



Growfer

OCTOBER 5, 2026

GROWFER RESEARCH · PRIMARY STUDY REPORT

The Split-Pricing Study: **Final Report**

When a small business publishes its pricing on the public web, can AI answer engines surface it — and when they can't, why not?

Public research edition

Revision 6

1,123 US SMBs · 2 engines



PUBLIC RESEARCH EDITION

This document is a public, redacted edition of an internal Growfer Research report. Business identities, domains, URLs, row-level excerpts, and internal repository references have been removed or anonymized. Aggregate findings, methodology, statistical summaries, limitations, and research conclusions are preserved from the cited report revision (Revision 6, October 5, 2026). Where an example is needed to explain a pattern, it appears as an anonymous pattern case retaining only the attributes necessary for the methodological point.

01 – SUMMARY

Executive summary

Research question. When a small or medium-sized business publishes its pricing somewhere on the public web, can AI answer engines retrieve and surface that pricing when asked directly about the business — and when they cannot, why not?

The study froze a population of **1,123 independent US SMBs** (10 business categories × 10 metropolitan areas) by human-audited evidence, classified every business by **where its pricing lives**, and then — after the classification was frozen — asked two AI answer engines the same fixed question about every business, exactly once, with no follow-ups. A parallel set of category-level discovery questions (no business named) was run per city × category cell.

The three pricing-location cohorts (frozen before any engine measurement):

COHORT	DEFINITION	N	SHARE
on-site	pricing verified on the business's own website	647	57.6%
platform-hosted	pricing found only on an external booking/ordering platform	32	2.8%
not-found	no public pricing found by the study's crawl-and-audit process	444	39.5%

Headline result (branded queries — the business asked about by name; "pricing visible" = the engine answered that current pricing for the queried service was visible on the pages it relied on):

COHORT	N	PRICING VISIBLE – GEMINI	PRICING VISIBLE – PERPLEXITY
on-site	647	52.2% (338) · 95% CI [48.4%, 56.1%]	48.4% (313) · CI [44.5%, 52.2%]
platform-hosted	32	37.5% (12) · small n, counts only	40.6% (13) · small n, counts only
not-found	444	27.7% (123) · CI [23.7%, 32.0%]	22.1% (98) · CI [18.5%, 26.2%]

Businesses with verified on-site pricing were surfaced at roughly **twice** the not-found rate (rate ratio **1.9×** Gemini, **2.2×** Perplexity — an unadjusted **association**, not a measured causal effect). Yet even among businesses whose pricing was verified present on their own sites, the engines failed to surface it about **half the time**. And publishing is not *necessary* either: in the not-found cohort, engines surfaced pricing for 123 (Gemini) and 98 (Perplexity) businesses that the study's own crawl-and-audit process could not find — answers assembled from third-party surfaces the businesses never linked.

HEADLINE · BRANDED QUERIES

Publishing roughly doubles the rate — and still fails to surface an answer about half the time.

52%

On-site pricing surfaced
Gemini · 48.4% Perplexity

28%

Not-found pricing surfaced
Gemini · 22.1% Perplexity

~2×

On-site vs not-found
1.9× / 2.2× rate ratio, unadjusted

The engines also disagree structurally: Gemini declined to answer 286 of 1,123 branded queries; Perplexity declined 1. On identical queries they returned different verdicts, cited different sources, and sometimes surfaced different prices. **AI visibility was not a single answer in this study: even with the prompt held constant, the two engines produced materially different views of the same business.**

MAJOR LIMITATIONS, UP FRONT

two engines measured (not the full field); a single measurement day (a snapshot, not a behavioral model); official APIs rather than consumer applications; unadjusted cohort comparisons; a small platform-hosted cohort (n=32) reported as counts; a high Gemini refusal rate that shapes its headline rates; and price-accuracy findings limited to the human-audited subset (Section 11). Full limitations in Section 14.

PART 01

Study design and ground truth

Who was studied, how pricing was located, and why the ground truth was frozen before any engine was asked.



