

Websearch Synthesis Optimization

Sarunyu Thongjarast

June 2026

Abstract

Company-activity web-search queries ask whether a company performed a specific activity during a given time period, and answering them requires evidence from the public web. These queries are common in sales and consulting research. Because the work is repetitive and evidence-driven, it is a natural target for automation with AI agents.

We introduce **ask-web**, an agent that answers company-activity web-search queries with a static search-then-synthesize workflow. Given a query, the agent expands it into multiple search queries, retrieves candidate pages, ranks and crawls the most promising ones, and synthesizes a final answer from their contents.

To measure performance, we construct a benchmark of 612 company-activity queries with adjudicated ground truth. On this benchmark, we compare **ask-web** with commercial web-search agents, including ChatGPT, Claude, Exa, Parallel, and Perplexity. **ask-web** reaches 91.3% accuracy at an average cost of about 1.03 cents per query.

1 Introduction

Sales and consulting teams often need to know whether a company matches a specific activity profile before reaching out to it. For example, a salesperson offering relocation support may need to find companies that recently expanded into another country, while a consultant may need to check whether a prospective client was acquired by a private equity firm this year. We call these questions *company-activity web-search queries*: each asks whether a given company performed a given activity in a given timeframe, and the answer must be found on the public web. A human researcher answers such a query by issuing search-engine queries, opening promising results, and judging the evidence on those pages. The loop continues until the researcher finds concrete evidence or exhausts the search results. Because this work is slow and repetitive, we aim to automate it with AI agents while optimizing for accuracy and cost.

Our agent, **ask-web**, has two main functions: *choosing* which web pages to read and *synthesizing* an answer from their contents. This search-then-synthesize structure is common in commercial web-search systems because it is simple and effective. For example, Perplexity’s production search system retrieves and ranks web content before passing it to the answering model.¹ Exa’s **answer** endpoint “performs an Exa search and uses an LLM to generate” the final answer,² and its **deep** search mode expands a query into multiple parallel searches before “synthesizing results with citations.”³ Parallel’s deep-research tasks likewise conduct “multi-step web exploration” and then “synthesize findings” into a report.⁴

Unlike the human loop described above and unlike agentic deep-research systems, **ask-web** uses a *static, single-pass* workflow: it generates multiple search queries up front, retrieves and ranks the resulting pages once, and synthesizes the answer in one pass. This design gives up iterative search in exchange for lower cost and more predictable behavior.

To evaluate **ask-web** and compare it with commercial providers, we construct a benchmark of company-activity web-search queries with adjudicated ground truth. The benchmark tracks average accuracy and cost per query, which are the main production constraints for this workload; latency is not measured.

¹<https://research.perplexity.ai/articles/architecting-and-evaluating-an-ai-first-search-api>

²<https://exa.ai/docs/reference/answer>

³<https://exa.ai/blog/exa-deep>

⁴<https://docs.parallel.ai/task-api/examples/task-deep-research>

2 Benchmark

The benchmark is a collection of *(query, ground truth)* pairs. Each query is a company-activity web-search query, and the ground truth is its correct answer. We restrict the benchmark to *binary* queries: every prompt is phrased as a yes/no question, so every ground truth is either TRUE or FALSE. Prompts that require open-ended answers, such as a company’s current valuation, are excluded. Each query is represented as a pair *(prompt, company)*, where the prompt states the activity question and the company is the target. A sample is shown in Table 1.

Query		Ground Truth
Prompt	Company	
Determine whether the company currently has operations in Ontario, Canada, as of April 30, 2026	OpenAI	FALSE
Determine whether the organization had any divestiture, merger, acquisition, or strategic investment activity as of April 30, 2026	Snowflake	TRUE
Determine whether the company explicitly mentioned a new AI solution, AI feature, or AI product initiative in investor-facing or regulatory materials from November 1, 2024 through April 30, 2026	Danaher	TRUE

Table 1: Three example queries and their ground-truth answers.

We choose prompts and companies as follows.

- Companies are selected to cover major industries, including semiconductors, life sciences, IT/infrastructure, pharmaceuticals, financial services, and software-as-a-service (SaaS).
- Prompts are randomly sampled from Rox’s database. We remove prompts that are too specific or that usually do not have a clear answer.

We select 36 companies and 19 prompts, for a total of $36 \times 19 = 684$ candidate queries.

To label each query, we use a candidate-judge procedure:

1. **Candidates.** Six frontier web-search providers — **Exa**, **Parallel**, **Perplexity**, **Gemini**, **ChatGPT**, and **Claude Opus 4.8**, each with web-search tools enabled — independently answer every query. Each response contains a TRUE/FALSE answer, reasoning, and a supporting URL.
2. **Judges.** Three judges, each an instance of **Claude Opus 4.8**, independently review all candidate responses for a query and issue a verdict: TRUE, FALSE, or INCONCLUSIVE.
3. **Adjudication.** If all three judges return TRUE, or all three return FALSE, that unanimous verdict becomes the query’s ground truth. Otherwise, the query is marked *inconclusive* and excluded from the benchmark.

Of the 684 candidate queries, 72 are inconclusive, leaving 612 queries in the final benchmark, as shown in Table 14; the full prompt text behind each prompt code is listed in Table 13.

To evaluate an agent, we run it for multiple trials over the 612 queries; each (query, trial) evaluation is a *cell*. If a run errors — that is, if the agent fails to output TRUE or FALSE — the run is discarded and the cell is rerun. Discarded runs count toward neither accuracy nor cost. This protocol applies to every result and every agent, including **ask-web** and the commercial providers.

3 The ask-web Agent

ask-web implements the static, single-pass search-then-synthesize workflow described in the introduction.

3.1 Architecture

The workflow has five steps:

1. **Query Expansion.** An LLM (the *query expander*, abbreviated **qexp**) reads the input query (*prompt, company*) and generates **num-queries** Google search queries, called *subqueries*.
2. **Query Search.** Each subquery is sent to BrightData, a search API provider, which returns 10 URLs per subquery, each with a title and a short snippet. This gives at most $10 \times \text{num-queries}$ candidate URLs in total; some may be duplicates.
3. **URL Ranking.** A second LLM (the *URL ranker*, abbreviated **urlr**) reads all candidate URLs with their titles and snippets, and selects the **max-num-crawl** URLs most likely to answer the query.
4. **Crawl.** The selected URLs are crawled — by Tavily if the URL points to a PDF, and by BrightData otherwise — returning each page’s content. Each crawl result is preprocessed and trimmed to **trim-token** tokens.
5. **Synthesis.** A third LLM (the *synthesizer*, abbreviated **synt**) reads the query and the preprocessed crawl results, then outputs the final TRUE/FALSE answer.

The abbreviations **qexp**, **urlr**, and **synt** are used throughout the tables in this paper to denote the model chosen at each of the three LLM stages.

None of the three LLMs in steps 1, 3, and 5 has a web-search tool. Each receives only its system and user prompts and generates text. All web access happens through BrightData and Tavily in steps 2 and 4. The full pipeline is shown in Figure 1.

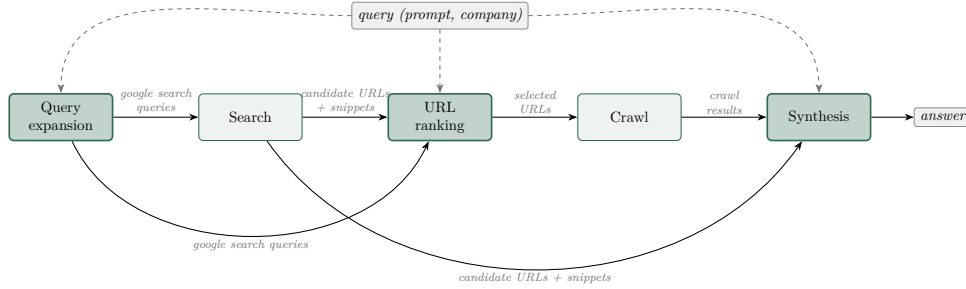


Figure 1: The **ask-web** dataflow. A module’s inputs may come from any upstream module, not only the previous one. Dashed edges carry the original query (reused by the three shaded LLM modules); solid edges carry produced data.

