At least 85% of restaurants are **invisible to AI.**

We measured an entire market to find out what separates the ones that aren't.

4,776
cafés, restaurants and bars — every venue in two Bali markets

2,208
questions asked, the way real customers ask

26,993
sources the AI systems cited — every one logged and analyzed

1.4M
venue-level exposure opportunities in the statistical model

Based on: Pitenin, V. (2026). *Invisible to the Machine: Auditing AI Restaurant, Café, and Bar Recommendation Against a Complete Market Census.* arXiv:2608.07069 — arxiv.org/abs/2608.07069. Every figure in this report is traceable to that paper; nothing here goes beyond what the study measured.

EXECUTIVE SUMMARY

Five things we now know that nobody knew before

We mapped every food-and-drink venue in two of Bali's busiest markets, asked four AI assistants 2,208 realistic customer questions over seven days, and traced all 12,439 venues they named back to the map. It is, to our knowledge, the first complete-market audit of AI dining recommendations ever run — and the results rearrange several beliefs the industry treats as obvious.

01 **Invisibility is the market's default state. 85.6% of venues were never recommended once** — by any system, in any answer, all week. Even among established venues with 50+ Google reviews, 72.6% never appeared. And 85.6% is the floor: counting the small, uncategorized venues no register can fully see, the true figure may run as high as 92%. → Section 01

02 **AI recommends in two stages, and each stage runs on different fuel.** Getting *into* an answer is driven by documentation — website (**1.9×**), review volume (**1.6×**), listed prices, web mentions — and fresh recent reviews lean the same way. Your star rating does nothing at that door. It starts working only at stage two: which of the admitted venues gets named *first*. → Section 02

03 **The famous AI problem — making things up — is not the real risk. Staleness is.** In 12,439 venue mentions we found evidence of just **one** likely-invented venue (0.08%). But the systems recommended permanently closed venues **93 times**. AI doesn't invent restaurants; it remembers ones that no longer exist. → Section 03

04 **There is no single "AI ranking."** Any two systems agree on only **33–54%** of their top-20 venues; just 8 venues made all four lists. And every system's answers change from one asking to the next — so a one-off "does AI recommend me?" check measures luck, not visibility. → Section 04

05 **Your own website can outperform the giant platforms.** AI reads TripAdvisor — but it reads everything else too, and venue-owned content competes on equal terms. Of 26,993 logged citations, the single most-cited source in the whole study was not a platform: it was one venue's own website. Businesses that write about themselves get read. → Section 05

> **Every number in this report comes from the published, pre-registered study** — and every headline conclusion survived a battery of stress tests: two independent statistical models, 500 bootstrap re-runs, and three sensitivity re-analyses. The method callouts throughout explain how, in plain language.

THE STUDY

We counted everyone first — the step every other study skips

To say "85% are invisible," you must first know every venue that exists. Most AI-visibility studies skip this step — they check a list of famous brands and report who ranks higher. We went the other way, and that decision is what makes every number in this report possible.

1.4M	12,439	26,993	7+14
venue-level exposure opportunities in the primary statistical model	AI venue mentions, each one verified against the map	sources the AIs cited — we logged every single one	days of repeated asking, plus a re-test two weeks later

The census. Every café, restaurant and bar listed in greater Canggu and greater Ubud — **4,776 venues**, each with its public profile: rating, review count, website, price level, hours. This register is what makes the word *never* measurable.

The questions. Real people don't type "best restaurant Canggu." They ask with a budget, a companion, a dietary need, a laptop that needs wi-fi. We asked as eight kinds of guests, six phrasings each, in both areas — 96 fixed questions, locked before the study began and published verbatim in the paper.

The digital nomad
wi-fi, outlets, and tolerance for a three-hour flat white

The couple
a date night that has to land

The business host
a table that won't embarrass anyone

The budget backpacker
the best cheap warung, honestly priced

The family
kids, prams, no patience for waits

The vegan
done with menus that offer one sad salad

The coffee connoisseur
knows what a flat white should taste like

The late-night group
food, drinks, and a reason to stay out

The systems. ChatGPT, Claude, Gemini and Perplexity, each in its live, search-connected form — 2,208 answers across seven days, plus a re-test two weeks later to check the picture holds. It does.

