RESEARCH · VERSION 1.1

# What AI answer engines cite when Malaysian SMEs ask for marketing help

## A two-window audit of Gemini and ChatGPT, September and October 2026

Yik Jinn, Head of SEO and AI Search, ALIQ Studio

Reviewed by Ricky Wong, Co-founder and CEO

Published by ALIQ Studio · 6 October 2026

ALIQ Studio Research

**Read online and download the data**

https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/

**Reuse**

This paper and its dataset are published under the Creative Commons Attribution 4.0 International licence (CC BY 4.0). You may copy, share, quote, adapt and build on them for any purpose, including commercially, as long as you give appropriate credit, link to the licence and say if you changed anything, without suggesting that ALIQ Studio endorses you or your use. The licence does not restrict uses that the law already allows without permission. The licence: https://creativecommons.org/licenses/by/4.0/

Suggested credit: Yik Jinn, ALIQ Studio (2026). What AI answer engines cite when Malaysian SMEs ask for marketing help, v1.1. https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/. CC BY 4.0.

## Executive summary

**Research question:** When a Malaysian SME asks an AI assistant who to hire for marketing, what sources does the answer cite?

**Main answer:** Mostly pages that marketing providers publish on their own websites. In ALIQ Studio's audit of 42 English marketing-services questions asked through the Gemini and ChatGPT APIs between 26 September 2026 and 4 October 2026, pages owned by agencies and freelancers made up 87.4% of Gemini's citations and 59.9% of ChatGPT's. For questions about whom to hire, the shares were 86.1% and 84.8% (an exploratory split, Table 1). The study's first working hypothesis, that third-party pages would outnumber providers' own pages, did not hold on either engine. These are questions written for the study and asked in two windows a week apart; the results are not a measurement of the consumer apps or of every question businesses ask.

**Evidence base:** 42 English questions about 7 kinds of marketing service in Malaysia, asking whom to hire, how a service works or what it costs, each put to Gemini and to ChatGPT 3 times in each of two windows a week apart (26 September 2026 to 27 September 2026, and 4 October 2026). That gave 252 answers from each engine, with 1,513 citations from Gemini and 932 from ChatGPT. Every cited page was coded by source type and publisher country. A blind re-code of a random 20% sample of all the citations in both languages (569 rows) agreed with the first coding on source type for 92.6% of them (Cohen's kappa, an agreement score corrected for chance, 0.87). A small set of Bahasa Malaysia questions and a capture of Google's own AI answers are reported separately as exploratory.

**Key findings:**

- **In this audit, both engines almost always searched before answering whom to hire.** ChatGPT searched the web before every one of its 84 answers to those questions, and Gemini before 82 of 84. For questions about how a service works, Gemini searched in 13.1% of answers and ChatGPT in 86.9% (Table 3).
- **Third-party pages were the minority, and differed by engine.** Directories and review platforms were 1.2% of Gemini's citations and 4.2% of ChatGPT's. ChatGPT also cited government and association pages (17.1%) and Google's own help and documentation pages (10.9%), mostly for questions about how a service works, against 0.1% and 0.0% for Gemini (Table 2).
- **Some "best agency" lists were written by agencies, and two more were agency-supplied rankings on media sites.** Lists ranking or comparing providers, published on agencies' own websites, were 13.0% of Gemini's citations and 0.9% of ChatGPT's. Two cited pages were rankings of AI SEO agencies published on media sites as agency-supplied content, one marked as sponsored and the other placed in a news site's "Announcement" section. None of the answers that cited them said so.
- **The cited pages were mostly Malaysian.** In this audit, pages from Malaysian businesses and publishers made up 91.1% of Gemini's citations and 81.0% of ChatGPT's. "Malaysian" means a stated Malaysian base or legal entity, or a .my domain, and a platform page about one business takes that business's country (Table 5). Publisher country was the less consistently coded of the two checked fields (Cohen's kappa 0.57).
- **About half the cited websites recurred a week later.** When the same questions were asked again a week later, 55.8% (Gemini) and 54.2% (ChatGPT) of the websites cited for each question in the first window were cited for it again, so the study's working hypothesis that fewer than half would recur

ALIQ Studio Research

did not hold. Part of the change that did occur is ordinary variation between runs rather than change over time, and two windows show no trend.

**Main limitation:** The questions were written for this study, not sampled from what businesses ask, and the engines ran through their developer APIs with fixed model versions rather than through the consumer apps. For ChatGPT, the audit counts the inline citations attached to each answer, not every page it read. The results describe this panel on these dates and show what was cited, not why.

**Business implication:** In this audit, most of what the answers cited was providers' own material, so an assistant's recommendation may be closer to a summary of how providers describe themselves than to an independent assessment; the engines may also have had few independent Malaysian sources to draw on. A business choosing a provider should treat a name in an AI answer as a lead to check, not a verdict: open the cited sources, see who published them, and look for independent evidence the answer did not show. For a business that wants to be found, the provider pages cited here were usually not homepages, and because the cited set varied between runs and between two windows a week apart, a fixed set of questions asked on a schedule gives a fairer picture of visibility than a single check. These steps form a proposed checklist, not a tested intervention.

## What this paper examines

This paper looks at the sources an AI assistant's answer cites when a Malaysian business asks who should build its website, run its social media or manage its advertising. It is written for the owner or marketing lead of a Malaysian small or medium-sized enterprise (SME) who is about to hire outside marketing help, or who is deciding whether being visible in AI answers deserves a budget.

**The primary question is: when a Malaysian SME asks an AI assistant who to hire for marketing, what sources does the answer cite?**

Six supporting questions follow from it:

1. How often does the assistant search the web before answering, and does that differ by the kind of question?
2. What share of the cited sources are pages owned by marketing providers, and what share are third-party pages?
3. Which third-party platforms recur?
4. Are the cited pages Malaysian or foreign?
5. A week later, how much of the cited set is the same?
6. When a provider's own page is cited, is it a homepage or a page built for that service or industry?