This workflow leaves the following variables to be chosen:

- **Models:** the three LLMs — the query expander, the URL ranker, and the synthesizer. Each can be a different model.
- **Prompts:** the system and user prompts of the three LLMs.
- **Hyperparameters:** **num-queries**, the number of subqueries generated per query; **max-num-crawl**, the number of URLs selected for crawling; and **trim-token**, the token limit of each crawl result. Tables abbreviate these as **n-q**, **n-cr**, and **trim**.
- **Synthesis preprocessing:** the crawl output format and how each crawl result is transformed before the synthesizer reads it; for example, Markdown passed through as-is, or raw HTML cleaned first.
- **Tolerant output parsing:** whether the synthesizer’s verdict is read leniently, as described in Section 3.3.

3.2 Chosen Configuration

Our chosen configuration for **ask-web** is as follows.

- **Query expander:** `gpt-5-0-mini` (medium reasoning effort).
- **URL ranker:** `gpt-oss-120b` (high reasoning effort), an open-source model that we self-host on Modal.
- **Synthesizer:** `gpt-oss-120b` (high reasoning effort), on the same self-hosted Modal deployment.
- **Prompts:** the `base` prompts for the query expander and the URL ranker; for the synthesizer, the `oss-minimal-notools` prompt. All prompts are in Appendix H.
- **Hyperparameters:** `num-queries = 5`, `max-num-crawl = 5`, `trim-token = 5000`.
- **Synthesis preprocessing:** crawl in raw HTML and clean it with a script.
- **Tolerant output parsing:** on.

Throughout the paper, **ask-web** refers to this chosen configuration. Any other configuration of the pipeline is an **ask-web** *variant*.

3.3 Fallback Policy

The following error-handling rules apply to every result in this paper.

- At the benchmark level, the rerun protocol of Section 2 applies to every agent, commercial or ours: a run that fails to produce a TRUE/FALSE verdict is discarded and the cell is rerun until it succeeds. This mainly compensates for timeouts caused by high evaluation concurrency.
- Internally, **ask-web** handles transient failures of its web providers, mirroring the production system’s policy:
 - every BrightData search call is retried with exponential backoff (waiting 5, 10, 20, then 40 seconds);
 - crawling falls back per URL from BrightData to Tavily when BrightData cannot serve the page; Tavily returns Markdown directly, bypassing the HTML-cleaning step. PDFs skip BrightData and go straight to Tavily, as described in step 4;
 - a web call surfaces as an errored run only after exhausting this retry-and-fallback chain.
 - LLM calls are not retried inside the pipeline: any failed call fails the run immediately and is handled by the benchmark-level rerun above.
 - Some open-source models, specifically `gpt-oss-120b`, are prompted without an enforced schema. By default, the synthesizer is called with a JSON schema that constrains its output to a bare TRUE/FALSE; however, some open-source models consistently fail under this constraint. Tolerant parsing lets the model reply in plain text and then extracts a TRUE/FALSE verdict from the response.

3.4 Result

We run **ask-web** on the benchmark for 5 trials, giving $5 \times 612 = 3060$ query evaluations. Calls to closed models are priced at the provider’s published per-token rates (Table 8). Calls to the self-hosted open-weight models, used for URL ranking and synthesis, are priced with the fitted per-token rates from the saturated-throughput measurements in Appendix A.2 (Table 10).

Table 2 compares **ask-web** against commercial web-search agents on our benchmark.

Table 2: **ask-web** vs. commercial web-search agents on **company-eval** (3 trials per commercial agent, 5 for **ask-web**).

Provider	Accuracy	Cost/query (¢)	Queries
ask-web	91.3%	1.03	3060
Exa	96.9%	11.4	1836
Exa (minimal effort)	95.5%	1.2	1836
ChatGPT 5.5 (medium reasoning effort)	96.9%	8.3	1836
Claude	94.7%	16.7	1836
Parallel	91.4%	2.5	1836
Perplexity	89.7%	0.9	1836

4 Optimization

With the five-step architecture fixed, choosing an agent configuration reduces to choosing the variables listed in the previous section: models, prompts, hyperparameters, and synthesis preprocessing. The objective is to maximize accuracy while minimizing cost per query.

We justify the chosen configuration with two analyses:

- **Local optimality.** We perturb the chosen configuration along each group of variables — hyperparameters, models, and synthesis preprocessing — and show that no perturbation gives a better accuracy–cost trade-off. Prompts are the one variable we do not perturb: whenever a model is swapped in, it uses the prompt variant tuned for that model (Appendix H).
- **Comparison against the global ceiling.** We estimate the best accuracy this architecture can reach by substituting the strongest available model into each pipeline step. The ceiling — about 94%, reached through the synthesizer — is within 3 percentage points of the chosen configuration, at roughly 15× the cost.

4.1 Local Perturbation Study

4.1.1 Method

We evaluate the following perturbations of the chosen configuration independently.

- **Hyperparameters.** We perform a grid search on the three hyperparameters:
 - `num-queries` = 3, 5, 8;
 - `max-num-crawl` = 2, 5, 8;
 - `trim-token` = 2500, 5000, 10000.

This gives 26 variants in total, excluding the original configuration. We run each variant on the benchmark for 3 trials.

- **Models.** We swap each model in the chosen configuration to cover the following four models: `gpt-5-0-nano` (high effort), `gpt-5-0-mini` (medium effort), `Qwen3.5-122B`, and `gpt-oss-120b` (high effort). We abbreviate them as `nano`, `mini`, `qwen`, and `oss`. This gives 9 variants in total. We run each variant on the benchmark for 3 trials.
- **Synthesis preprocessing.** We change the crawl output used for synthesis from cleaned raw HTML to Markdown. The variant is run for 5 trials.

4.1.2 Result

The grid search over the three hyperparameters is reported in Table 3 and visualized as an accuracy–cost trade-off in Figure 2.

Table 3: Grid search over the three **ask-web** hyperparameters: **n-q** = **num-queries**, **n-cr** = **max-num-crawl**, **trim** = **trim-token** (per-URL token limit); all other components fixed. Accuracy (%) over 3 trials (5 for the bold row); *Queries* = number of evaluated (query, trial) points. Bold row = chosen configuration.

n-q	n-cr	trim	Acc. (%)	Std (%)	Cost (c)	Queries
3	2	2.5k	86.2	2.0	0.71	1836
3	2	5k	87.5	1.3	0.72	1836
3	2	10k	88.4	1.2	0.73	1836
3	5	2.5k	87.6	0.8	0.82	1836
3	5	5k	88.9	0.6	0.83	1836
3	5	10k	89.4	0.7	0.86	1836
3	8	2.5k	89.6	1.1	0.90	1836
3	8	5k	91.2	0.3	0.93	1836
3	8	10k	90.8	1.2	0.95	1836
5	2	2.5k	90.4	0.9	0.90	1836
5	2	5k	90.9	0.9	0.90	1836
5	2	10k	91.0	1.2	0.92	1836
5	5	2.5k	91.2	0.5	1.01	1836
5	5	5k	91.3	1.1	1.03	3060
5	5	10k	91.8	0.5	1.06	1836
5	8	2.5k	91.5	0.8	1.10	1836
5	8	5k	92.8	0.2	1.12	1836
5	8	10k	92.0	0.7	1.17	1836
8	2	2.5k	90.1	0.9	1.12	1836
8	2	5k	90.0	1.0	1.11	1836
8	2	10k	91.1	0.7	1.13	1836
8	5	2.5k	91.9	1.4	1.24	1836
8	5	5k	91.8	0.7	1.25	1836
8	5	10k	93.1	0.6	1.28	1836
8	8	2.5k	93.0	0.7	1.34	1836
8	8	5k	93.2	0.4	1.37	1836
8	8	10k	91.6	1.2	1.41	1836