HOW BIAS WAS DESIGNED OUT

Before collecting the main data, we wrote down publicly what we would test and how — a practice from clinical research called **pre-registration**. It makes moving the goalposts impossible: the hypotheses, the questions, and the analysis plan were frozen first, results came second. Extraction and name-matching were validated against independent annotation with published error rates (mention-level precision 97.6%, recall 99.1%). Exploratory trial runs were kept strictly separate; no number in this report touches them.

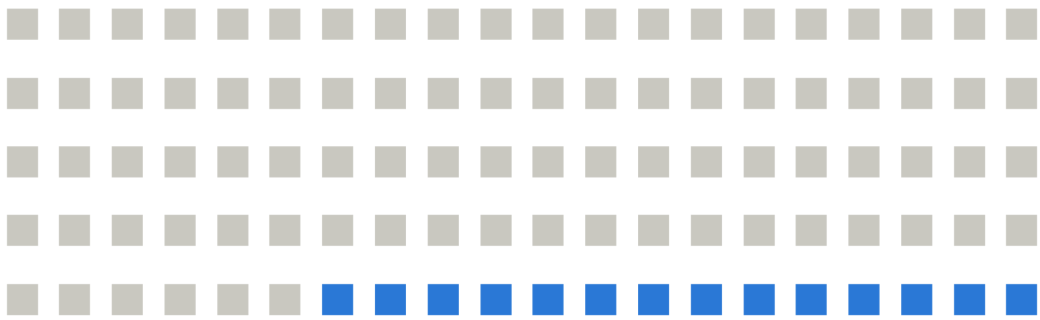
One more thing, said plainly: Norly builds visibility software for hospitality — this question is our day job, which is exactly why we answered it with science instead of a marketing post. Vendors publish claims; we published evidence, on arXiv, where any researcher can take it apart — including the two places the data proved *us* wrong (Section 06).

Most of the market doesn't exist to AI

Across 2,208 answers, the four systems produced 9,791 recommendations — and kept landing on the same small slice of the market. Only 689 of 4,776 venues were ever recommended at all.

Out of every 100 venues on the map, about 86 were never once recommended

4,087 of 4,776 venues, across 2,208 questions to four AI systems • gray = never recommended



Each square = 1% of the census. Source: study Fig. 1; data: chart_data/invisibility_storefronts.csv.

The obvious objection: maybe the invisible venues are marginal — stalls, ghosts, places nobody would recommend anyway. We checked. Among **established venues with at least fifty Google reviews** — real businesses with real customer bases — **72.6% were still never recommended.** Invisibility is not a quality filter. It swallows most of the legitimate market too.

The wall is at the door, not at the top: the most-recommended venue holds just 1.9% of all recommendations.

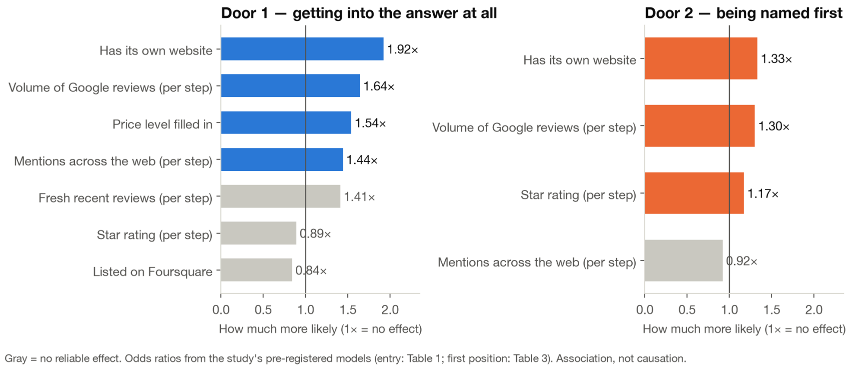
Among the visible 689 there is no cartel — the top five venues together hold under 8% of recommendations. Getting recommended *at all* is the scarce event. That is bad news for the invisible majority and genuinely good news for any single venue: there is no incumbent monopoly to displace. The question is what opens the door — which is Section 02.