### Why it matters for Malaysian SMEs

Micro, small and medium enterprises (MSMEs) produced 39.7% of Malaysia's gross domestic product and 48.7% of its employment in 2025, according to the Department of Statistics Malaysia (DOSM) [1]. The Economic Census 2023 counted 1,069,831 MSME establishments in 2022 [2]. Under the national definition, a services business is an SME if its sales are up to RM20 million or it has up to 75 full-time employees, and a manufacturer if its sales are up to RM50 million or it has up to 200 employees [3]. Nearly everyone in Malaysia is online: 98.3% of individuals used the internet in 2025, and 94.6% of internet users searched online for information about goods and services [4].

ALIQ Studio Research

AI-generated answers are now part of search at a very large scale. Google said in May 2026 that AI Overviews had over 2.5 billion monthly users worldwide and AI Mode over 1 billion [5]. In July 2026 it said it had brought AI Overviews and AI Mode together into one Search experience [6]. When an assistant names or recommends a marketing provider, the business owner reading the answer is relying on whatever sources the answer was built from. This paper measures those sources.

### Scope

The audit covers English questions about seven kinds of marketing service (websites, search engine optimisation, social media management, content production, paid advertising, branding, and visibility in AI answers), asked of Gemini and ChatGPT through their developer interfaces (APIs), with a small exploratory set of Bahasa Malaysia questions and a small exploratory capture of Google's own AI answers. Local questions name Kuala Lumpur, Selangor or Petaling Jaya.

It does not test Perplexity, Claude, Microsoft Copilot, Meta AI or Chinese-language assistants, so it says nothing about what they cite. It does not name any brand in its questions. It does not assess the quality of any provider, and it is not a measurement of ALIQ Studio's own visibility: ALIQ Studio's pages are identified only in one disclosure paragraph.

### Definitions

- **Generative engine optimisation (GEO)** is the term a 2023 research paper introduced for methods that aim "to aid content creators in improving their content visibility in generative engine responses" [7].
- **Citation:** a web address the engine returned in its own source field, not one that appears only in the answer text. The method section gives the exact definition for each engine.
- **Provider-owned page:** a page on the website of an agency or freelancer that sells marketing, web, creative or advertising services. **Third-party page:** every other cited page, such as a directory, a government page, a news article or a social profile. A provider's own profile on a social network or a directory counts as a third-party page, because the coding goes by the website that hosts the page.
- **Agency-published list:** a page on an agency's own website that ranks or compares several providers, such as a top-ten list of agencies in Kuala Lumpur.

### What this paper adds

This paper adds a pre-registered, two-window audit of the sources Gemini and ChatGPT cite when asked Malaysian marketing-services questions, with its source taxonomy, coding rules, agreement figures and an anonymised trial-level dataset. It is meant to help Malaysian SME owners and marketing leads decide how much weight to give an AI assistant's recommendation, and where to spend effort if they want to be found in such answers.

A small earlier study, published on 21 July 2026, checked eight commercial queries on Malaysia-localised Google results [8]. We found no pre-registered, two-window, two-engine audit of AI citations for Malaysia with a published dataset. The searches we ran are listed in the supporting material.

## Methodology and evidence base

The paper combines three kinds of work, each labelled where it appears. The **citation audit** is original research: data ALIQ Studio collected and coded. The sections on how the platforms treat sources and on what other research has measured are a **research synthesis**. The external sources are the platforms' own documentation, read at source and dated; official Malaysian statistics; and published studies whose

method, sample and period are stated. Company-run studies are labelled as such. One study was excluded because its conclusion changed between versions and its method got around a platform's usage limits, and another because its main measure is not defined. The checklist for SMEs is an **expert framework**: proposed, practice-informed and untested.

### Pre-registration

The research questions, prompts, analysis plan and coding rules were fixed and recorded before any data was collected. The protocol was frozen on 26 September 2026, before the first call, and is published with this paper; it was not lodged with an external registry. Every later departure from them is listed with its date and reason in the published protocol.

### The question panel

The panel has 42 English prompts covering 7 service categories, 3 kinds of question (choosing a provider, understanding how a service works, and cost) and 2 phrasings of each question. Of the English prompts, 13 are local: they name Kuala Lumpur, Selangor or Petaling Jaya; the other 29 ask about Malaysia as a whole. A further 6 Bahasa Malaysia prompts are reported separately as exploratory. The panel was frozen before the first call (version 1.0, 26 September 2026). Some prompts reproduce the shape of real queries that reached ALIQ Studio's website, some paraphrase questions prospects asked on sales calls, and the rest are editorial. Each is tagged in the published prompt file, no prompt reuses a caller's wording, and no prompt names a brand.

### Engines and settings

Gemini ran as gemini-3.8-flash through Google's Gemini API with Google Search grounding. ChatGPT ran as gpt-5.5-2026-04-23 through OpenAI's Responses API with web search and an approximate Kuala Lumpur location. Our Gemini configuration set no location, so Gemini answered every prompt without one. Neither engine is the consumer app, and no account or personal history was involved.

The consumer apps can run other models. As of 6 October 2026, OpenAI said GPT-5.6 Luna was becoming the default for ChatGPT's Free and Go users [9], and Google said Gemini 3.8 Flash was available in the Gemini app to Pro and Ultra subscribers [10]. The API results are a proxy for the apps, not a measurement of them.

### Repetition and timing

Each prompt was asked 3 times per engine in each of 2 windows at least seven days apart. W1 ran on 26 to 27 September 2026 and W2 on 4 October 2026 (Kuala Lumpur time). In W1, 75 ChatGPT answers failed when the API account ran out of credit; they were re-run the same day and marked as retries, and the failed attempts are reported alongside the results.

### What counts as a citation

A citation is a web address the engine returned in its own source field: for Gemini, the grounding sources listed in the response, followed to the final page [11]; for ChatGPT, the inline citations attached to the answer [12]. Addresses that appear only in the answer text are not counted. OpenAI says its inline citations show "only the most relevant references", and that a separate field lists every page the model consulted [12]. This audit counts the inline citations a user sees, not every page the model read. Shares count each cited page once per answer, and measures about websites count each website once per answer, so a page an answer cites twice is counted once.