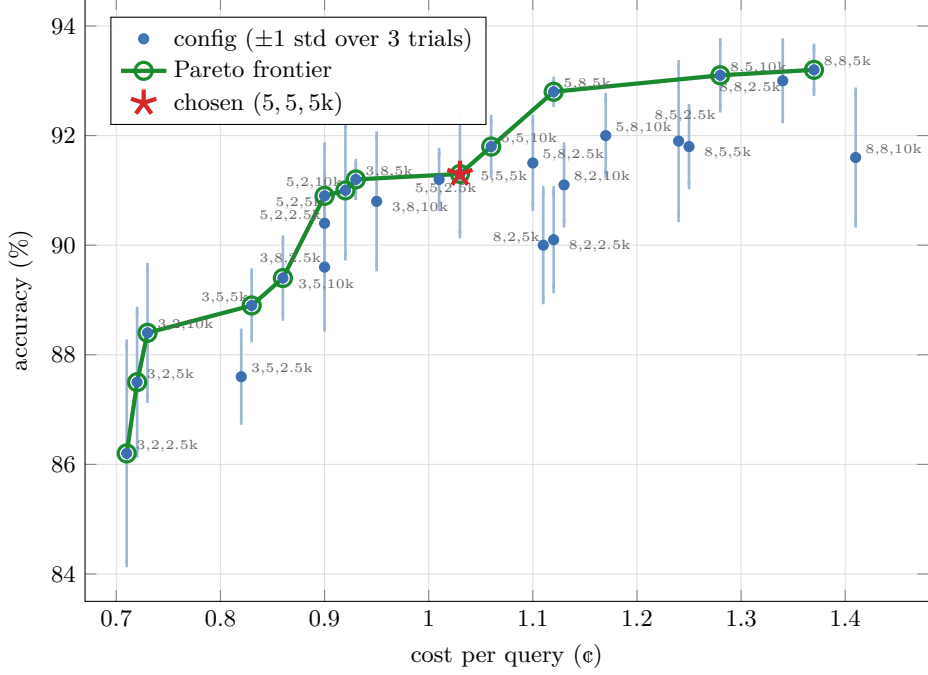


Figure 2: Accuracy versus cost per query for the hyperparameter grid search. Each point is one configuration; error bars = ± 1 standard deviation over 3 trials. Green line = Pareto frontier; red star = chosen configuration (5, 5, 5k).

The model perturbation results are reported in Table 4.

Table 4: Model perturbation of `ask-web`. Columns `qexp`, `urlr`, and `synt` are the models at the three stages (codes defined in the text). Bold row = chosen configuration; each later row swaps one stage (“–” = unchanged). Accuracy (%) over 3 trials (5 for the first row).

<code>qexp</code>	<code>urlr</code>	<code>synt</code>	Acc. (%)	Std (%)	Cost (€)	Queries
mini	oss	oss	91.3	1.1	1.03	3060
nano [†]	–	–	91.3	1.1	0.96	1836
qwen	–	–	91.3	1.5	1.14	1836
oss	–	–	87.3	1.4	0.69	1836
–	nano	–	89.7	2.3	1.30	1836
–	qwen	–	90.3	0.7	1.23	1836
–	mini	–	91.7	0.8	1.45	1836
–	–	nano	89.9	1.2	1.05	1836
–	–	mini	90.6	0.2	1.36	1836
–	–	qwen	89.3	0.3	1.31	1836

[†]Measured under a rerun-on-failure policy: $\sim 3.6\%$ of first attempts on this arm failed at the expansion stage (including 1.4% empty-query outputs); the chosen expander produced no failures. See the discussion.

The synthesis-preprocessing perturbation is reported in Table 5. Because both formats were run on the same (query, trial) points, we also compare them pairwise: McNemar’s exact test over the 3,060 paired points finds no significant difference (166 vs. 170 discordant points, $p = 0.87$).

Table 5: Synthesis-preprocessing perturbation: the crawl format used to build the synthesis input. Accuracy (%) over 5 trials; bold row = chosen configuration.

Crawl format	Acc. (%)	Std (%)	Cost (€)	Queries
HTML, cleaned by script	91.3	1.1	1.03	3060
Markdown, passed through as-is	91.4	1.3	1.06	3060

4.1.3 Discussion

Hyperparameters. The chosen configuration (5, 5, 5000) lies on the Pareto frontier of the grid (Figure 2). The grid contains more accurate points — for example, (5, 8, 5000) reaches 92.8% — but only at higher cost.

Models. No model swap significantly improves the accuracy–cost trade-off. All three synthesizer swaps reduce accuracy while costing the same or more. All URL-ranker swaps raise cost; the only accuracy gain, from `gpt-5-0-mini`, costs 40% more per query. At the query-expansion stage, replacing `gpt-5-0-mini` with `gpt-oss-120b` reduces accuracy, while replacing it with `qwen` preserves accuracy but increases cost.

The one seemingly favorable swap is `gpt-5-0-nano` (high reasoning effort) as the query expander: 91.3% at 0.96¢ vs. `ask-web`’s 91.3% at 1.03¢. However, the accuracy difference is not significant: paired McNemar over 1,836 points, 105 vs. 99 discordant, $p = 0.73$.

Synthesis preprocessing. The two crawl formats are statistically indistinguishable in accuracy ($p = 0.87$) and essentially tied in cost. We keep the cleaned-HTML format.

Summary. Every perturbation either reduces accuracy, is statistically indistinguishable from the chosen configuration, or raises cost. This supports the claim that `ask-web` is locally optimal under the accuracy–cost metric.

4.2 Comparison Against the Global Ceiling

The pipeline has three LLM-backed components: query expansion, URL ranking, and synthesis. To estimate the best accuracy this architecture can reach without cost constraints, we replace each component in turn with the strongest available model and measure the gain.

4.2.1 Component Substitution

We run three experiments, one per component. Each experiment measures the potential accuracy gain when that component is not cost-constrained. For the query expander and URL ranker, we swap the model to `xhigh`. For the synthesizer, we swap the model to `xhigh` and set `trim-token` = 10000. The results are reported in Table 6.

Table 6: Global-ceiling component substitutions; model codes as in Table 4. Bold row = chosen configuration; each later row upgrades one component (“–” = unchanged). The synthesizer row doubles `trim` to 10k so the stronger reader is not input-limited; the last row is the trim-matched control from Table 3. The `xhigh` rows are single trials (no standard deviation).

qexp	urlr	synt	trim	Acc. (%)	Std (%)	Cost (¢)	Queries
mini	oss	oss	5k	91.3	1.1	1.03	3060
<code>xhigh</code>	–	–	5k	92.0	–	4.59	612
–	<code>xhigh</code>	–	5k	93.5	–	16.16	612
–	–	<code>xhigh</code>	10k	94.1	–	15.31	612
–	–	–	10k	91.8	0.5	1.06	1836

Every substitution raises the cost per query by an order of magnitude (4–16×), so none is a production candidate.