Growfer

02 — DESIGN

Research question and study design

The study separates two things that are routinely conflated:

- **Pricing being publicly available** — a property of the business's public web presence, established by crawling and human audit.
- **An AI engine surfacing that pricing** — a property of the interaction between that public presence and the engine's retrieval-and-generation behavior, established by measurement.

Two question shapes were measured:

- **Branded queries** — one per business: does the engine, asked directly about this business by name, find and surface current pricing for a fixed, category-typical service? This measures retrieval of known information about a known business.
- **Discovery queries** — two per city × category cell, naming no business: which local businesses does the engine recommend for the service, and what do they charge? This measures membership in recommendation sets — a conceptually different question with a different denominator (Section 12).

Why ground truth was frozen first. Defining "visibility" using the same systems being evaluated would make the study circular. The population, its cohort classification, and its human-audited evidence were therefore frozen on **September 22, 2026**, before any engine measurement used in the results was run; no engine output ever influenced any classification decision (enforced structurally in the pipeline: the measurement stage had no write access to any classification artifact). An earlier exploratory precheck (September 2026; 1,496 grounded Gemini calls) informed design only and served as an analysis covariate, never as classification input.

03 — POPULATION

Population construction

- **Source.** Business records were sampled from **Foursquare Places** across 10 categories × 10 US metropolitan areas.
- **Eligibility.** Independent (non-chain) businesses with a discoverable primary website; chain locations were excluded by design so that pricing surfaces reflect individual business decisions rather than corporate templates.
- **Crawl population.** 1,484 businesses with verified-reachable primary websites formed the crawl population.

- **Qualification.** From the crawl population: **305** businesses were excluded on audit because their primary offering did not match the sampled category; **56** businesses with no working primary website were excluded and remain outside every figure in this report (their situation — platform presence without a functioning primary site — is noted qualitatively where relevant). The final frozen study set: **1,123 businesses**.
- **Cities studied:** New York, Los Angeles, Chicago, Houston, Phoenix, Philadelphia, Miami, Atlanta, Denver, Seattle.
- **Categories studied:** restaurants, dentists, hair salons & barbershops, nail salons, medical spas, massage spas, pet groomers, fitness studios & gyms, auto detailing, property management.

04 — COHORTS

Pricing cohorts

Every one of the 1,123 businesses was classified — by crawl evidence plus human audit — into exactly one cohort:

COHORT	N	OPERATIONAL DEFINITION
on-site	647	Quoted pricing found on the business's own website (its own domain).
platform-hosted	32	Quoted pricing found only on an external platform surface (a booking, ordering, or scheduling tenant page), not on the business's own site.
not-found	444	No quoted public pricing found by the study's crawl-and-audit process.

What "not-found" means — and does not mean. It is an operational finding about this study's process: polite crawling of the business's own site and its linked pages, plus human-audited re-checks, found no publicly quoted pricing. It is **not** proof that no pricing exists anywhere on the web. The engines themselves demonstrably found pricing for a substantial subset of these businesses (Section 9) — the not-found cohort is best read as "no public pricing found by the study's crawl-and-audit process," a verified census of *crawler-discoverable* pricing.

Persistence of the split-pricing pattern. Of the 32 platform-hosted businesses, **31 did not replicate the platform pricing onto their own sites** — unchanged when re-audited roughly nine months later; the platform-hosted pricing itself remained live in **30 of 32 cases (94%)**. The typical mechanism is a branded tenant page on the platform's domain: the business's own site markets, the platform prices.

Category texture. Restaurants were the most transparent category (65% on-site) and also the split-pricing category (14 of the 32 platform-hosted businesses, concentrated in the ordering/booking-platform ecosystem). Dentists were the most opaque (61% not-found; 104 of 172),

consistent with consultation-pricing norms in dentistry. Two categories contained zero platform-hosted rows, so category-level platform inference is out of scope; the analysis pools categories.

05 — METHOD

Crawl and ground-truth methodology

Crawl. Each business's primary site was crawled with an LLM-assisted navigation pass (keyword- and link-guided page selection under per-site budgets) plus deterministic pricing-signal detection on rendered pages: quoted-price patterns, pricing/menu/services page structures, and booking-widget presence. Rendered (headless) captures were used throughout. A pricing-platform taxonomy emerged from the crawl evidence itself rather than from any pre-baked platform list: **38 pricing hosts**, 33 of them discovered empirically. Linked external surfaces (booking tenants, ordering platforms, menus) were pursued where the business's own pages referenced them.