> **WHY "AT LEAST" 85%**
>
> No census of a living market is perfect, so we attacked our own: we cross-checked the register against an independent venue dataset, hand-audited the gap, and used the AIs themselves as an adversarial probe — any venue they could name that we had missed would show up as an unmatched mention. The result: venues absent from our census are almost certainly never-recommended too, so every counting error makes the true invisibility rate *higher* than 85.6%, not lower. Under the broadest defensible market definition it exceeds 92%.

Paper §5.2, Figures 1–2, Appendix B. Established-venues denominator: 2,173 venues with ≥50 ratings.

Documentation gets you in. Rating gets you named first.

This is the study's central discovery, and we haven't seen it measured anywhere before: AI recommendation works in two stages — and each stage rewards different factors.



Pre-registered models; entry = study Table 1, first position = Table 3. Data: chart_data/two_doors_factors.csv.

Stage one — getting into the answer. The systems first have to *find* you in the sources they read. Here, everything that matters is documentation: an own website (1.9× the odds), review volume (1.6× per step), a filled-in price level (1.5×), mentions across the web (1.4×). Your star rating? Statistically invisible at this stage — a 4.9-star venue with a thin footprint loses to a 4.3-star venue with a thick one.

Stage two — being named first. Once a venue is inside the answer, the logic flips. Now rating is the strongest quality signal (1.17× per step), review volume still helps, and raw web presence stops mattering. The model composes its shortlist from what it retrieved — and orders it by merit signals.

Your rating decides where you rank. Your paper trail decides whether you exist.

What this means for your business: the industry's default obsession — pushing the star rating from 4.5 to 4.7 — targets stage two. But most venues never reach stage two. If AI doesn't name you, the first suspect is not your quality; it's your documentation: no website, sparse profile, few reviews, no trace in blogs and guides. All fixable — and unlike your rating, fully under your control.

READING THE NUMBERS

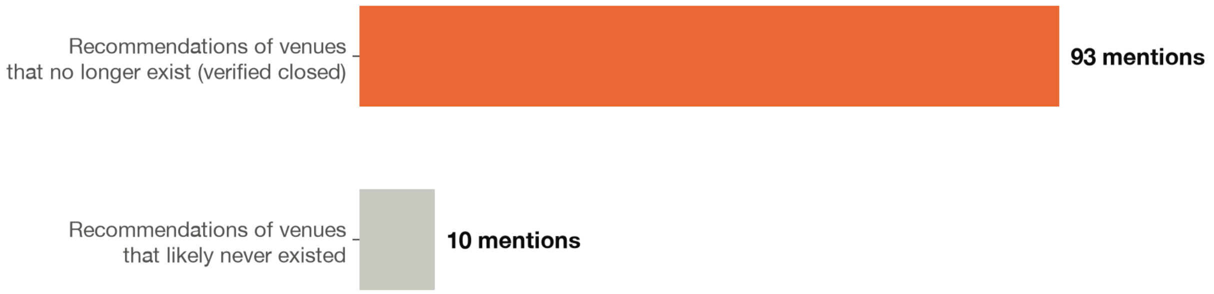
"1.9×" is an **odds ratio**: venues with their own website had nearly twice the odds of being recommended in a given answer, holding rating, review volume and the other factors constant. Estimates come from a regression over 1.4 million venue–question opportunities, corrected for testing multiple factors at once; 1.0 means no effect, and gray bars could not be distinguished from it — strong, robust associations, not experiments.

Paper §5.3–5.4, Tables 1 & 3, Figure 4. Robustness: two alternative models and three sensitivity refits agree on every conclusion.

AI almost never invents restaurants. It remembers dead ones.

Ask anyone what's wrong with AI recommendations and you'll hear one word: hallucinations. We went looking for them — every single name the systems produced, checked against the real world. The myth did not survive contact with the data.

The AI recommendation problem nobody talks about is staleness, not invention



Study §5.8, Figure 10. Data: chart_data/staleness_vs_fabrication.csv.