ALIQ Studio Research

### Coding the sources

Every cited page was assigned one of 14 source types, a publisher country and a page type, under rules fixed before coding and published with this paper. The first coding was done by the AI assistant that ran the study, under the author's direction. The reviewer, Ricky Wong, then re-coded a random 20% sample of all the citations in both languages without seeing the first codes. Agreement was 92.6% on source type (Cohen's kappa 0.87) and 88.9% on publisher country (kappa 0.57), over 569 sampled rows. Had agreement fallen below 80%, the rules would have been revised and the set re-coded. Each disagreement was then settled by discussion, and the settled codes are the ones reported. Three rules on publisher country were clarified in that step and applied to every citation, not only the sample: a platform page about one business takes that business's country, a platform company's own pages take the company's country, and a publisher counts as foreign where it states a base outside Malaysia or is a well-known company or publisher based outside Malaysia. The agreement figures were measured before these clarifications.

### Google's own AI answers (exploratory)

A small exploratory sub-panel of 12 prompts was captured once per window from Google's standard results page and from its AI Mode tab, in the author's signed-in Chrome browser, which is personalised, including by location. Google says it has brought AI Overviews and AI Mode together into one Search experience [6]; on 27 September 2026 the AI answer on the standard results page was labelled "AI Mode reply" for all 11 sub-panel prompts captured that day (the twelfth was captured on 26 September). On 4 October 2026 the standard results page also showed a visible "AI Overview" label above the same AI answer. The capture method was adjusted in one line so that it read the answer rather than the label, and both versions were tested on the two page layouts. The W2 captures were taken with the browser window off screen; the links still loaded, but the one W2 AI Mode answer with no sources could not be checked by eye. These captures are reported separately and are not comparable with the API panel.

### Analysis

All results are descriptive: shares with their denominators, overlap between the two windows, and the number of websites that account for half of all citations. There are no significance tests. The prompts were written for this study, not sampled from all the questions people ask, so the results describe this panel on these dates. They show what was cited, not why.

## How the platforms say they choose and show sources

This section is a research synthesis of the platforms' own documentation, read at source on 26 and 27 September 2026; the entries that change often were re-read on 6 October 2026.

**Google.** Google says a site does not need AI text files, special markup, "chunking", an ideal page length or AI-specific writing to appear in its generative AI features [13]. To be shown as a supporting link in AI Overviews or AI Mode, a page must be indexed and eligible to appear in Google Search with a snippet [14]. For Google, being findable in Search is therefore the precondition for being cited. Since 31 August 2026, site owners everywhere can include or exclude their site from AI Overviews and AI Mode in Google Search Console; inclusion is the default, and the setting does not govern AI training [15, 16]. Search Console now reports how often a site's links were shown in these AI features, although Google notes the report is not yet available to every site. It reports impressions, not clicks [17].

**OpenAI.** Sites that block OpenAI's search crawler, OAI-SearchBot, are not shown in ChatGPT search answers, though they can still appear as navigational links. Blocking OpenAI's training crawler, GPTBot, is a separate choice [18, 19]. OpenAI describes ChatGPT's ranking of search results only in general terms: it

"ranks search results using multiple factors", and "Placement is not guaranteed." [19] As noted above, the inline citations a ChatGPT user sees are the most relevant subset of the pages consulted [12].

**Microsoft.** Bing Webmaster Tools, in a feature still in public preview, reports when a site is cited in Microsoft Copilot and Bing's AI summaries, and Microsoft says a citation is not a click or a visit [20, 21].

**Cloudflare.** Cloudflare's documentation, last updated on 2 November 2026, says that for new domains search, agent and training crawlers are all allowed by default [22]. That differs from its July 2026 announcement that, from 15 September 2026, new domains would block AI training and agent crawlers on pages that show ads, while search crawlers stayed allowed [23]. Cloudflare's own pages therefore disagree about what a new site gets, so the setting that applies to a site is the one shown in its own Cloudflare dashboard.

## What other research has measured

This section summarises published research. None of it was collected by ALIQ Studio, and most of it is not about Malaysia.

The original GEO study found that changing a page's content could raise its visibility in AI answers, but it measured that effect among sources that had already been retrieved for the question [7]. A 2026 survey of 45 studies, published as a preprint, notes that those gains "establish neither organic discoverability nor durable traffic effects". It found topical relevance and position within the retrieved content to be the most reproducible levers, and noted that commercial audits report "low source overlap" and "substantial run-to-run variability" [24].

Being cited does not mean being visited. In the United States, Google users who saw an AI summary clicked a traditional result in 8% of visits, against 15% when there was no summary (Pew Research Center, March 2025 browsing data) [25, 26]. A separate study of US desktop users found ChatGPT conversations led to a click on an outside website in 5.2% of sessions [27]. AI chatbots still send a small share of web traffic: under half of one percent of visits to the tens of thousands of websites in one vendor's panel in 2026 [28, 29].

Two company-run studies bear directly on this audit. A study by the SEO software company Ahrefs found that "best X" blog lists were 40.64% of the pages ChatGPT cited for agency-recommendation prompts (750 prompts across three categories, November 2025, with agency examples set in London) [30]. Its page-type definitions differ from this audit's, so only the direction is compared here, and for ChatGPT it differs: in this audit, lists on agencies' own websites were 0.9% of ChatGPT's citations (and 13.0% of Gemini's). Semrush found that being cited as a source and being named as a recommended brand are different outcomes: in its US data, only 21% of the most-cited websites in a category were also the most-mentioned brand [31].

In the one earlier Malaysian study we found, the local marketing-services queries on Google showed a map of local businesses rather than an AI Overview [8]. That study used short keyword queries on a single day in July 2026, before Google's merge of AI Overviews and AI Mode. This audit's question-form prompts are a different query style.

ALIQ Studio Research

## What the answers cited

The results below pool both windows (W1 from 26 September 2026 to 27 September 2026, and W2 on 4 October 2026) and cover the English panel unless a section says otherwise. Shares count each cited page once per answer. They describe this panel on these dates, asked through the engines' APIs rather than their consumer apps, and show what was cited, not why. A blind re-code of a random 20% sample of all the citations in both languages agreed with the first coding on source type for 92.6% of the sampled citations (Cohen's kappa 0.87).