Human audit. Machine classifications were treated as drafts, never as truth. A human-audit program then ran across five workstreams: a 100% manual audit of the platform-hosted cohort (all 32 businesses); a 60-business re-check sample of the on-site cohort; a 60-business random sample of the not-found cohort specifically hunting for hidden platform pricing; a targeted 60-business candidate list (robotics-blocked sites and rows with known-platform links); and re-render plus deep-re-pursuit passes for degraded captures and high-platform-prevalence categories. In total, **307 audit judgments over 277 unique businesses** (some businesses were sampled by more than one workstream) produced quoted-evidence records: for each audited business, the specific pages, quotes, and verdicts a human reviewer verified.

Key audit results (verification meta-findings).

- The not-found cohort is real, not a crawling artifact: the 60-business random sample found **0 hidden platform-pricing cases (95% CI 0–6.0%)**; the targeted candidate list converted 2 of 60. Best estimate of hidden-pricing rows across all 444: on the order of ten or fewer.
- The on-site cohort is stable: **59 of 60** re-checks re-confirmed on fresh renders.
- Apparent "pricing withdrawn" cases were mostly capture artifacts: **9 of 11** dissolved when the specific pages carrying the pricing signal (rather than homepages alone) were re-rendered; the 2 genuine withdrawals were thin promotional offers — the least durable evidence class.
- Every machine misfire encountered (mismatched businesses, app-store copy, B2B directories, cookie-documentation pages) was caught at human adjudication; no misclassification entered the frozen set.

Freeze. The adjudicated population — $647 / 32 / 444 = 1,123$ — was frozen on **September 22, 2026**. Engine measurement began only after the freeze, and engine output was structurally barred from feeding back into classification (Section 2). An earlier exploratory precheck had shown why this

discipline matters: of 227 businesses where a grounded AI precheck claimed visible pricing despite no on-site signals, follow-up verification confirmed only **27 (11.9%)**; conversely, **204** businesses with verified quoted on-site pricing had received a "no pricing visible" answer. AI claims about pricing, in both directions, required independent grounding.

06 — MEASUREMENT

Engine measurement

- **Engines and configurations.** Google **Gemini** (`gemini-2.5-flash`, grounded with Google Search, temperature 0) and **Perplexity** (`sonar`, web search size pinned to `medium`). Both measured through their **official APIs**, not consumer applications. ChatGPT and Claude were not measured: every table in this report is a two-engine picture of the four-engine study design, and generalization beyond the measured pair is unsupported (Section 14).
- **Measurement date and windows.** All queries were asked on **September 30, 2026**: Gemini 19:13–21:05 UTC; Perplexity 21:06–22:05 UTC. Single pass per query — no repeat polling, no re-prompting of any kind.
- **Volume.** 1,123 branded queries + 200 discovery queries per engine = **2,646 recorded engine responses, zero failed records**. Retries (on transient errors) were rare (~1% of calls).
- **The exact branded query** (fixed grammar, identical for every business and both engines; one category-typical service per category, listed in Section 15):

```
What does {businessName} ({market}) charge for {verticalService}?
Answer only from {domain} or pages that clearly host its pricing;
cite the URLs you rely on.
End with a final line exactly: VERDICT: yes|no|refused
(yes = current pricing for that service is visible on the pages you relied on).
```

- **Verdicts.** `yes` = current pricing for the queried service visible on the pages the engine relied on; `no` = no such pricing surfaced; `refused` = the engine declined to answer. Verdicts were parsed from the required final line (with a structured-field fallback where the API provided one). **A refusal is a verdict, never re-prompted** — re-prompting would contaminate comparability. Primary rates keep refusals in the denominator; an answered-only sensitivity view is reported separately and labeled as derived (Section 10).
- Because Gemini's grounded search mode does not support strictly structured output, its verdicts were elicited via the fixed final-line instruction and parsed from answer text — one reason raw answer text was preserved for every record.