Some substitutions still produce meaningful accuracy gains. Upgrading the query expander adds +0.7pp (92.0%), and upgrading the URL ranker adds +2.2pp (93.5%). The largest gain comes from synthesis: the `xhigh` synthesizer with a 10000-token input reaches 94.1%, the highest accuracy of any configuration in this paper. This suggests that synthesis is the main bottleneck in the `ask-web` pipeline.

4.3 Synthesis Model Variants

Motivated by the ceiling analysis, we examine the synthesis step more closely.

4.3.1 Method

To isolate the effect of synthesis, we freeze the earlier pipeline stages and evaluate synthesizers directly.

In the resulting benchmark, **url-eval**, each provider receives a fixed input consisting of the prompt and at most **max-num-crawl** crawl results, then outputs a TRUE/FALSE verdict. Each crawl result is extracted using BrightData, or Tavily for PDFs, and trimmed to 10,000 tokens, matching **ask-web**’s synthesis step.

We construct **url-eval** as follows.

1. Uniformly sample n (query, ground truth) pairs from the main benchmark.
2. For each sampled query, run the **ask-web** pipeline m times to generate m tuples of 5 crawl results, each trimmed to 10,000 tokens.
3. This yields $n \times m$ **url-eval** queries, each a tuple of 5 crawl results; the ground truth carries over from the original benchmark query.

Because LLM outputs are stochastic, we evaluate a provider by sampling its verdict t times on each query — $n \times m \times t$ calls in total — and report the fraction of correct verdicts.

We use $n = 40$, $m = 8$, and $t = 5$. The sampled queries and their retrieved URLs are listed in Table 15. The evaluated providers are **nano**, **xhigh**, **oss**, **qwen**, and **qwen(max)**, standing for **gpt-5-0-nano** (high reasoning effort), **GPT-5.5** (**xhigh** reasoning effort), **gpt-oss-120b** on our Modal deployment, **Qwen3.5-122B** on our Modal deployment, and **Qwen3.7-Max** on Together.

We use the same fallback policy for **url-eval**: if a provider fails to output a valid answer, we rerun the query until it succeeds. This compensates for model timeouts caused by high evaluation concurrency.

4.3.2 Result

Table 7 reports overall accuracy by provider. Table 16 gives the main **url-eval** summary: five synthesizer rows and forty (query, company) columns (Q1–Q40), with each cell shaded by mean accuracy over the eight trials. Per-trial breakdowns are in Table 17.

Table 7: **url-eval** full benchmark: overall synthesis accuracy per provider over all 320 trials. All providers read the same retrieved URLs; only the synthesizer differs.

Provider (synthesizer)	Accuracy
xhigh (GPT-5.5)	97.1%
qwen (Qwen3.5-122B)	95.6%
qwen(max) (Qwen3.7-Max)	94.7%
nano (GPT-5-nano)	94.3%
oss (gpt-oss-120b)	93.5%

4.3.3 Discussion

xhigh performs significantly better than the other providers, including the current **ask-web** synthesizer, **oss**. This initially suggests a possible improvement path: a model with **oss**-level cost but stronger synthesis accuracy.

We therefore inspect model traces to understand why **xhigh** performs better. The full trace analysis is presented in Appendix B, together with a harder follow-up benchmark that magnifies the gap between **xhigh** and weaker models.

The trace analysis suggests that such a model is unlikely, for two reasons.

- **xhigh** has stronger world knowledge, giving it useful priors when the crawl evidence is incomplete.
- **xhigh** performs deeper analysis over the crawl input, allowing it to extract evidence that weaker models miss.

The first property generally requires more pretraining and more parameters. The second requires longer reasoning traces. Both increase cost.

4.3.4 Conclusion

xhigh remains the strongest synthesizer because it combines deeper reasoning over the crawl input with stronger priors for useful inferences. Models with both properties are expensive, due to model size and reasoning-token cost. We therefore keep **oss** as the **ask-web** synthesizer.

5 Conclusion

We introduced **ask-web**, a static, single-pass agent for company-activity web-search queries. It expands the query, retrieves and ranks candidate pages, crawls the most promising ones, and synthesizes a TRUE/FALSE verdict from their contents. Because each stage is modular, models and hyperparameters can be tuned independently to trade accuracy against cost. To evaluate the design, we built a benchmark of 612 company-activity queries with adjudicated ground truth, tracking both accuracy and cost per query.

Our chosen configuration of **ask-web** reaches 91.3% accuracy at about 1.03 cents per query, which is cheaper than the commercial agents of comparable accuracy (Table 2).

We use two analyses to support the chosen configuration. First, local perturbation shows that no nearby change improves the accuracy–cost trade-off. Second, global ceiling analysis shows that this architecture reaches 94.1% accuracy when the strongest available model, **xhigh**, is used at the synthesis step with a longer crawl input. Further trace analysis suggests that the gap comes from **xhigh**’s stronger world knowledge and deeper reasoning over crawl evidence. These capabilities are costly because they require larger models and more reasoning tokens.

Overall, **ask-web** occupies a favorable point on the accuracy–cost frontier for this production workload.

A Cost Calculation

A.1 API-Billed Models and Services

For every externally billed service, we use the provider’s published prices, reproduced in Table 8. Services differ in how they are billed.

- Closed-source LLMs charge per token. Reasoning tokens are billed as output tokens, and cached input tokens are billed at the cheaper cached rate.
- Hosted web-search tools charge a flat fee per search, on top of the token cost.
- Web-search/crawl services and commercial agents charge per request.

Exa and Perplexity are exceptions: each of their responses reports its actual dollar cost, so we record that reported cost directly instead of computing it from a rate.

A.2 Self-Hosted Open-Source Models

The open-source models in this experiment are *self-hosted* on Modal, a serverless GPU cloud. Modal bills per GPU hour, and GPU usage depends on token counts. This relationship is not linear: the model generates autoregressively, with each new token attending to the tokens before it, so GPU time grows roughly quadratically with sequence length. We therefore estimate token cost with a quadratic regression.

1. Mix sampling. A *mix* is a collection of pairs (input token count, output token count). Each mix becomes one data point in the price regression. We build 15 mixes per model.

- Three *real* mixes, one per pipeline stage (query expansion, URL ranking, synthesis). Their pairs are sampled uniformly from the recorded input/output token counts of that stage’s calls in finished benchmark runs.
- Twelve *synthetic* mixes with fixed token counts: every pair in the mix is identical. The shapes cover prefill-heavy (1000×60 up to 64000×60), decode-heavy (300×500 up to 300×3000), and mixed (4000×2000 up to 32000×2500).

Table 8: Prices of the API-billed models and services used in this paper, as recorded by the benchmark’s cost accounting (July 2026).

	Input (\$/Mtok)	Cached input (\$/Mtok)	Output (\$/Mtok)	Per request (\$)
<i>Closed-source LLMs (reasoning tokens billed as output)</i>				
gpt-5-nano (synthesis)	0.05	0.005	0.40	–
gpt-5-mini (query expansion, URL ranking)	0.25	0.025	2.00	–
gpt-5.5 ($\leq 272k$ context, ChatGPT/ xhigh)	5.00	0.50	30.00	–
gpt-5.5 ($> 272k$ context)	10.00	1.00	45.00	–
Claude Opus 4.8 (agent, candidate, judges)	5.00	0.50	25.00	–
Gemini 3 Flash (candidate)	0.50	0.05	3.00	–
<i>Hosted web-search tool surcharges (billed on top of tokens)</i>				
OpenAI web_search (per call)	–	–	–	0.010
Anthropic web_search (per search)	–	–	–	0.010
Gemini Google-Search grounding (per search query)	–	–	–	0.014
<i>Web search and crawl services (used by ask-web)</i>				
BrightData search (per SERP call)	–	–	–	0.00035
BrightData crawl (per URL)	–	–	–	0.00025
Tavily crawl (PDF URLs only, per URL)	–	–	–	0.00066
<i>Commercial agents (Table 2)</i>				
Exa /answer	–	–	–	0.005
Exa agent (auto effort)	–	–	–	reported per run
Exa agent (minimal effort)	–	–	–	0.012
Parallel Task API (core processor)	–	–	–	0.025
Perplexity (sonar)	–	–	–	reported per call

For example, the synthesis mix (real) and the 8000×1200 mix (synthetic) start with

$$\begin{aligned} \text{synthesis} &: [(21379, 678), (23700, 1018), (35266, 673), (19339, 695), (41303, 721), \dots] \\ 8000 \times 1200 &: [(8000, 1200), (8000, 1200), (8000, 1200), (8000, 1200), (8000, 1200), \dots]. \end{aligned}$$

The synthetic mixes are needed because the three real mixes have similar input:output ratios; a regression over them alone cannot separate the input price from the output price.

