2023 anchor. Four N4 and N5 expressions in the summarize, find, and choose families remain below detectability through April 2026. P3/P1 intensification ratios for N3 forms reach extreme values worldwide: ×33.3 for "please help me write" (N3a, write), ×23.5 for "can you help me understand" (N3b, understand), ×15.8 for "can you help me choose" (N3b, choose). These magnitudes reflect near-floor P1 baselines (mean SFI between 1.1 and 1.9), indicating a later worldwide baseline adoption of these intermediate forms, rather than stronger intensification at comparable baselines: for forms with established P1 volume on both panels, ratios are of the same order (×8.3 worldwide versus ×9.6 US for "help me write"; ×5.5 versus ×4.1 for "can you help me find").

Supplementary Table 34. P3 emergence timing and intensification ratios for the worldwide nesting cascade (EXACT 10kw, monthly). First date: month of first non-zero SFI in P3. Lag: months from June 2023 to first detection. P3 detection: percent of P3 months (N = 24) with non-zero SFI. Forms with no detection through April 2026 are shown with a hyphen.

Family	Level	Expression	First detection	Lag (months)	P3 detection (%)
summarize	N2	help me summarize	Sep 2024	15.0	85.7
summarize	N3b	can you help me summarize	Oct 2025	28.0	4.8
choose	N3a	please help me choose	Jan 2026	31.0	9.5
write	N4	can you please help me write	Jan 2026	31.0	14.3
write	N5	can you please help me write a	Feb 2026	32.1	4.8
choose	N5	can you please help me choose the best	-	-	0.0
find	N5	can you please help me find the best	-	-	0.0
summarize	N3a	please help me summarize	-	-	0.0
summarize	N4	can you please help me summarize	-	-	0.0

I.10.3 Broad versus exact gap (worldwide)

The broad-versus-exact detectability gap for N3b forms collapses more completely in worldwide data than in the US (Supplementary Table 35). For "can you help me write" (N3b, write), the gap falls to 0 percentage points in P3 worldwide versus 25 pp in the US. For "can you help me understand" (N3b, understand), the gap is 4.8 pp worldwide versus 29 pp in the US. At N4 and N5 imbrication, the gap remains wide in both geographies, consistent with the imbrication-stratified standardization documented for the US panel (Section H.2).

Supplementary Table 35. Broad-versus-exact detectability gap by period, worldwide panel (EXACT 10kw, monthly). Gap = broad detection percentage minus exact detection percentage, in percentage points. Positive values indicate broader semantic adoption outpacing exact syntactic standardization.

Family	Level	Expression	Gap P1 (pp)	Gap P2 (pp)	Gap P3 (pp)
canyou	N4	can you please help me	0.0	0.0	0.0
canyou	N5	can you please help me with	0.0	0.0	0.0
choose	N3a	please help me choose	100.0	100.0	90.5
choose	N3b	can you help me choose	96.6	100.0	61.9
choose	N4	can you please help me choose	82.8	92.9	100.0
choose	N5	can you please help me choose the best	17.2	7.1	85.7
find	N3a	please help me find	0.0	0.0	0.0
find	N3b	can you help me find	0.0	0.0	0.0
find	N4	can you please help me find	82.8	78.6	33.3
find	N5	can you please help me find the best	79.3	100.0	100.0
summarize	N3a	please help me summarize	0.0	14.3	28.6
summarize	N3b	can you help me summarize	58.6	78.6	95.2
summarize	N4	can you please help me summarize	3.4	0.0	14.3
summarize	N5	can you please help me summarize this	-3.4	0.0	4.8
understand	N3a	please help me understand	69.0	85.7	9.5
understand	N3b	can you help me understand	79.3	57.1	4.8

understand	N4	can you please help me understand	96.6	100.0	95.2
understand	N5	can you please help me understand the	96.6	100.0	100.0
write	N3a	please help me write	86.2	35.7	0.0
write	N3b	can you help me write	96.6	14.3	0.0
write	N4	can you please help me write	100.0	100.0	85.7
write	N5	can you please help me write a	96.6	100.0	95.2



Supplementary Figure 32. Worldwide nesting detectability matrix across six verb families (find, understand, write, summarize, choose, canyou) and six imbrication levels (N1 to N5, with N3a/N3b distinction where applicable). Upper sub-panel: broad match. Lower sub-panel: exact match. Cell shading: percent of non-zero months within each period (P1, P2, P3). The qualitative pattern mirrors the US matrix (Fig. 9 of the Main), with attenuation in absolute detection rates and longer emergence lags as documented in I.10.1 to I.10.3.

I.10.4 US versus Worldwide synthesis

Supplementary Table 36 summarizes the principal contrasts between the US and worldwide nesting cascades. The P2 null result for N4 to N5 forms is preserved across the two extractions. P3 emergence timing is systematically delayed worldwide (median lag 31 months versus 22 months in the US, over P1-absent forms emerging in P3). N3 forms reach extreme P3/P1 intensification ratios worldwide, reflecting near-floor P1 baselines, and show tighter broad-versus-exact convergence at N3b. Taken together, the worldwide nesting cascade preserves the qualitative structure of the US cascade while documenting an attenuation pattern consistent with the panel-level results reported in I.1 to I.7.