### What sources does the answer cite?

Mostly pages that marketing providers publish on their own websites. Pages owned by agencies and freelancers made up 87.4% of Gemini's citations (1,322 of 1,513) and 59.9% of ChatGPT's (558 of 932). The study's first working hypothesis, that third-party pages would outnumber providers' own pages, did not hold on either engine. Lists ranking or comparing providers, published on agencies' own websites, were 13.0% of Gemini's citations and 0.9% of ChatGPT's.

**Table 1. Providers' own pages and third-party pages among the citations**

| | Gemini | ChatGPT |
|---|---|---|
| Providers' own pages | 87.4% (1,322) | 59.9% (558) |
| Third-party pages | 12.6% (191) | 40.1% (374) |
| All citations | 1,513 | 932 |
| *Providers' own pages by kind of question (exploratory): share of that kind of question's citations* | | |
| Whom to hire | 86.1% (627 of 728) | 84.8% (307 of 362) |
| How a service works | 85.3% (64 of 75) | 18.8% (57 of 303) |
| Cost | 88.9% (631 of 710) | 72.7% (194 of 267) |

Scope: English questions about marketing services in Malaysia, asked of Gemini and ChatGPT through their APIs in two windows (26 to 27 September and 4 October 2026). Each cited page counts once per answer.

Source: Yik Jinn, ALIQ Studio (2026). What AI answer engines cite when Malaysian SMEs ask for marketing help, v1.1. https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/. CC BY 4.0.

Split by kind of question (an exploratory cut, added after the planned analysis), the gap between the engines comes mostly from questions about how a service works. For questions about whom to hire, providers' own pages were 86.1% of Gemini's citations and 84.8% of ChatGPT's.

The third-party pages differed by engine, as the study's second working hypothesis (that the two engines would cite different mixes of source types) expected. Government and industry-association pages were 17.1% of ChatGPT's citations and 0.1% of Gemini's. Google's own pages were 10.9% of ChatGPT's and 0.0% of Gemini's. ChatGPT cited both mostly for questions about how a service works. News and media pages were 3.0% of Gemini's citations and 0.3% of ChatGPT's; 24 of Gemini's 45 were the two agency-supplied rankings described below. Directories and review platforms were a small share on both: 1.2% for Gemini and 4.2% for ChatGPT.

ALIQ Studio Research

**Table 2. Source type of each cited page, both windows**

| Source type | Gemini | ChatGPT |
|---|---|---|
| Provider's own page (service, homepage or article) | 70.9% (1,073) | 57.1% (532) |
| Provider's own list of providers | 13.0% (196) | 0.9% (8) |
| Freelancer's own site | 3.5% (53) | 1.9% (18) |
| Independent list of providers | 0.3% (5) | 0.0% (0) |
| Directory or review platform | 1.2% (18) | 4.2% (39) |
| Press release | 0.0% (0) | 0.5% (5) |
| Social network | 2.0% (31) | 0.1% (1) |
| Job board or company profile | 0.1% (1) | 1.3% (12) |
| Services marketplace | 0.5% (8) | 1.6% (15) |
| News and media | 3.0% (45) | 0.3% (3) |
| Government or association | 0.1% (2) | 17.1% (159) |
| Google's own pages | 0.0% (0) | 10.9% (102) |
| Software or platform vendor | 3.7% (56) | 2.8% (26) |
| Other or unresolvable | 1.7% (25) | 1.3% (12) |
| All citations | 1,513 | 932 |

Scope: English questions about marketing services in Malaysia, asked of Gemini and ChatGPT through their APIs in two windows (26 to 27 September and 4 October 2026). Each cited page counts once per answer.

Source: Yik Jinn, ALIQ Studio (2026). What AI answer engines cite when Malaysian SMEs ask for marketing help, v1.1. https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/. CC BY 4.0.

"Other or unresolvable" includes pages on independent sites that are not lists, such as a how-to guide on a blog that sells no marketing services, as well as dead or blocked links. The codebook gives the full rules.

For ChatGPT, these are the inline citations attached to each answer, which OpenAI describes as the most relevant subset of the pages consulted, not every page the model read [12].

For these questions, an assistant's recommendation mostly cited what providers publish on their own websites, so it may be closer to a summary of self-description than to an independent assessment. The engines may also have had few independent Malaysian sources to draw on. Either way, a provider named in an answer is a lead to check, not a verdict (checklist item 1).

## How often does the assistant search before answering?

For choosing a provider and for cost, almost always. For how a service works, it depends on the engine. ChatGPT searched the web before every one of its 84 answers to questions about whom to hire, and Gemini before 82 of 84. For questions about how a service works, Gemini searched in 13.1% of answers and ChatGPT in 86.9%. An answer given without a search has no citations.

**Table 3. Answers that searched the web before answering, by kind of question**

| Kind of question | Gemini | ChatGPT |
|---|---|---|
| Choosing a provider | 82 of 84 (97.6%) | 84 of 84 (100.0%) |
| How a service works | 11 of 84 (13.1%) | 73 of 84 (86.9%) |
| Cost | 78 of 84 (92.9%) | 84 of 84 (100.0%) |
| All questions | 171 of 252 (67.9%) | 241 of 252 (95.6%) |

Scope: English questions about marketing services in Malaysia, asked of Gemini and ChatGPT through their APIs in two windows (26 to 27 September and 4 October 2026).

An answer counts as searched when the API's response records a web search: a search query for Gemini, a web search call for ChatGPT.

Source: Yik Jinn, ALIQ Studio (2026). What AI answer engines cite when Malaysian SMEs ask for marketing help, v1.1. https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/. CC BY 4.0.

What was cited also differed by kind of question, mostly in ChatGPT's answers (an exploratory cut). For questions about how a service works, government and association pages made up 40.9% of ChatGPT's citations and Google's own pages 26.4%, against 0.0% and 0.0% for Gemini, which searched for few of these questions (Table 3). For questions about whom to hire, ChatGPT's shares were 4.4% and 0.8%.

ChatGPT's heavier use of government, Google and directory pages may reflect how often it searched for questions about how a service works, or a different search system behind it. This audit cannot tell which.