2. Running a mix. For each pair (**in**, **out**) in the mix, we generate a prompt of exactly **in** random-word tokens, send it to the model, and let the model generate exactly **out** tokens. We then run a fixed 180-second window on a single pinned GPU container with 64 requests in flight at all times. We measure beforehand that 64 concurrent requests saturate the GPU for both models. Finally, we record how many input and output tokens are processed in the window.

Random words are sufficient here because throughput depends on token counts, not on the content of the prompt.

3. Price per mix. From each window, we record the totals processed by the model: $\sum \text{in}$, $\sum \text{out}$, $\sum \text{in}^2$, $\sum \text{out}^2$, and $\sum \text{in} \cdot \text{out}$ (Table 9). The window’s price follows from the GPU rate: $\$4.56/\text{hr} \times 180\text{s}/3600 \approx \0.228 . Each mix is therefore one regression data point: token totals on one side, and the $\$0.228$ cost on the other.

4. Quadratic regression. We regress window cost on token counts with the quadratic model

$$\text{cost } (\$/\text{call}) = \frac{a_1 \text{ in} + b_1 \text{ out}}{10^6} + \frac{a_2 \text{ in}^2 + b_2 \text{ out}^2 + c (\text{in} \cdot \text{out})}{10^{12}} + \frac{d}{10^3},$$

We fit this model with non-negative least squares (NNLS). The fitted prices are given in Table 10. Applying these prices to the ranking and synthesis token profiles yields the per-query self-hosting costs quoted for **ask-web** in Table 2.

Figures 3 and 4 show each fitted quadratic surface over the request-shape domain (input $\times$ output tokens), with the 15 measured saturation windows overlaid. For **gpt-oss-120b**, the surface curves

Table 9: The 15 cost-fit windows per model: the data points of the price regression. Each window ran 180s at concurrency 64 on the pinned container; Price = the window’s measured GPU cost (\$); n = completed requests; $\sum \text{in} / \sum \text{out}$ = token totals (ktok); $\sum \text{in}^2 / \sum \text{out}^2 / \sum \text{in} \cdot \text{out}$ = squared-token sums (Gtok²), the regressors of the quadratic terms. The first three mixes replay stage profiles; the other twelve are synthetic shapes.

Mix	n	$\sum \text{in}$ (ktok)	$\sum \text{out}$ (ktok)	$\sum \text{in}^2$ (Gtok ²)	$\sum \text{out}^2$ (Gtok ²)	$\sum \text{in} \cdot \text{out}$ (Gtok ²)	Price (\$)
gpt-oss-120b							
query expansion	267	207	704	0.16	1.9	0.55	0.228
URL ranking	97	604	443	3.8	2.2	2.8	0.228
synthesis	96	3,005	131	114	0.28	3.8	0.228
1000 × 60	8,694	8,694	522	8.7	0.03	0.52	0.228
6000 × 60	2,543	15,258	153	91.5	0.0092	0.92	0.228
16000 × 60	913	14,608	55	234	0.0033	0.88	0.228
32000 × 60	375	12,000	23	384	0.0014	0.72	0.228
64000 × 60	144	9,216	9	590	0.0005	0.55	0.228
300 × 500	1,182	355	591	0.11	0.30	0.18	0.228
300 × 1500	439	132	659	0.04	0.99	0.20	0.228
300 × 3000	241	72	723	0.02	2.2	0.22	0.228
4000 × 2000	260	1,040	520	4.2	1.0	2.1	0.228
8000 × 1200	306	2,448	367	19.6	0.44	2.9	0.228
16000 × 2500	129	2,064	323	33.0	0.81	5.2	0.228
32000 × 2500	67	2,144	168	68.6	0.42	5.4	0.228
qwen3-5-122b-a10b-fp8							
query expansion	89	69	220	0.05	0.56	0.17	0.228
URL ranking	49	305	199	1.9	0.94	1.2	0.228
synthesis	79	2,570	96	99.1	0.15	3.1	0.228
1000 × 60	2,376	2,376	143	2.4	0.0086	0.14	0.228
6000 × 60	700	4,200	42	25.2	0.0025	0.25	0.228
16000 × 60	289	4,624	17	74.0	0.0010	0.28	0.228
32000 × 60	170	5,440	10	174	0.0006	0.33	0.228
64000 × 60	86	5,504	5	352	0.0003	0.33	0.228
300 × 500	496	149	248	0.04	0.12	0.07	0.228
300 × 1500	136	41	204	0.01	0.31	0.06	0.228
300 × 3000	83	25	249	0.0075	0.75	0.07	0.228
4000 × 2000	116	464	232	1.9	0.46	0.93	0.228
8000 × 1200	152	1,216	182	9.7	0.22	1.5	0.228
16000 × 2500	56	896	140	14.3	0.35	2.2	0.228
32000 × 2500	60	1,920	150	61.4	0.38	4.8	0.228

Table 10: Fitted quadratic per-token prices for the self-hosted models (GPU \$4.56/hr). Linear terms in \$/Mtok; quadratic terms in \$/Ttok²; per-request bias in \$/10³ requests. “Mean |err|” = mean absolute residual over the 15 windows (linear-only fit: 25.4% and 7.5%).

Model	Input (\$/Mtok)	Output (\$/Mtok)	Bias (\$/k req)	in ²	out ²	in-out	Mean err (%)
gpt-oss-120b	0.0083	0.3185	0.000	0.249	0.000	26.01	6.2
qwen3-5-122b-a10b-fp8	0.0421	0.9197	0.000	0.000	23.36	5.02	6.7

up sharply for long-prompt, long-generation requests, driven by the in · out attention term. For qwen3-5-122b, the cost is nearly planar over the measured range.

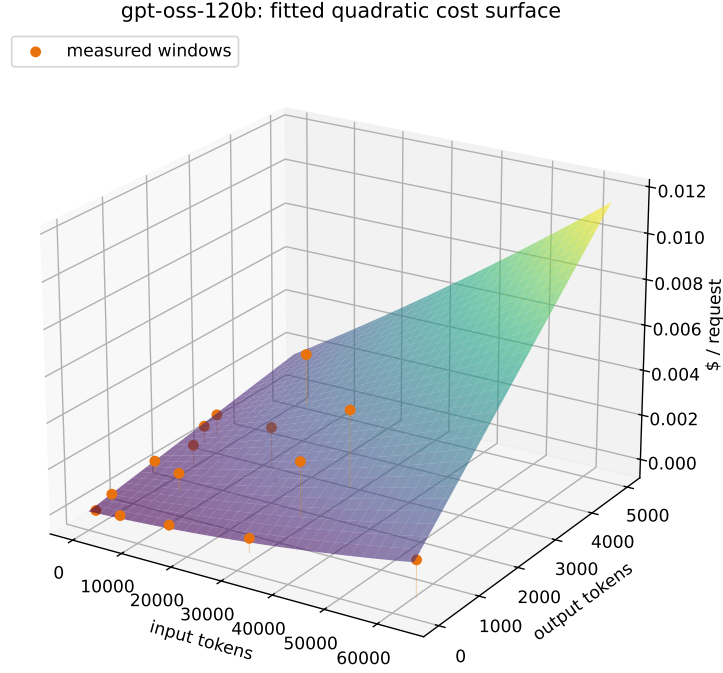


Figure 3: gpt-oss-120b: fitted per-request cost surface (\$/request) over input and output tokens. Orange points = the 15 measured windows.