PART 02

What the engines answered

Surfacing rates, sources, divergence between engines, price accuracy and discovery behaviour.



Growfer

07 — RESULTS

Primary results

The headline table appears in Section 1. Full verdict distributions:

COHORT	VERDICT (YES / NO / REFUSED) – GEMINI	VERDICT (YES / NO / REFUSED) – PERPLEXITY
on-site	338 / 183 / 126	313 / 334 / 0
platform-hosted	12 / 15 / 5	13 / 19 / 0
not-found	123 / 166 / 155	98 / 345 / 1

Relative likelihood (unadjusted). The on-site ÷ not-found visible-rate **rate ratio** is **1.9×** for Gemini (52.2% vs 27.7%) and **2.2×** for Perplexity (48.4% vs 22.1%). These are associations, not measured causal effects: no confounder control (category/city mix) was performed. The strict odds ratios are different, larger statistics ($\approx 2.9\times$ and $\approx 3.3\times$ respectively); the two must not be conflated.

Confidence intervals are Wilson 95% score intervals. The platform-hosted cohort (n=32) is reported as **counts only**: at that sample size an interval is wide enough to mislead, and the cohort's analytic role is mechanism evidence (how platform pricing surfaces behave), not population prevalence. Slices under n=100 carry bare counts throughout.

08 — SOURCES

Source and citation behavior

Every cited URL in every answer was recorded (full citation lists; redirect resolution capped at the first five citations per record — Section 14). Aggregating cited hosts across the census shows where engine answers actually come from:

Perplexity's five most-cited hosts across all 1,323 of its queries: yelp.com (438 citations) · fresha.com (214) · trustanalytica.org (201) · tripadvisor.com (165) ·groupon.com (163). **No booking/ordering-platform tenant host appears in its top 20** (2,230 citations, all review sites, directories, and aggregators).

Gemini's most-cited hosts: google.com (151) · ubereats.com (58) · facebook.com (36) · grubhub.com (34) — and, unlike Perplexity, platform tenant hosts do appear: order.online (18) · toasttab.com (18) · mindbodyonline.com (17) · square.site (8) · zenoti.com (5) — 66 citations, 14.6% of its top-20 volume.

Three readings, each bounded by its evidence:

1. **AI engines routinely construct answers about a business from information the business does not control.** For platform-hosted businesses specifically, the business's own tenant page is a minority source: engines answering "what does this business charge" leaned on third-party summaries of the business more than on the business's own pricing surface — on either domain.
2. **Platforms are not created equal for AI visibility.** Restaurant ordering-platform pages were cited (mostly by Gemini); salon/spa booking-tenant pages were cited far less; one heavily cited booking platform (fresha.com, 214 Perplexity citations) sits outside the crawl-emergent taxonomy entirely — the taxonomy was built from businesses' own link graphs, and engines demonstrably reach surfaces those graphs never contained. A taxonomy's silence about a host is evidence about crawls, not about engines.
3. **The engines' citation behavior did not consistently privilege a business's own page over third-party sources.** Empirically: own-domain citations appear in most answers, yet third-party hosts dominate citation volume (Section 8 aggregates), and negative answers frequently cite the business's own domain alongside — or instead of — the pages that carry its pricing (Section 10; companion analysis). Vendor documentation describes citations as automated retrieval attributions rather than editorial judgments (Google Gemini API grounding documentation, <https://ai.google.dev/gemini-api/docs/google-search> · Perplexity, "How does Perplexity work?", <https://www.perplexity.ai/help-center/en/articles/10352895-how-does-perplexity-work>); beyond those references, this report relies on the empirical statement above.

09 — DIVERGENCE

Crawler-visible versus engine-visible information

The study's sharpest divergence: businesses the crawl-and-audit process classified **not-found** — no public pricing found on their own sites or on the linked external surfaces the process pursued — where an engine nonetheless answered **yes, with citations**:

ENGINE	NOT-FOUND BUSINESSES ANSWERED "YES"	SHARE OF COHORT (N=444)
Gemini	123	27.7%
Perplexity	98	22.1%

Roughly **1 in 5 to 1 in 4** not-found businesses had pricing surfaced by an engine. Anonymous pattern examples from the row-level record:

- A restaurant whose menu lived on a **dynamically rendered ordering-platform surface** (client-side routing that defeated every headless capture attempt during the audit) drew "yes" with five citations from **both** engines.
- A property-management company drew a "yes" from Perplexity citing **ten listing and directory portals** (apartment-listing and real-estate platforms) with three price figures — pricing that exists in the world, attached to the business, on surfaces its own site never links.