Out of 12,439 venue mentions, exactly **one name** — ten mentions, 0.08% — survived our scrutiny as likely invented. That's the hallucination problem: nearly nonexistent. But the systems recommended venues that have **permanently closed** 93 times, across 14 confirmed-dead establishments. A traveler following those answers arrives at a locked door.

The systems aren't lying. Their sources are out of date — and the same digital trail that creates visibility keeps it alive after the venue dies.

Why does this happen? Because of Section 02's logic. A closed café's reviews, blog mentions and listicle entries don't vanish when the doors shut — and those are exactly the signals that get a venue into answers. The documentation trail outlives the business. For the industry this reverses a priority: the practical quality problem in AI dining recommendations is not fact-checking the AI. It's the freshness of the web the AI reads.

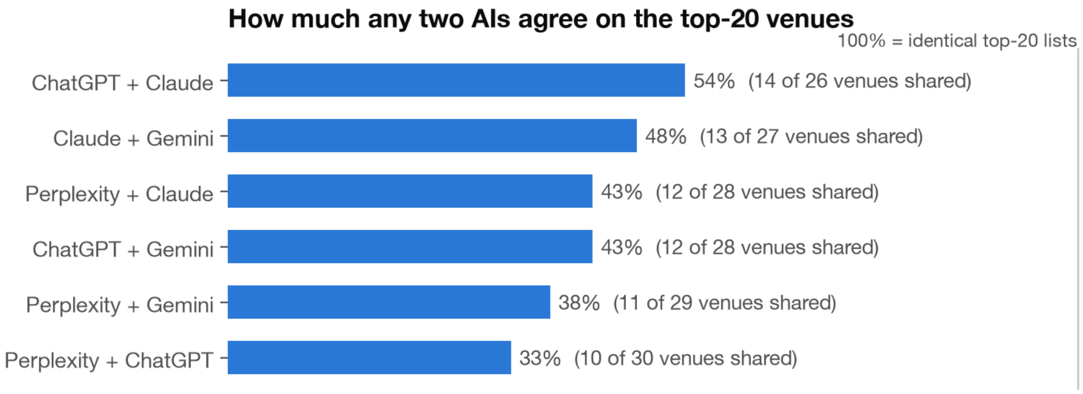
HOW WE VERIFIED THIS

Every AI-named venue that didn't match our census went through a per-name, web-verified audit with a conservative evidence bar: name variants and rebrands were recovered (941 mentions traced back to registered venues), closures were confirmed against public signals, and a name was labeled "likely invented" only when searches produced no trace of it ever existing. The full taxonomy, with error rates, is in the paper.

Paper §5.8, Figure 10. The one likely-invented name, for the curious: "La Blossom."

There is no single "AI ranking" — and one-off checks measure luck

If AI recommendations were one league table, the four systems would name roughly the same venues. They don't. Not even close.



Top-20 most-recommended venues per system, pairwise overlap. Study §5.6, Figure 6. Data: chart_data/cross_ai_agreement.csv.

Of the 37 venues that appear in any system's top-20, only **8 made all four lists** — and 15 appear on exactly one system's list. Being ChatGPT-famous does not make you Gemini-famous. Each system reads a different slice of the web, weighs it differently, and produces its own market.

It gets more interesting: even *the same system* answers differently each time you ask. Identical repeated questions returned venue lists that overlap only 22–45%. But — and this was one of our favorite results — when we re-ran the study two weeks later, the overall picture was just as stable as it was day-to-day. The churn is noise around a stable signal, not a trend.

Visibility is a stable property, measured through a noisy channel. A single spot-check tells you almost nothing.

What this means for your business: if you ask ChatGPT about your venue once and celebrate (or panic), you've measured a coin flip. Real visibility measurement needs repetition, multiple phrasings and multiple systems — the exact protocol this study used. And any vendor selling you a "your AI rank" screenshot from one query is selling you noise.