## Which third-party platforms recur?

Google's own documentation and Malaysian public bodies, more than review platforms. The most-cited non-provider website was google.com, cited in 46 answers across 12 questions, all of them ChatGPT's, mostly to Google's Search documentation and its Google Ads and Business Profile help pages. The public bodies cited most often were the Personal Data Protection Department (in 35 answers), the Companies Commission of Malaysia (22), the Intellectual Property Corporation of Malaysia (17) and the Ministry of Health (11), all of them in ChatGPT's answers.

ALIQ Studio Research

**Table 4. The ten most-cited named websites that are not marketing providers**

| Website | Source type | Answers citing it | Questions | Gemini | ChatGPT |
|---|---|---|---|---|---|
| google.com | Google's own pages | 46 | 12 | 0 | 46 |
| pdp.gov.my | Government or association | 35 | 9 | 0 | 35 |
| clutch.co | Directory or review platform | 24 | 11 | 6 | 18 |
| ssm.com.my | Government or association | 22 | 6 | 0 | 22 |
| facebook.com | Social network | 18 | 9 | 17 | 1 |
| myipo.gov.my | Government or association | 17 | 5 | 0 | 17 |
| malaysiakini.com | News and media | 17 | 4 | 17 | 0 |
| google.cn | Google's own pages | 12 | 6 | 0 | 12 |
| moh.gov.my | Government or association | 11 | 3 | 0 | 11 |
| handld.org | Services marketplace | 9 | 5 | 0 | 9 |

Each website is counted once per answer, both engines together. Agencies and freelancers are excluded here and anonymised everywhere. So are 9 other websites that sell marketing, web, creative or advertising services but are coded otherwise in the codebook (for example a hosting company, a business directory, or a site that could not be reached at coding).

Scope: English questions about marketing services in Malaysia, asked of Gemini and ChatGPT through their APIs in two windows (26 to 27 September and 4 October 2026).

Source: Yik Jinn, ALIQ Studio (2026). What AI answer engines cite when Malaysian SMEs ask for marketing help, v1.1. https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/. CC BY 4.0.

All 17 answers that cited malaysiakini.com cited one page: an agency-supplied ranking of AI SEO agencies in its "Announcement" section (see the section on agency-written lists below).

Directories and review platforms were a small share of all citations (Table 2), so a buyer who wants an independent view of a provider will usually have to look beyond the answer (checklist item 2).

## Are the cited pages Malaysian?

Mostly. Pages from Malaysian businesses and publishers made up 91.1% of Gemini's citations and 81.0% of ChatGPT's. The rest were foreign publishers (for ChatGPT, these include Google's own help and documentation pages), platform pages about many businesses (a directory category page or a discussion thread), or publishers whose base could not be established. "Malaysian" here means a stated Malaysian base or legal entity, or a .my domain; a platform page about one business takes that business's country.

**Table 5. Publisher country of each cited page**

| Publisher | Gemini | ChatGPT |
|---|---|---|
| Malaysian | 91.1% (1,379) | 81.0% (755) |
| Platform page about many businesses | 2.4% (36) | 2.5% (23) |
| Foreign | 5.8% (88) | 15.5% (144) |
| Could not be established | 0.7% (10) | 1.1% (10) |
| All citations | 1,513 | 932 |

Scope: English questions about marketing services in Malaysia, asked of Gemini and ChatGPT through their APIs in two windows (26 to 27 September and 4 October 2026). Each cited page counts once per answer.

"Malaysian" means a stated Malaysian base or legal entity, or a .my domain; a platform page about one business takes that business's country.

Source: Yik Jinn, ALIQ Studio (2026). What AI answer engines cite when Malaysian SMEs ask for marketing help, v1.1. https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/. CC BY 4.0.

Publisher country was the less consistently coded of the two checked fields (Cohen's kappa 0.57 in the blind re-code, measured before three country rules were clarified; see Coding the sources).

### Is the cited set the same a week later?

About half of it was. Per question, the median overlap (Jaccard index, from zero for no shared websites to one for the same set) between the websites cited in W1 and in W2 was 0.40 for Gemini and 0.35 for ChatGPT, and 55.8% and 54.2% of the websites cited for a question in W1 were cited for it again in W2. The overlap is computed over the questions with a citation in either window: 29 for Gemini and 42 for ChatGPT.

The study's third working hypothesis, that fewer than half of the websites cited for a question in W1 would be cited for it again in W2, therefore did not hold on either engine.

Each window pooled three runs of every question, so part of this change is the ordinary variation between runs rather than change over time, and two windows a week apart show no trend. A 2026 survey of published studies notes that commercial audits report "substantial run-to-run variability" [24].

A single check of what an assistant says is a snapshot. A fixed set of questions asked on a schedule gives a fairer picture (checklist item 6).

### When a provider is cited, which page is it?

Usually not the homepage. When an agency's own page was cited, it was the homepage in 4.3% of Gemini's cases and 18.5% of ChatGPT's. Service or industry pages made up 40.3% and 50.6%, and articles or lists 54.2% and 29.8%. Page type was coded by the AI assistant and was not part of the blind re-code.

**Table 6. Page type among agencies' own cited pages**

| Page type | Gemini | ChatGPT |
|---|---|---|
| Homepage | 4.3% (54) | 18.5% (100) |
| Service or industry page | 40.3% (512) | 50.6% (273) |
| Article or list | 54.2% (688) | 29.8% (161) |
| Other | 1.2% (15) | 1.1% (6) |
| All agency-owned citations | 1,269 | 540 |

Scope: English questions about marketing services in Malaysia, asked of Gemini and ChatGPT through their APIs in two windows (26 to 27 September and 4 October 2026). Each cited page counts once per answer.

Source: Yik Jinn, ALIQ Studio (2026). What AI answer engines cite when Malaysian SMEs ask for marketing help, v1.1. https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/. CC BY 4.0.

For a business that wants to be found, the provider pages cited here were usually not the homepage. This shows which pages were cited, not that adding a page leads to citation (checklist item 3).