"There are three internets." What renders for humans, what renders for polite crawlers, and what engines retrieve and cite are demonstrably different views of the same web. Two consequences, stated carefully: the 444-business not-found count remains a **verified census of crawler-discoverable pricing** (the 0-of-60 hidden-platform audit, Section 5) — and it is simultaneously **not** the count of businesses whose pricing engines cannot see. The 56 excluded businesses with no working primary site sit at the extreme of this gap: platform presence without a primary domain, invisible to both this study's population and, largely, to engine answers.

10 — DISAGREEMENT

Engine disagreement and answer variability

Refusal behavior differs fundamentally. Gemini declined 286 of 1,123 branded queries (25.5%); Perplexity declined 1. Primary rates (Section 7) include refusals in the denominator — a declined answer is a real consumer outcome. The labeled secondary view, **answered-only sensitivity** (visible ÷ answered), shows what the rates would be if refusals were excluded:

COHORT	GEMINI ANSWERED-ONLY (CHANGE VS PRIMARY)	PERPLEXITY ANSWERED-ONLY (CHANGE)
on-site	64.9% (+12.6pp)	48.4% (+0.0pp)
platform-hosted	44.4% (+6.9pp)	40.6% (+0.0pp)
not-found	42.6% (+14.9pp)	22.1% (+0.0pp)

Among queries Gemini answered, it surfaced pricing at rates comparable to or above Perplexity's: **the primary-rate gap between these engines is substantially a refusal-behavior gap, not solely a retrieval gap.** Which layer matters more for a consumer is an interpretation this study flags but does not settle.

These figures do not rank engine quality. Different refusal policies, retrieval stacks, and citation behaviors make "better at finding pricing" claims underdetermined by this design. What the study establishes is narrower and more useful: **engine behavior differs materially under identical queries** —

one engine may answer less often but surface pricing more often when it answers; another answers nearly everything and surfaces pricing less often.

Same business, different answers. Anonymous pattern cases from the record:

- A platform-hosted **medical spa** with non-identical prices across two of its external surfaces: one engine answered **yes**, citing the business's own site *and* its booking-platform tenant page and quoting ten service prices; the other answered **no** while citing twelve hosts — including the business's own site and shop pages — and quoting three prices. Same business, same day, divergent surfaces, opposite verdicts.
- A large **public market venue**: one engine refused; the other answered no, citing twelve third-party hosts (a newspaper, travel platforms, review sites) and a single price figure.
- A **pet-care service** in the not-found cohort: both engines said no; one cited only the business's own domain, the other cited fourteen lookalike and directory hosts and quoted nothing.

When a business's pricing exists on multiple surfaces that disagree, **which price an AI answer carries is engine-specific and unpredictable from the business's side**. Consistency across surfaces behaves like a visibility property, not a bookkeeping nicety.

11 — ACCURACY

Price-accuracy subset

Where human auditors recorded **quoted prices** (not merely a verdict), engine-quoted prices were compared against them with a fixed five-way vocabulary, computed offline with zero normalization: **exact-match** (the engine's quoted price appears verbatim in the audited quotes) · **numeric-close** (within $\pm 20\%$, *flagged*, *never equated*) · **mismatch** (prices quoted, none match) · **unverifiable** (no audited quotes exist for the business) · **no-claim** (the engine quoted no pricing).