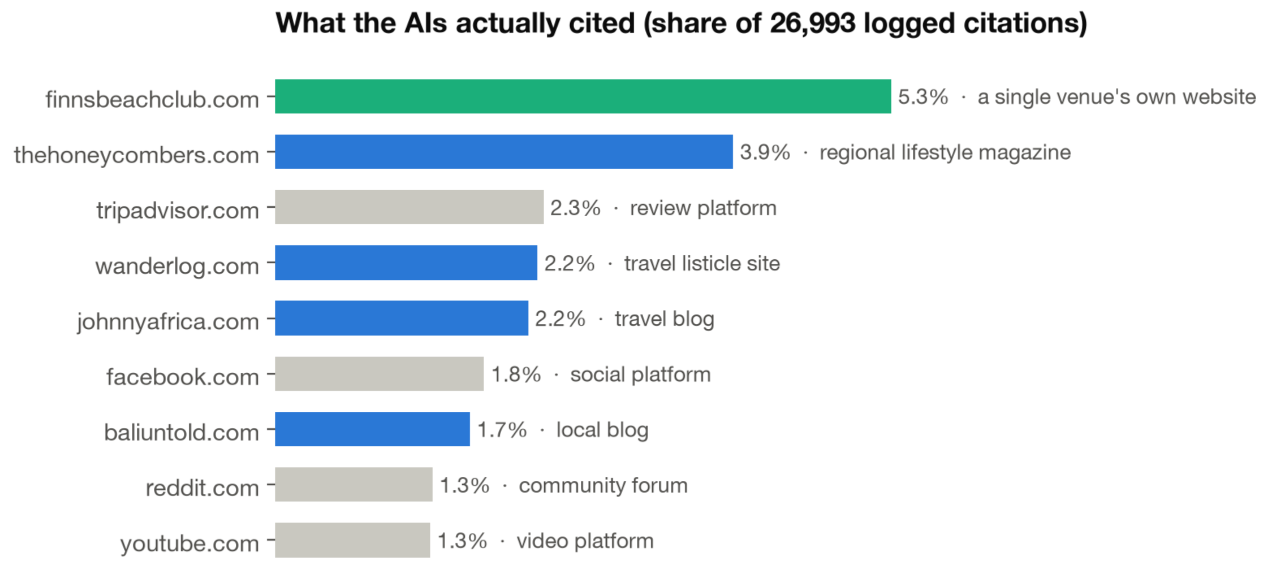
THE RE-TEST

Two weeks after the main study we re-ran a fixed slice of it — same questions, same systems, same pipeline. Answer-to-answer overlap across the two-week gap was statistically the same as overlap within a single week: time added no extra instability. That is what lets us call visibility a persistent property rather than a snapshot.

Paper §5.5–5.6, Figures 5–8, Appendix A.

The open web outweighs the platforms

Every time a system answered, we logged what it cited — 26,993 citations across 986 different websites. The reading list is not what most owners expect.



Share of all logged citations, pooled across the four systems. Study §5.7, Figure 9. Data: chart_data/what_ai_reads.csv.

TripAdvisor is there — but it's **third**, at 2.3%. Ahead of it: a regional lifestyle magazine, and in first place, **a single venue's own website**, cited more than any platform in the entire study. Around them: travel blogs, listicle sites, community forums. The AIs graze the open web, and venue-owned content competes there on equal terms.

The most-cited source in the study wasn't a platform. It was one venue's own website.

This closes the loop with Section 02: the reason an own website nearly doubles your odds of being recommended is sitting right here in the citation logs. The systems don't just check that your site exists — they read it, and they cite it. Every page you publish is a document AI can retrieve; every guide or blog that mentions you is another. Venues with rich trails appear in many sources; venues without appear in none — and Section 01 shows where that ends.

Paper §5.7, Figure 9. One measurement caveat for Gemini's citations is documented in the paper.

The data proved us wrong twice. We're publishing both.

Findings you can trust come from studies willing to report the results they didn't want. We have two — and they're among the most instructive things we learned.

The bet we lost: Foursquare

We went in personally convinced that presence in open point-of-interest databases — the structured datasets many AI products ingest — was a hidden visibility lever. It's a popular theory in the GEO world, and we designed part of the study specifically to test it. The result: **nothing.** Not presence, not data quality, not tips, not popularity scores — no positive effect at either stage, once ordinary documentation is accounted for. In early, messier data the effect had looked strongly positive; better name-matching erased it entirely. Which taught us a second lesson worth publishing: **sloppy matching manufactures findings** — one reason to distrust quick-and-dirty visibility studies, including the one we almost accidentally ran ourselves.