### How concentrated are the citations?

Fairly concentrated. Counting each website once per answer, 24 websites made up half of all the times Gemini cited a website, although 249 different websites were cited. For ChatGPT, 23 websites made up half, out of 219. In this panel, those few websites accounted for half of the times a website was cited, so their pages may sit behind many of these answers, although the audit shows only what was cited, not how each answer used it. This describes this panel, not market share or quality.

ALIQ Studio Research

### Agency-written and agency-supplied lists among the sources

Two of the cited pages were rankings of AI SEO agencies published on media sites as agency-supplied content. One, on a trade-media site, is marked "This post is sponsored by" an agency. The other sits in a news site's "Announcement" section and is marked as content provided by an agency. Together they drew 24 of Gemini's citations and 0 of ChatGPT's. Other "best agency" lists the engines cited were written by agencies ranking their own field. None of the answers that cited the two media-site rankings said they were agency-supplied.

### Exploratory: Bahasa Malaysia prompts and Google's own AI answers

In the six exploratory Bahasa Malaysia prompts, providers' own pages were again most of the citations: 90.8% of Gemini's and 62.4% of ChatGPT's. Six prompts cannot stand for Malay-language search in general.

In the exploratory Google sub-panel, every captured page showed an AI answer in both windows. Providers' own pages were 77.7% of the links in the answer on the standard results page and 79.8% in AI Mode, where 2 answers (one in each window) had no links to outside websites. These captures came from one personalised browser, once per window, and are not comparable with the API panel.

## What businesses should do with these findings

The checklist below is a **proposed, practice-informed framework**. It has not been tested as an intervention. Each item names the evidence it rests on, who it is for and how to judge whether it worked, and, where they apply, what it needs, its trade-off and when it does not fit.

### If you are choosing a provider an AI assistant named

**1. Open the sources and check who published them.** Before contacting a provider an AI assistant named, open the sources under the answer and check who published each one. A provider's own page, or its own list of the "best" agencies, is self-description, not a review. *Evidence:* in this audit, run through the engines' APIs, most cited pages were providers' own pages (87.4% of Gemini's citations and 59.9% of ChatGPT's), and agencies' own lists of providers were 13.0% of Gemini's. *For:* SMEs choosing a provider. *Needs:* an answer that shows its sources. *Trade-off:* a few minutes per shortlisted name. *Evaluate:* count how many of the cited sources were independent of the provider.

**2. Look for evidence the answer did not show.** Look for client work you can verify, reviews on platforms the provider does not control, and the company's registration. *Evidence:* directories and review platforms were a small share of citations in this audit (1.2% for Gemini and 4.2% for ChatGPT), and Google does not show review stars for businesses that mark up reviews they control about themselves [32]. *For:* SMEs choosing a provider. *Needs:* time to check two or three sources. *Trade-off:* slower shortlisting. *Does not fit:* very small one-off jobs. *Evaluate:* whether at least one independent source confirms the provider's claims.

### If you want your own business to be found

**3. Give each service you sell a clear page of its own.** *Evidence:* when a provider's own page was cited in this audit, it was usually not the homepage (4.3% of Gemini's cases and 18.5% of ChatGPT's); service or industry pages were 40.3% and 50.6% (Table 6). This shows which pages were cited, not that adding a page leads to citation. Google advises making pages for your audience, "not just for generative AI search"

ALIQ Studio Research

[13]. *For:* businesses that sell more than one service. *Needs:* real content for each page. *Trade-off:* writing and upkeep. *Does not fit:* single-service businesses, or thin copies of one page. *Evaluate:* the queries and AI-feature impressions each page receives in Search Console [17].

**4. Skip AI-specific files; make sure search crawlers can reach you.** Google says it does not use AI text files or special markup [13]. Instead, check that search crawlers can reach your site, including OpenAI's search crawler [18], and that your hosting or content delivery network is not blocking them [22]. *For:* any business with a website. *Needs:* access to your site's crawler rules and hosting settings. *Does not fit:* sites whose owners choose to block AI crawlers. *Evaluate:* your crawler rules and hosting settings allow the named crawlers.

**5. Do not mass-produce pages or "best agencies" lists to get cited.** Google treats using generative AI tools to produce many pages without adding value as spam [33], and a careful reader can see who wrote a list. *For:* providers tempted to publish "best agencies" lists on their own websites. *Trade-off:* fewer pages. *Evaluate:* each published page adds information a buyer could not get elsewhere.

**6. Measure visibility with the platforms' own reports, and do not mistake it for traffic.** Use Search Console's generative AI report [17], Bing's AI Performance report [20] and a fixed set of questions asked on a schedule, because the cited set varied between runs and between two windows a week apart (in this audit, 55.8% and 54.2% of the websites cited for a question in the first window were cited for it again in the second). Treat a citation as visibility, not as traffic or enquiries [21]. *For:* anyone paying for AI-visibility work. *Needs:* access to Search Console and Bing Webmaster Tools. *Trade-off:* setup time. *Does not fit:* sites that do not have these reports yet. *Evaluate:* enquiries traced to AI answers, recorded separately.

**7. Keep reviews and your Business Profile within the rules.** Google prohibits fake or undisclosed incentivised reviews on a page or in its markup [32], and its Business Profile rules require the listed name to match the business's real-world name, and adding extra words to the name can get a profile suspended [34]. *For:* businesses with reviews or a Business Profile. *Evaluate:* a profile and markup check against the rules.

ALIQ Studio Research

## Limitations and remaining questions

- **A designed panel, not a sample.** The prompts were written for this study, so the results describe this panel on these dates, not every question Malaysian businesses ask.
- **APIs, not apps.** The engines ran through their developer interfaces with fixed model versions, while the consumer apps can run other models. The results are a proxy for what app users see, not a measurement of it.
- **ChatGPT's citations are a subset.** The audit counts the inline citations attached to each answer, which OpenAI describes as the most relevant subset of the pages consulted.
- **Location.** ChatGPT's answers used an approximate Kuala Lumpur location. Our Gemini configuration set none, and the Google sub-panel ran in one personalised browser.
- **One week apart.** Two windows a week apart describe the short term, not change over months.
- **English first.** The Bahasa Malaysia set is small and exploratory, and no Chinese-language prompts were run.
- **Coding.** An AI assistant did the first coding. A human reviewer re-coded a random sample for source type and publisher country, and agreement is reported in the method section. Page type was coded by the AI assistant only.
- **Interruptions.** W1 was paused for a day and resumed within its window, and some ChatGPT answers were re-run after the API account ran out of credit. Both are recorded in the protocol.