The powering caveat comes first. Audited price quotes exist for only a subset:

COHORT	ROWS	HUMAN-AUDITED RECORDS	RECORDS WITH QUOTED PRICES
on-site	647	77	77 (11.9%)
platform-hosted	32	32	32 (100%)
not-found	444	168	7 (1.6%)

161 of the 168 not-found audit records are **absence audits** — they verify the cohort assignment (no public pricing to quote) and cannot power accuracy claims. Accuracy findings are therefore reported

over quoted subsets only, and **this study makes no overall "AI price-accuracy rate" claim for the full population.**

Distributions:

COHORT	ENGINE	EXACT	CLOSE (±20%, FLAGGED)	MISMATCH	NO-CLAIM	UNVERIFIABLE
on-site	Gemini	37	5	12	23	570
on-site	Perplexity	33	8	16	20	570
platform-hosted	Gemini	9	5	5	13	0
platform-hosted	Perplexity	6	7	8	11	0
not-found	Gemini	0	0	5	102	337
not-found	Perplexity	0	1	3	82	358

What the powered slices support:

- **Platform-hosted (the only fully verifiable cohort, n=32):** where engines quoted prices for these businesses, they were right-or-close more often than wrong — Gemini 9 exact + 5 close vs 5 mismatch; Perplexity 6 + 7 vs 8. Small n; counts, not rates.
- **On-site quoted subset (n=77):** exact-match dominates when the business publishes its own pricing (Gemini 37, Perplexity 33) — and mismatch is non-trivial even there (12–16 businesses).
- **Not-found:** effectively unpowered (≤7 quoted rows). The handful of mismatch/close rows here (5 Gemini, 4 Perplexity) are the Section 9 phenomenon inside the accuracy data: engines quoting prices that nothing in the study's rendered evidence contained.

12 – DISCOVERY

Discovery-query findings

Two fixed discovery phrasings per city × category cell (100 cells × 2 phrasings × 2 engines = **400 queries**), naming no business — one service-seeking ("which local businesses can I choose from, and what do they charge?"), one price-seeking ("what does this service typically cost at local businesses?"). Exact templates in Section 15.

Results. Known businesses from the frozen study population were mentioned, with pricing context, **92 times (Gemini)** and **43 times (Perplexity)** across all discovery answers — with near-zero overlap between the engines' top-mentioned businesses. In one illustrative cell containing 20 known businesses, each discovery query surfaced at most one of them. Mention matching is mechanical (name/domain substring with the matched span recorded); a 7-sample human spot-check found 7 of 7 spans were genuine in-context business mentions. Discovery answers were dominated by businesses **outside** the study population and by aggregator-sourced content — consistent with Section 8: the engines' discovery answer space is a third-party-surface space.

Why this is not an AI-visibility score. Branded and discovery queries answer fundamentally different questions. A branded query asks: *can the engine retrieve known information about a known business?* A discovery query asks: *out of every business the engine could possibly recommend, does it happen to choose one from our sample?* The denominators are not comparable — low mention counts do not, by themselves, mean the sampled businesses have "poor AI visibility." What determines whether a small business enters an AI recommendation set is a separate research question requiring its own experimental design.

PART 03

Interpretation and limits

What the record supports, what it does not, and how to reproduce it.



Growfer

13 – INTERPRETATION

Interpretation

The central conclusion, bounded to the measured record: **AI answers about a business reflect a broader public information environment, not simply the business's own website.** The engines in this study drew on review sites, directories, aggregators, listing portals, ordering and booking platforms — and on the business's own pages with varying reliability. The study's most defensible findings, ranked by how well they survive skeptical scrutiny:

1. Publishing pricing on your website roughly doubles the likelihood that these AI engines can surface it — but they still failed to surface a price answer about half the time (association, not measured causation).
2. AI engines routinely obtain business pricing from third-party sources rather than the business itself.
3. About one quarter of businesses for which the crawl-and-audit process found no pricing nevertheless had pricing surfaced by an AI engine.
4. Different engines materially disagree about whether pricing exists, where it comes from, and sometimes what it costs.
5. Therefore **AI visibility did not behave like a fixed property of a website, or even of a business.** Under the conditions tested, it behaved like an interaction between a business's broader public web presence and an engine's retrieval behavior: the business has an information environment; the engine has a retrieval-and-generation process; the answer is the interaction of the two — and because either side can vary, the answer varies.