The number that flattered us — and was fake

One variable ("opening hours listed") showed a strong *negative* link to visibility. Strange — until we traced it to our own data collection: a subset of venues we added to the census through a different route was missing the hours field by construction. The variable was measuring how venues entered our register, not reality. We excluded it and documented the trap, because every audit of AI systems is exposed to it, and almost nobody checks.

The data said no — and that "no" is one of the study's most useful results.

HOW WE STRESS-TESTED EVERY FINDING

The same discipline ran through the whole study. The headline results come from a **main model** — a regression over 1.4 million venue–question opportunities that weighs all factors simultaneously, so no factor gets credit that belongs to another. We then rebuilt the conclusions three more ways: a **second, independent statistical model** with different assumptions agreed with the main one on all nine factors; a **cluster bootstrap** re-ran the analysis 500 times on resampled question groups, and the estimates barely moved; and three **sensitivity re-analyses** — dropping the noisiest AI arm, keeping only the highest-confidence name matches, switching the outcome definition — changed no conclusion. When a result survives all of that, it isn't an artifact of one modeling choice. That is the standard this report's numbers are held to.

Paper §5.3, §6.2, §6.5. Both episodes are reported in full in the published study.

SECTION 07 · IMPLICATIONS

What the visible venues have in common

Ranked by the strength of the association in our models, the venues AI recommends share a profile. We state it as what it is — the measured anatomy of visibility, not a guaranteed recipe.

- **They have their own website.** The single strongest entry-stage factor (1.9×) — and Section 05 shows why: the systems read and cite venue sites directly.
- **They accumulate reviews continuously.** Volume beats score at the entry stage (1.6× per step); fresh recent reviews lean positive too. The habit that builds volume — asking every happy guest — is also the one that eventually lifts the rating, which pays off at stage two.
- **Their profiles are complete.** Price level filled in: 1.5×. Mundane, controllable, and associated with getting through the door.
- **They exist beyond the platforms.** Mentions across blogs, guides and local media: 1.4× per step — and at the ranking stage, coverage in food-and-travel editorial specifically is what correlates with being named first.
- **They don't obsess over the decimal.** Rating matters — but only after retrieval, and volume matters more, sooner. 4.6 with 900 reviews beats 4.9 with 40, twice over.

THE FRESHNESS FACTOR

Review freshness deserves its own note. At 1.41× per step, the share of recent reviews just missed the study's deliberately strict statistical bar ($p = .054$ against a .05 threshold, after correction for testing nine factors at once) while clearing the conventional uncorrected one. We report it exactly as the paper does — suggestive rather than confirmed — but the direction and size make it a poor factor to ignore: a steady stream of new reviews keeps a venue's documentation alive, which is precisely the currency of stage one. All factors here are associations measured under controls; the intervention experiment that turns them into a recipe is our next study.

Visibility, it turns out, is not luck. It's a system — and now it's a measured one.

We ran this study for a market. We run it for venues, too.

Norly builds visibility software for hospitality — review management and AI visibility for the people who run the room. This research is the foundation the product stands on: we wanted our advice to rest on measurement, not folklore. Now it does, and it's public.

If you want to know where **your venue** stands — whether AI mentions you, on which systems, against which documentation gaps — we offer an **AI visibility research** for individual venues — built on the same multi-system, repeated-measurement protocol as this study (because, as Section 04 showed, single checks measure noise). And if you'd rather just talk it through first, a conversation costs nothing.

dino@norly.co · norly.co

What's next for the research

This study covered one region and one language, deliberately — a bounded market we could census completely. Replication in a structurally different market is next on the roadmap, along with the intervention study the correlation caveat calls for. The research continues; so will the publishing.

Cite this work

Pitenin, V. (2026). *Invisible to the Machine: Auditing AI Restaurant, Café, and Bar Recommendation Against a Complete Market Census.* arXiv:2608.07069 (arxiv.org/abs/2608.07069). Study design pre-registered; full method, all 96 queries, figures and derived data published with the paper. This report quotes only published numbers; charts carry their source data in machine-readable form.