- **Citations are not clicks.** The audit measures what was cited, not what users clicked or whether anyone became a customer.

**Remaining questions:**

- What do the consumer apps cite for the same questions?
- How do Malay and Chinese questions compare?
- How does the cited set change over months rather than a week?
- Which properties of a page make it more likely to be retrieved and cited?

This design cannot answer the last question: it describes what was cited and does not test what makes a page cited.

## References


1. Department of Statistics Malaysia. (2026, July 30). *Malaysia's Micro, Small and Medium Enterprises (MSMEs) Performance 2025* [media statement]. DOSM. https://www.dosm.gov.my/uploads/release-content/file_20260730113107.pdf (accessed 27 September 2026; re-read 6 October 2026).
2. Department of Statistics Malaysia. (2024, September 19). *Economic Census 2023: Profile of Small and Medium Enterprises* [media statement]. DOSM. https://www.dosm.gov.my/site/downloadrelease?id=economic-census-2023-profile-of-small-and-medium-enterprises&lang=English (accessed 27 September 2026; re-read 6 October 2026).
3. SME Corporation Malaysia. (n.d.). *Official Definition of SME*. SMEinfo Portal. https://www.smeinfo.com.my/official-definition-of-sme/ (accessed 27 September 2026).
4. Department of Statistics Malaysia. (2026, April 23). *ICT Use and Access by Individuals and Households Survey Report, 2025* [media statement]. DOSM. https://www.dosm.gov.my/uploads/release-content/file_20260422202124.pdf (accessed 27 September 2026; re-read 6 October 2026).
5. Pichai, S. (2026, May 19). *Google I/O 2026: Sundar Pichai's opening keynote*. Google, The Keyword. https://blog.google/innovation-and-ai/sundar-pichai-io-2026/ (accessed 27 September 2026; re-read 6 October 2026).
6. Pichai, S. (2026, July 22). *Alphabet earnings call Q2 2026: Sundar Pichai remarks*. Google, The Keyword. https://blog.google/company-news/inside-google/message-ceo/alphabet-earnings-q2-2026/ (accessed 27 September 2026).
7. Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K., & Deshpande, A. (2024). *GEO: Generative Engine Optimization*. Accepted to KDD 2024. arXiv:2311.09735, version 3, 28 June 2024. https://arxiv.org/abs/2311.09735 (accessed 26 September 2026).
8. Fong, M. (2026, July 21; updated 19 August 2026). *State of AI Search in Malaysia 2026: What Commercial Queries Reveal About Who AI Cites*. Ai Mode. https://aimode.my/state-of-ai-search-malaysia-2026/ (accessed 27 September 2026; re-read 6 October 2026).
9. OpenAI. (n.d.; updated September 2026). *GPT-5.6 and GPT-6 Pro in ChatGPT*. OpenAI Help Center. https://help.openai.com/en/articles/20001354-gpt-56-and-gpt-6-pro-in-chatgpt (accessed 27 September 2026; re-read 6 October 2026).
10. Doshi, T., & Popa, R. A. (2026, September 2). *Introducing Gemini 3.8 Flash and 3.8 Flash Cyber*. Google, The Keyword. https://blog.google/innovation-and-ai/models-and-research/gemini-models/3-8-flash-and-3-8-flash-cyber/ (accessed 27 September 2026; re-read 6 October 2026).
11. Google. (n.d.; last updated 23 September 2026). *Generating content* (Gemini API reference). Google AI for Developers. https://ai.google.dev/api/generate-content (accessed 26 September 2026).
12. OpenAI. (n.d.). *Web search*. OpenAI API documentation. https://developers.openai.com/api/docs/guides/tools-web-search (accessed 26 September 2026).

13. Google Search Central. (n.d.; last updated 10 July 2026). *Optimizing your website for generative AI features on Google Search*. Google for Developers. https://developers.google.com/search/docs/fundamentals/ai-optimization-guide (accessed 26 September 2026).
14. Google Search Central. (n.d.; last updated 10 December 2025). *AI features and your website*. Google for Developers. https://developers.google.com/search/docs/appearance/ai-features (accessed 26 September 2026).
15. Google. (n.d.). *Search generative AI control*. Search Console Help. https://support.google.com/webmasters/answer/16908024 (accessed 26 September 2026; re-read 6 October 2026).
16. Loew, M. (2026, June 3; updated 31 August 2026). *New opportunities, control and insights for website owners*. Google, The Keyword. https://blog.google/products-and-platforms/products/search/new-controls-website-owners/ (accessed 26 September 2026).
17. Google. (n.d.). *Generative AI performance report (Search)*. Search Console Help. https://support.google.com/webmasters/answer/16984139 (accessed 26 September 2026).
18. OpenAI. (n.d.). *Overview of OpenAI Crawlers*. OpenAI developer documentation. https://developers.openai.com/api/docs/bots (accessed 26 September 2026).
19. OpenAI. (n.d.; updated August 2026). *Searching the web with ChatGPT*. OpenAI Help Center. https://help.openai.com/en/articles/9237897-searching-the-web-with-chatgpt (accessed 27 September 2026).
20. Madhavan, K., Merchant, M., Canel, F., & Nigam, S. (2026, February 10). *Introducing AI Performance in Bing Webmaster Tools Public Preview*. Microsoft Bing Webmaster Blog. https://blogs.bing.com/webmaster/2026/2/Introducing-AI-Performance-in-Bing-Webmaster-Tools-Public-Preview/ (accessed 26 September 2026).
21. Microsoft. (n.d.). *AI Performance*. Bing Webmaster Tools Help. https://www.bing.com/webmasters/help/ai-performance-9f8e7d6c (accessed 27 September 2026).
22. Cloudflare. (n.d.; last updated 2 November 2026). *AI Crawler and Agent access policies*. Cloudflare Docs. https://developers.cloudflare.com/bots/additional-configurations/block-ai-bots/ (accessed 26 September 2026; re-read 5 November 2026).
23. Cloudflare. (n.d.; entry dated 1 July 2026). *Changelog: Bots*. Cloudflare Docs. https://developers.cloudflare.com/bots/changelog/ (accessed 27 September 2026; re-read 5 November 2026).
24. Martinez, O. (2026, July 15). *Optimizing Visibility in Generative Engines: A Critical Survey of Generative Engine Optimization (2023-2026)*. Preprint, arXiv:2607.14035. https://arxiv.org/abs/2607.14035 (accessed 26 September 2026).
25. Chapekis, A., & Lieb, A. (2025, July 22). *Google users are less likely to click on links when an AI summary appears in the results*. Pew Research Center. https://www.pewresearch.org/short-reads/2025/07/22/google-users-are-less-likely-to-click-on-links-when-an-ai-summary-appears-in-the-results/ (accessed 26 September 2026; re-read 6 October 2026).
26. Chapekis, A., Lieb, A., Shah, S., & Smith, A. (2026, August 5). *Investigating Click Behaviors On Google Search Result Pages That Produce an AI Overview*. Extended abstract, arXiv:2608.04831. https://arxiv.org/abs/2608.04831 (accessed 26 September 2026; re-read 6 October 2026).
27. Shi, Q., Zhu, K., & Gu, K. (2026, July 8). *Answering Without Referring: How AI Search Rewrites the Web's Economic Bargain*. Preprint, arXiv:2607.07652. https://arxiv.org/abs/2607.07652 (accessed 26 September 2026; re-read 6 October 2026).
28. Makosiewicz, M. (2026, May 15; updated 12 August 2026). *AI Chatbot Traffic: What It Is, and How to Get More*. Ahrefs blog. https://ahrefs.com/blog/ai-chatbot-traffic/ (accessed 26 September 2026; re-read 6 October 2026).