Holding the query constant did not produce a stable business representation across engines. Any single AI response is best understood as **a sample from a variable retrieval environment, not an authoritative visibility state** — which is why a point-in-time "AI visibility score" is inherently conditional on the measurement setup (engines chosen, phrasings, refusal handling, retrieval paths, measurement window), and why adding measurements can add false precision rather than stability.

A useful definition of AI visibility emerges from the record: not "can a chatbot find my website?" but **"what information will an AI system present about my business, where does that information come from, and is it correct?"** — spanning the business's own site, marketplaces, review sites, ordering and booking systems, and aggregators.

None of these conclusions requires any claim about revenue, conversion, or consumer behavior; this study measured no economic outcomes. An **exploratory companion analysis** — *"On-site pricing not surfaced as an answer"* — examines why published pricing often fails to be surfaced (retrieval vs extraction vs semantic matching). Its findings are labeled exploratory and are deliberately **not** part of this report's audited result set; readers are pointed there for that analysis.

14 — LIMITATIONS

Limitations

Stated once, applying everywhere:

- **Two engines of four.** ChatGPT and Claude were not measured. Cross-engine generalizations beyond the measured pair are unsupported.
- **Single measurement window.** All answers were collected on September 30, 2026 (Gemini 19:13–21:05 UTC; Perplexity 21:06–22:05 UTC). No repeat polling: engine answers are a snapshot, not a behavioral model. Temporal drift versus the ground-truth window (September 22, 2026) is a covariate.
- **API, not consumer applications.** Both engines were measured through official APIs with fixed configurations. Consumer-app behavior (personalization, UI-mediated follow-ups, different retrieval settings) may differ and was not measured.
- **Unadjusted cohort comparisons.** The $1.9\times/2.2\times$ rate ratios carry no confounder control (category/city mix) and support no causal claim; strict odds ratios ($\approx 2.9\times/3.3\times$) are different, larger statistics.
- **Category/city imbalance within cohorts.** Cohort composition varies by category and city (e.g., the platform-hosted stratum is thin or empty in some categories); category-level platform inference is out of scope and the analysis pools categories.
- **Small platform-hosted cohort (n=32).** Reported as counts; mechanism evidence, not population prevalence.
- **Gemini's refusal rate (286 of 1,123).** Primary rates include refusals; the answered-only view is labeled derived and never headlined. The two engines' rates are not like-for-like without that context.
- **Limited price-accuracy ground truth.** Quoted-price audits cover 77 on-site, 32 platform-hosted, and 7 not-found businesses; no overall accuracy rate exists or should be quoted (Section 11).
- **One service query per business.** Each branded query asked about a single category-typical service (Section 15). Businesses offering that service under different naming, or not at all, could be answered "no" correctly — the query measures answerability of one fixed question, not the business's full pricing surface.
- **Label and market mismatches.** A small share of answers referenced a different metro than queried ($\sim 4\%$ of Gemini "no" answers; none of Perplexity's), and some businesses' category labels did not match their actual offerings; both patterns are quantified in the companion analysis. Cohort classification itself was unaffected (pricing location, not label accuracy, defined cohorts).

- **Citation resolution capped at five per record** (network-politeness budget). Full citation *lists* were recorded, so host aggregations are complete; redirect-resolution coverage is top-5 only.
- **Platform-hosted prevalence is a systematic undercount.** Dynamically rendered platform surfaces (client-side ordering/menu routes) defeat polite headless capture; Section 9's 123/98 is the engine-side lower bound on the same phenomenon.
- **Discovery mention matching is mechanical** (substring + span), spot-checked on 7 samples (7/7 genuine); single-token business names carry over-match risk that was not exhaustively audited.
- **Tenant URLs are not durable.** A platform's URL restructure between the audit and measurement windows broke at least one recorded tenant link; row-level citation hosts are window-bound.
- **No economic measurement.** No revenue, conversion, booking, or customer-behavior outcome was measured or is implied.

15 — APPENDIX

Reproducibility and methodology appendix

Cities (10): New York · Los Angeles · Chicago · Houston · Phoenix · Philadelphia · Miami · Atlanta · Denver · Seattle.