B Synthesis Capability Analysis

We first inspect traces from the main url-eval run (Section 4.3) to identify why xhigh outperforms nano, oss, qwen, and qwen(max). The examples below illustrate the main failure modes.

- ai-initiative, Danaher; ground truth True; trials 1, 8**
Average accuracy: nano 20% ; xhigh 0% ; oss 0% ; qwen 0% ; qwen(max) 0% .
 The crawl results do not provide enough information.
- ai-initiative, Danaher; ground truth True; trial 5**
Average accuracy: nano 20% ; xhigh 100% ; oss 0% ; qwen 0% ; qwen(max) 0% .
 xhigh has the prior knowledge that Innovaccer is a healthcare-AI company; its full chain-of-thought is reproduced in Listing 13. Other providers do not have that prior and thus answer incorrectly.
- canada-power-signals, Cisco; ground truth False; trial 1**
Average accuracy: nano 40% ; xhigh 100% ; oss 80% ; qwen 100% ; qwen(max) 100% .
 The failing models lose focus on the company Cisco: the power signals come from Newfoundland & Labrador Hydro, not from Cisco.
- geographic-expansion, Dell Technologies; ground truth True; trial 7**
Average accuracy: nano 60% ; xhigh 0% ; oss 20% ; qwen 0% ; qwen(max) 0% .
 Based solely on the crawl results, the answer should be false (incorrect) because there is not sufficient information. The models that answer true (correct) are actually weaker, as they count “a soft, non-binding government MoU (training/knowledge-transfer) in an existing market” as geographic expansion although it is clearly not (Listing 14).
- ontario-operations, Workday; ground truth True; trial 7**
Average accuracy: nano 80% ; xhigh 100% ; oss 0% ; qwen 100% ; qwen(max) 100% .

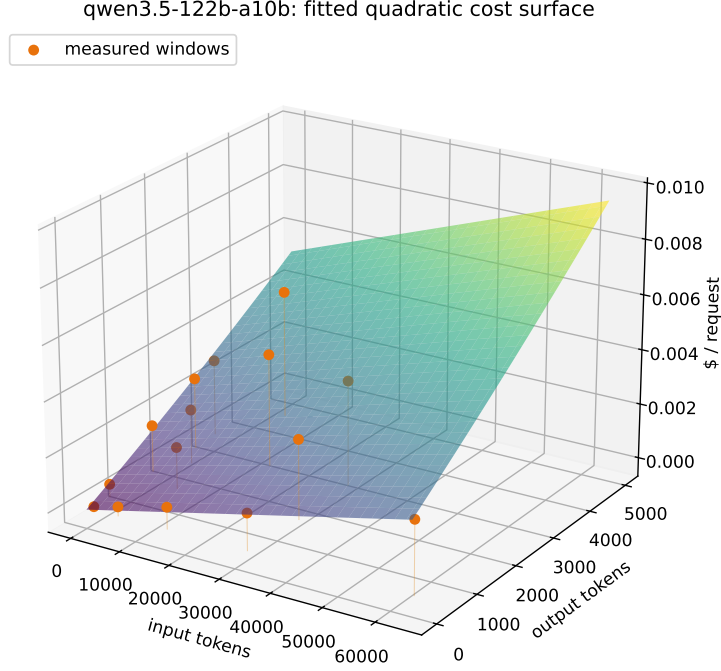


Figure 4: qwen3-5-122b-a10b-fp8: fitted per-request cost surface (\$/request) over input and output tokens. Orange points = the 15 measured windows.

The crawl contains no evidence to justify a TRUE answer. However, stronger synthesizers answer true because they know Workday has a Toronto office (Listing 15).

To focus on cases where weaker models fail, we build a harder benchmark using the same `url-eval` structure. Instead of sampling uniformly, we select queries on which weaker models perform poorly. The selected queries, crawl results, and ground truth are shown in Table 18. The benchmark has only 6 problems, but each has 16 crawl-result tuples. Per-query accuracy is summarized in Table 11, with per-trial breakdowns in Table 19.

Table 11: `url-eval` mini benchmark: mean synthesis accuracy (%) per query cell (legend below); **ovr** = mean over the six cells.

provider	Q1	Q2	Q3	Q4	Q5	Q6	ovr
nano	51	86	44	34	36	64	52
xhigh	91	99	98	82	48	100	86
oss	51	92	64	24	39	86	59
qwen	51	100	81	25	44	94	66
qwen(max)	52	100	97	15	52	88	67

Legend: Q1=c-suite-change/Smartsheet; Q2=c-suite-change/GE Vernova; Q3=canada-power-signals/Medline; Q4=geographic-expansion/Hewlett Packard Enterprise; Q5=investment-announcement/Discover Financial Services; Q6=ma-activity/Intel.

This run uses the same provider slate as the main `url-eval`: `nano`, `xhigh`, `oss`, `qwen`, and `qwen(max)`. On this harder sample, the field separates sharply. Only `xhigh` performs well: 86% overall, 17 percentage points ahead of the next provider. `qwen` and `qwen(max)` form the middle band, while `oss` and `nano` are weakest. `nano`’s low score is expected because the sample is constructed from queries it fails. We now examine the appendix traces query by query.

1. c-suite-change, Smartsheet; ground truth True; trials 4, 9–13

Average accuracy: nano 0% ; xhigh 93% ; oss 0% ; qwen 0% ; qwen(max) 0% .

The direct evidence in the crawl result is either too early or too late, so most models answer

incorrectly. **xhigh**, by contrast, infers from an embedded URL in the crawl result that a C-suite change occurred within the time frame. **oss** performs poorly here.

2. **c-suite-change, Smartsheet; ground truth False; trial 16**

Average accuracy: nano 20% ; **xhigh** 100% ; **oss** 0% ; **qwen** 20% ; **qwen(max)** 0% .

A page published on 5 May 2026 mentions a “C-suite change” happening a week earlier, so the change falls between Jan 31 and Apr 30, as the query requires. Weaker models fail to infer that this relative date lies within the time frame.

3. **c-suite-change, GE Vernova; ground truth False; trial 1**

Average accuracy: nano 0% ; **xhigh** 80% ; **oss** 40% ; **qwen** 100% ; **qwen(max)** 100% .

The incorrect models are confused by the time frame: the source reports a C-suite change in January, but the window is Jan 31–Apr 30, so this source should not count as evidence — yet the incorrect models treat it as such.

4. **canada-power-signals, Medline; ground truth False; trials 6–10, 15, 16**

Average accuracy: nano 0% ; **xhigh** 97% ; **oss** 29% ; **qwen** 57% ; **qwen(max)** 97% .

The source mentions Canada power signals but never mentions Medline. The incorrect models ignore the company constraint and answer “true”.

5. **geographic-expansion, Hewlett Packard Enterprise; ground truth True; all trials**

Average accuracy: nano 34% ; **xhigh** 82% ; **oss** 24% ; **qwen** 25% ; **qwen(max)** 15% .

xhigh’s reasoning is based on an article about HPE closing its \$14B Juniper Networks acquisition on July 2, 2025. Only **xhigh** interprets this as geographic expansion. The other models look for a more literal match, whereas **xhigh** reads the acquisition as expansion and uses the prior that Juniper operates worldwide.