ALIQ Studio Research

29. Ahrefs. (n.d.; data June 2025 to August 2026). *AI vs Search Traffic Analysis 2025*. chatgpt-vs-google.com. https://chatgpt-vs-google.com/ (accessed 26 September 2026).
30. Allsopp, G. (2025, December 4). *Do Self-Promotional "Best" Lists Boost ChatGPT Visibility? Study of 26,283 Source URLs*. Ahrefs blog. https://ahrefs.com/blog/best-lists-research/ (accessed 26 September 2026; re-read 6 October 2026).
31. Loktionova, M. (2026, July 20). *AI visibility is a topic-level game: A study of 50,000 brands in ChatGPT*. Semrush blog. https://www.semrush.com/blog/chatgpt-topic-authority-study/ (accessed 26 September 2026).
32. Google Search Central. (n.d.; last updated 8 September 2026). *Review snippet (Review, AggregateRating) structured data*. Google for Developers. https://developers.google.com/search/docs/appearance/structured-data/review-snippet (accessed 26 September 2026).
33. Google Search Central. (n.d.; last updated 28 August 2026). *Spam policies for Google web search*. Google for Developers. https://developers.google.com/search/docs/essentials/spam-policies (accessed 26 September 2026).
34. Google. (n.d.). *Guidelines for representing your business on Google*. Google Business Profile Help. https://support.google.com/business/answer/3038177 (accessed 26 September 2026).

## Supporting material

These files are published with the paper at https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/, under the same licence:

- The pre-registered protocol, with its list of deviations.
- The prompt file (version 1.0).
- The codebook.
- The generated tables.
- The anonymised trial-level dataset, in which every agency appears only as [agency] and every unnamed website as [unnamed].
- The record of the literature search behind the section on earlier research.
- A README describing the package: what the dataset covers, what is withheld and why, and what it can and cannot reproduce.

### Version history

- **Version 1.1, 5 November 2026.** Corrects the paragraph on Cloudflare in the section on how the platforms say they choose and show sources: Cloudflare's documentation, updated 2 November 2026, now says new domains allow AI crawlers by default, unlike the July 2026 announcement that version 1.0 reported. No figure or finding of the study changed. The version 1.0 PDF stays available, labelled superseded.
- **Version 1.0, 6 October 2026.** First publication.

### How to cite this paper

Yik Jinn, ALIQ Studio (2026). What AI answer engines cite when Malaysian SMEs ask for marketing help, v1.1. https://studio.aliqgroup.com/en/research/ai-answer-citations-malaysia-2026/. CC BY 4.0.

ALIQ Studio Research

### Licence and reuse

This paper and its dataset are published under the Creative Commons Attribution 4.0 International licence (CC BY 4.0). You may copy, share, quote, adapt and build on them for any purpose, including commercially, as long as you give appropriate credit, link to the licence and say if you changed anything, without suggesting that ALIQ Studio endorses you or your use. The licence does not restrict uses that the law already allows without permission. Suggested credit: as in How to cite this paper, above. Quotations from other publishers in this paper remain under their owners' terms.

## About the contributors and disclosures

ALIQ Studio brings branding, content, websites and performance marketing together under one coordinated team, helping businesses in Kuala Lumpur and Selangor access professional execution without traditional agency overhead.

**Yik Jinn**, Head of SEO and AI Search at ALIQ Studio, designed the study, ran the panel and wrote the paper. **Ricky Wong**, Co-founder and CEO of ALIQ Studio, reviewed it. His review is an internal review, not independent peer review.

An AI assistant (Claude) ran the instrument, coded the sources and drafted the text under the author's direction. The reviewer, Ricky Wong, re-coded a random sample of the coding without seeing the first codes, and signed the release checklist.

**Commercial interest.** ALIQ Studio sells AI-visibility services, so it has a commercial interest in this topic. Its own website was among the five most-cited agency websites in this audit. To keep the comparison fair, its pages were coded under the same rules as every other agency and are anonymised the same way in every table and in the dataset; this paragraph is the only place in the paper where they are identified. Competing agencies are not named; platforms, public bodies and publishers are.

ALIQ Studio offers AI-visibility work for Malaysian businesses: AI visibility for Malaysian businesses.

ALIQ Studio Research