Categories (10) and the fixed service queried per category (the branded query's `{verticalService}`):

CATEGORY	QUERIED SERVICE
Restaurants	a dinner main or tasting menu
Dentists	a cleaning or a crown
Hair salons & barbershops	a haircut or color
Nail salons	a gel manicure
Medical spas	a facial or injectable unit
Massage spas	a 60-minute massage
Pet groomers	a full groom
Fitness studios & gyms	a monthly membership

CATEGORY	QUERIED SERVICE
Auto detailing	a full detail
Property management	monthly rent for a listed unit

Branded query template (verbatim; `{market}` is the metro key, e.g. "miami"; `{domain}` the business's primary domain):

```
What does {businessName} ({market}) charge for {verticalService}?
Answer only from {domain} or pages that clearly host its pricing;
cite the URLs you rely on.
End with a final line exactly: VERDICT: yes|no|refused
(yes = current pricing for that service is visible on the pages you relied on).
```

Discovery query templates (verbatim; `{marketLabel}` is the display city name, e.g. "Miami"):

```
U1 (service-seeking):
I'm looking for {verticalService} in {marketLabel}. Which local businesses
can I choose from, and what do they charge? Cite your sources.

U2 (price-seeking):
What does {verticalService} typically cost at local businesses in
{marketLabel}? Name businesses with published pricing and cite the URLs.
```

Cohort definitions. on-site = quoted pricing verified on the business's own domain · platform-hosted = quoted pricing verified only on an external platform tenant surface · not-found = no quoted public pricing found by the crawl-and-audit process (operational; not proof of absence from the web).

Verdict definitions. `yes` / `no` / `refused` per the final-line instruction above; parsed from answer text with a structured-field fallback where the API provided one; refusals never re-prompted; verdicts recorded per response with the verbatim query, timestamp, full citation list, and raw usage.

Measurement protocol. Official APIs; Gemini `gemini-2.5-flash` grounded with Google Search (temperature 0); Perplexity `sonar` with web search size `medium`; single pass; one response per query per engine; retries ≤ 2 on transient errors with backoff; per-engine request budgets with hard fuses; resumable per-response records (interruptions cost no duplicate calls); engines run serially, within-engine request concurrency ≤ 4 .

Confidence-interval method. Wilson 95% score intervals ($z = 1.96$) for cohort-level proportions with $n \geq 100$; bare counts for smaller slices; the $n=32$ platform-hosted cohort reported as counts only. Derived

statistics (rate ratios) computed offline, byte-deterministically, and cross-checked against the primary tables (derived figures must match, never contradict, the primary artifacts).

Audit procedure. Machine-classified drafts → five human-audit workstreams (Section 5) → quoted-evidence records per audited business → adjudication with evidence requirements (class changes unsupported by quoted evidence were rejected by construction) → freeze. Population frozen before measurement; measurement structurally barred from writing to classification artifacts.

Methodological lessons for future studies (generalized from this study's conduct): validate dry-run selectors against real data before any live run; make measurement records resumable so interruptions are free; audit every derived table against the primary artifacts it summarizes (a derived table once labeled "audited" coverage where the underlying records supported only "quoted" coverage — caught before any conclusion consumed it); keep full source lists even when resolution budgets cap follow-up fetches; and separate *audited coverage* from *quoted coverage* explicitly in every accuracy claim.

Economics. Cash outlay was \$10.90 — the metered Perplexity leg of the census (1,323 calls: ~91k input + ~229k output tokens plus 1,323 searches at the medium search tier, at published API pricing). The Gemini leg ran on provider credits, so **\$10.90 is not the unsubsidized cost of the full two-engine run.** Even with both legs priced at published rates, measurement at this scale is economical enough that repeated, scheduled measurement is practical rather than exceptional.

Growfer Research — public research edition. The internal record preserves full row-level lineage, per-response raw records, and the complete revision and review history; aggregate findings, methodology, and conclusions above are unaltered from the cited revision.



CLOSING THOUGHT

**In this study, AI visibility
behaved like an interaction,
not a fixed property.**

NEXT STEP

**Read the companion
analysis.**

Why published pricing often fails to surface: retrieval,
extraction or semantic matching.

[Exploratory companion · Rev 2](#)

1,123 2,646

businesses frozen
before
measurement

engine responses,
zero failed