6. **investment-announcement, Discover Financial Services; ground truth False; trial 9**

Average accuracy: nano 0% ; **xhigh** 60% ; **oss** 0% ; **qwen** 20% ; **qwen(max)** 60% .

This query has a trap: a [Yahoo article](#) from Feb 24, 2026 suggests there was a “proposed deal,” but another [document](#) shows the deal predates the time frame.

7. **ma-activity, Intel; ground truth True; trial 10**

Average accuracy: nano 0% ; **xhigh** 100% ; **oss** 0% ; **qwen** 100% ; **qwen(max)** 40% .

The information is not stated in plain English but appears in SEC iXBRL tables; **xhigh** is able to recover it from those tables.

These examples show why **xhigh** outperforms the other providers:

- In Trace 1, **xhigh** is the only provider that analyzes an embedded URL to infer the time frame.
- In Trace 2, **xhigh** performs a date calculation to infer information not stated directly in the input, showing that it reasons carefully over details.
- In Trace 5, **xhigh** draws on a sharp world model.
- In Trace 6, **xhigh** is more careful about conflicting evidence.
- In the other trials, **xhigh** is simply better at extracting the relevant information, while the other models ignore otherwise obvious details.

From these examples, we conclude that **xhigh** does better because it combines stronger reasoning with a stronger world model.

C Additional Results

During development, we ran many exploratory **ask-web** variants to understand the design space. These runs informed our decisions but are not central to the main results, so we report them here only in summary form. Table 12 lists every variant across all experiments, aggregated by variant: its full configuration, pooled accuracy with per-trial breakdown and standard deviation, and cost per query on the benchmark.

Per-query results and full model traces are omitted for space. For qualitative insight, we embed representative traces directly in this report: Appendix E walks through a complete single-query trace — the query-expansion, URL-ranking, and synthesis inputs and outputs of two agent variants on the same query — and Appendix G reproduces a full raw synthesis input. We keep these examples in the document rather than linking to hosted traces, since hosted trace links can expire. The complete trace set is large, so we provide additional traces on request.

Table 12: Every ask-web variant run on company-eval, aggregated by variant (all trials pooled). qexp/urlr/synt are the stage models; standard codes, plus oss-m = oss at medium effort and qwn397 = Qwen397B. Hyperparameters as in Table 3. Crawl: md = BrightData Markdown, html = raw HTML cleaned by script. Prompt codes are defined in Appendix H; Tol. = tolerant output parsing. T1–T5 are per-trial accuracies (%) over the 612 queries; “–” = not run; n = scored (query, trial) points.

Models			Hyperparams				Prompts			Results										n	Note
qexp	urlr	synt	n-q	n-cr	trim	Crawl	qexp	synt	Tol.	Acc.	Std	T1	T2	T3	T4	T5	Cost/q (€)				
Chosen configuration																					
mini	oss	oss	5	5	5k	html	base	min-nt	on	91.3	1.1	90.4	90.0	92.5	92.2	91.5	1.03	3060	Chosen configuration: ask-web as specified in Section 3		
Hyperparameter grid around the chosen configuration (the (5, 5, 5k) point is the bold row above)																					
mini	oss	oss	3	2	2.5k	html	base	min-nt	on	86.2	2.0	84.0	86.8	87.9	–	–	0.71	1836			
mini	oss	oss	3	2	5k	html	base	min-nt	on	87.5	1.3	86.6	86.9	89.1	–	–	0.72	1836			
mini	oss	oss	3	2	10k	html	base	min-nt	on	88.4	1.2	89.5	88.6	87.1	–	–	0.73	1836			
mini	oss	oss	3	5	2.5k	html	base	min-nt	on	87.6	0.8	86.8	87.7	88.4	–	–	0.82	1836			
mini	oss	oss	3	5	5k	html	base	min-nt	on	88.9	0.6	89.4	88.2	89.1	–	–	0.83	1836			
mini	oss	oss	3	5	10k	html	base	min-nt	on	89.4	0.7	90.0	89.5	88.6	–	–	0.86	1836			
mini	oss	oss	3	8	2.5k	html	base	min-nt	on	89.6	1.1	88.7	89.2	90.8	–	–	0.90	1836			
mini	oss	oss	3	8	5k	html	base	min-nt	on	91.2	0.3	91.3	91.5	90.8	–	–	0.93	1836			
mini	oss	oss	3	8	10k	html	base	min-nt	on	90.8	1.2	92.2	90.5	89.9	–	–	0.95	1836			
mini	oss	oss	5	2	2.5k	html	base	min-nt	on	90.4	0.9	89.4	90.5	91.2	–	–	0.90	1836			
mini	oss	oss	5	2	5k	html	base	min-nt	on	90.9	0.9	91.3	89.9	91.5	–	–	0.90	1836			
mini	oss	oss	5	2	10k	html	base	min-nt	on	91.0	1.2	89.9	90.7	92.3	–	–	0.92	1836			
mini	oss	oss	5	5	2.5k	html	base	min-nt	on	91.2	0.5	90.7	91.3	91.7	–	–	1.01	1836			
mini	oss	oss	5	5	10k	html	base	min-nt	on	91.8	0.5	92.3	91.8	91.3	–	–	1.06	1836			
mini	oss	oss	5	8	2.5k	html	base	min-nt	on	91.5	0.8	91.0	91.0	92.5	–	–	1.10	1836			
mini	oss	oss	5	8	5k	html	base	min-nt	on	92.8	0.2	92.8	93.0	92.6	–	–	1.12	1836			
mini	oss	oss	5	8	10k	html	base	min-nt	on	92.0	0.7	92.0	92.6	91.3	–	–	1.17	1836			
mini	oss	oss	8	2	2.5k	html	base	min-nt	on	90.1	0.9	90.8	90.4	89.1	–	–	1.12	1836			
mini	oss	oss	8	2	5k	html	base	min-nt	on	90.0	1.0	90.3	90.8	88.9	–	–	1.11	1835			
mini	oss	oss	8	2	10k	html	base	min-nt	on	91.1	0.7	90.4	91.0	91.8	–	–	1.13	1836			
mini	oss	oss	8	5	2.5k	html	base	min-nt	on	91.9	1.4	92.3	90.4	93.0	–	–	1.24	1836			
mini	oss	oss	8	5	5k	html	base	min-nt	on	91.8	0.7	91.8	91.2	92.5	–	–	1.25	1836			
mini	oss	oss	8	5	10k	html	base	min-nt	on	93.1	0.6	93.6	92.5	93.3	–	–	1.28	1836			
mini	oss	oss	8	8	2.5k	html	base	min-nt	on	93.0	0.7	92.5	93.8	92.8	–	–	1.34	1836			
mini	oss	oss	8	8	5k	html	base	min-nt	on	93.2	0.4	93.6	92.8	93.1	–	–	1.37	1836			
mini	oss	oss	8	8	10k	html	base	min-nt	on	91.6	1.2	92.2	92.5	90.2	–	–	1.41	1836			
Per-stage model swaps around the chosen configuration																					
nano	oss	oss	5	5	5k	html	base	min-nt	on	91.3	1.1	92.5	90.8	90.5	–	–	0.96	1836			
qwen	oss	oss	5	5	5k	html	base	min-nt	on	91.3	1.5	92.5	89.7	91.8	–	–	1.14	1836			
oss	oss	oss	5	5	5k	html	base	min-nt	on	87.3	1.4	85.9	87.1	88.7	–	–	0.69	1836			
mini	qwen	oss	5	5	5k	html	base	min-nt	on	90.3	0.7	89.5	90.2	91.0	–	–	1.23	1836			
mini	mini	oss	5	5	5k	html	base	min-nt	on	91.7	0.8	92.6	91.2	91.2	–	–	1.45	1836			
mini	nano	oss	5	5	5k	html	base	min-nt	on	89.7	2.3	91.8	90.0	87.3	–	–	1.30	1836			
mini	oss	qwen	5	5	5k	html	base	base	–	89.3	0.3	89.1	89.7	89.2	–	–	1.31	1836			
mini	oss	mini	5	5	5k	html	base	base	–	90.6	0.2	90.4	90.7	90.7	–	–	1.36	1836			
mini	oss	nano	5	5	5k	html	base	base	–	89.9	1.2	90.8	88.6	90.4	–	–	1.05	1836			

continued on the next page

Table 12: Exploratory ask-web variants (continued).