Supplementary Table 36. Principal contrasts between the US and worldwide nesting cascades.

Metric	US	Worldwide	Note
P2 null result (P1-absent N4–N5 forms)	Preserved (zero new emergence among P1-absent forms)	Preserved (zero new emergence among P1-absent forms)	Strict robustness signal
P2 detection of N3 forms (write, understand)	0% (no P2 emergence)	Above zero (write N3a: 64.3%, write N3b: 85.7%, understand N3b: 42.9%)	Earlier WW convergence on intermediate complexity
Median P3 emergence lag	22 months (n = 10)	31 months (n = 5)	Median over P1-absent forms emerging in P3; consistent with WW attenuation
P1-absent forms reaching P3 detectability	10 of 18 candidates	5 of 9 candidates	Comparable closure of the cascade
P3/P1 ratio: help me write (N2)	×9.6	×8.3	Comparable pre-existing form; same order across geographies
P3/P1 ratio: can you help me find (N3b)	×4.1	×5.5	Comparable pre-existing form; same order across geographies
Broad-exact gap, can you help me write N3b (P3)	25 pp	0 pp	Tighter WW convergence at N3b
Broad-exact gap, can you help me understand N3b (P3)	29 pp	4.8 pp	Tighter WW convergence at N3b

SI Appendix J : AI Usage Declaration

AI tools, providers and versions: Claude (Anthropic) Sonnet 4.6 and Opus 4.8; Gemini (Google) 3.1.

Period of use: December 2025 to June 2026

Scope of AI use (summary): The authors used AI in a strictly limited and assistive capacity. Applications included writing and debugging analysis code, polishing English wording and formatting the Supporting Information. AI produced no research ideas, data, results, analyses, decisions, references, or images. All intellectual content, computations, and outputs belong to the authors. The authors edited, executed, and verified all materials.

J.1. Detailed Declaration of AI Usage

Area	Whether and how AI was used?	Specific details
Brainstorming & ideation	**No AI**	The authors solely developed the research question, the conceptual framework (five-register scale, three intent classes, three-epoch design, nesting cascade), and the hypotheses.

Data collection & sourcing	**No AI**	The authors extracted all series from Google Trends via the gtrendsR package (US and Worldwide; weekly and monthly). AI generated no data.
Conceptual framework & analysis	**Completed by Authors (AI assisted with code only)**	The authors designed, specified, and executed the TWFE, cointegration, ECM, Bai-Perron, three-epoch/AIO, Jenks, and nesting frameworks locally. AI assisted with coding and explaining specific R error messages. The authors delegated no estimates, decisions, or interpretations.
Statistical computation & code	**Partially by AI (Assistive debugging)**	Claude (Sonnet 4.6, Opus 4.8) and Gemini (3.1) helped write and debug the R code for the econometric pipeline and robustness checks. The authors executed and verified all code. All reported values originate from the authors's local runs.
Figures & data visualization	**Partially by AI (Code generation only)**	The authors produced all data figures (Fig. 2 to 9 and SI) via local R/ggplot2 scripts. No figure constitutes a generative-AI image. For the conceptual schematic (Fig. 1), the authors wrote an initial Python/matplotlib script based on a detailed textual framework and target graphical design. Claude and Gemini iteratively enhanced and debugged this code. The authors subsequently executed the final script locally.
Literature search & references	**No AI**	The authors selected, verified, and formatted all 85 references manually using Zotero. AI did not find, hallucinate, or suggest citations.
Manuscript writing	**No AI**	The authors wrote the structure, scientific content, claims, and core arguments of the manuscript and SI entirely.
Manuscript editing & language	**Partially by AI (Assistive polishing)**	Claude and Gemini provided grammatical corrections to meet PNAS academic standards. Example prompt: "Review this paragraph to ensure correct English academic grammar." The authors continuously refined and manually modified the AI outputs to produce the final version. The authors retained all final editorial decisions.
Supporting Information (SI)	**Partially by AI (Assistive structuring)**	The authors manually integrated all data tables and graphics into the 50-page SI document. The authors wrote all comments, legends, and titles manually. Subsequently, Claude optimized the overall structure, numbered the appendices (A to I), and checked grammatical structures. The authors verified all final formatting.
Supplementary aids	**Partially by AI (Assistive coding)**	The authors manually wrote the initial code for an internal study architecture map (Python/Graphviz/matplotlib). This map serves exclusively as a guide within the replication package to map data inputs, scripts, and outputs. The map does not appear in the Main manuscript or the Supporting Information. The authors provided the textual structure and the target design. Claude assisted in

		enhancing the visualization parameters. The authors iteratively reviewed, debugged, and executed the final code locally.

Declaration of AI-Assisted Technologies in the Research Process: 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 Figure 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 the 85 references manually using Zotero. The authors executed all computational scripts locally and reviewed all AI-assisted code and textual edits.
Authors take full responsibility for the integrity of the content and conclusions of this article.

SI References:

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.

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.

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

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.

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.

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

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.

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.

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.

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, 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.

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.

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.