Models			Hyperparams				Prompts			Results											n	Note
qexp	urlr	synt	n-q	n-cr	trim	Crawl	qexp	synt	Tol.	Acc.	Std	T1	T2	T3	T4	T5	Cost/q (€)					
Global ceiling: one stage on gpt-5.5 (xhigh reasoning effort)																						
xhigh	oss	oss	5	5	5k	html	base	min-nt	on	92.0	–	92.0	–	–	–	–	4.59	612				
mini	xhigh	oss	5	5	5k	html	base	min-nt	on	93.5	–	93.5	–	–	–	–	16.16	612				
mini	oss	xhigh	5	5	5k	html	base	base	–	92.0	–	92.0	–	–	–	–	12.51	612				
mini	oss	xhigh	5	5	10k	html	base	base	–	94.1	–	94.1	–	–	–	–	15.31	612				
Snippet fallback on the chosen configuration (both crawl formats)																						
mini	oss	oss	5	5	5k	html	base	min-nt	on	92.3	0.1	92.2	92.3	92.3	–	–	1.02	1836	snippet fallback on			
mini	oss	oss	5	5	5k	md	base	min-nt	on	90.1	0.8	89.7	91.0	89.5	–	–	1.06	1836	snippet fallback on			
Two self-hosted stages: candidate arms toward the chosen configuration																						
mini	oss	qwen	5	5	5k	md	base	base	–	90.0	–	90.0	–	–	–	–	1.19	612				
mini	oss	oss	5	5	5k	md	base	base	–	75.5	–	75.5	–	–	–	–	1.04	612				
mini	oss	oss	5	5	5k	html	base	base	–	76.1	–	76.1	–	–	–	–	1.01	612				
mini	oss	oss	5	5	5k	md	base	base	on	91.1	1.0	91.8	91.7	89.7	91.2	–	1.05	2448				
mini	oss	oss	5	5	5k	html	base	base	on	90.8	2.3	93.5	91.8	88.2	89.5	–	1.01	2448				
mini	oss	oss	5	5	5k	md	base	base-nt	on	100.0	–	100.0	–	–	–	–	1.11	7	partial run: only 7 of 612 cells completed (arm superseded), so the accuracy is not meaningful no completed runs at the current config version			
mini	oss	oss	5	5	5k	html	base	base-nt	on	–	–	–	–	–	–	–	–	–				
mini	oss	oss	5	5	5k	html	base	oss	on	89.8	1.3	88.1	90.2	91.2	89.9	–	1.02	2448				
mini	oss	oss	5	5	5k	html	base	min	on	91.3	0.8	90.8	92.0	90.5	92.0	–	1.02	2448				
mini	oss	oss	5	5	5k	md	base	min-nt	on	91.4	1.3	92.2	91.5	92.5	91.8	89.2	1.06	3060				
One self-hosted stage on the production configuration																						
oss-m	mini	nano	5	5	10k	md	base	base	–	87.4	–	87.4	–	–	–	–	1.09	612				
oss	mini	nano	5	5	10k	md	base	base	–	85.1	–	85.1	–	–	–	–	1.14	612				
oss	mini	nano	5	5	5k	md	base	base	–	85.6	–	85.6	–	–	–	–	1.11	612				
qwen	mini	nano	5	5	5k	md	base	base	–	86.8	–	86.8	–	–	–	–	1.57	612				
mini	oss-m	nano	5	5	10k	md	base	base	–	88.4	–	88.4	–	–	–	–	1.03	612				
mini	oss	nano	5	5	10k	md	base	base	–	90.0	–	90.0	–	–	–	–	1.12	612				
mini	oss	nano	5	5	5k	md	base	base	–	88.4	–	88.4	–	–	–	–	1.08	612				
mini	qwen	nano	5	5	5k	md	base	base	–	89.2	–	89.2	–	–	–	–	1.19	612				
mini	mini	oss	5	5	5k	md	base	base	–	75.5	–	75.5	–	–	–	–	1.43	611				
mini	mini	oss	5	5	5k	md	base	base	on	91.5	–	91.5	–	–	–	–	1.43	612				
mini	mini	qwen	5	5	10k	md	base	base	–	89.2	–	89.2	–	–	–	–	1.68	612				
mini	mini	qwen	5	5	5k	md	base	base	–	90.2	–	90.2	–	–	–	–	1.60	612				
mini	mini	qwn397	5	5	10k	md	base	base	–	87.3	–	87.3	–	–	–	–	1.27	612				
mini	mini	qwn397	5	5	5k	md	base	base	–	88.4	–	88.4	–	–	–	–	1.27	612				
Query-expansion prompt ablations (open-weight expanders)																						
oss-m	mini	nano	5	5	5k	md	base	base	–	82.5	–	82.5	–	–	–	–	1.01	612				
oss	mini	nano	5	5	5k	md	oss	base	–	86.6	–	86.6	–	–	–	–	1.13	612	v1 of the oss expansion prompt			

continued on the next page

Table 12: Exploratory ask-web variants (continued).

Models			Hyperparams				Prompts			Results										n	Note
qexp	urlr	synt	n-q	n-cr	trim	Crawl	qexp	synt	Tol.	Acc.	Std	T1	T2	T3	T4	T5	Cost/q (€)				
oss	mini	nano	5	5	5k	md	oss	base	–	85.6	–	85.6	–	–	–	–	1.16	612	v2 of the oss expansion prompt		
oss	mini	nano	5	5	5k	md	oss-tmpl	base	–	86.8	–	86.8	–	–	–	–	1.10	612			
oss	mini	nano	5	5	5k	md	oss-con	base	–	87.7	–	87.7	–	–	–	–	1.17	612			
oss	mini	nano	5	5	5k	md	oss-con2	base	–	84.2	–	84.2	–	–	–	–	1.15	612			
qwen	mini	nano	5	5	5k	md	qwen	base	–	89.1	–	89.1	–	–	–	–	1.67	612			
qwen	mini	nano	5	5	5k	md	qwen2	base	–	86.8	–	86.8	–	–	–	–	1.68	612			
Production configuration and its hyperparameter grid																					
mini	mini	nano	5	5	10k	md	base	base	–	–	–	–	–	–	–	–	–	–	no completed runs at the current config version; same config as the (5,5,10k) grid point below		
mini	mini	nano	3	2	2.5k	md	base	base	–	84.2	1.2	85.5	84.0	83.2	–	–	1.03	1836	trials 4–5 partially complete (541 and 594 of 612 cells)		
mini	mini	nano	3	2	5k	md	base	base	–	87.3	0.9	86.8	86.8	88.4	–	–	1.04	1836			
mini	mini	nano	3	2	10k	md	base	base	–	86.1	0.4	85.9	85.8	86.6	–	–	1.06	1836			
mini	mini	nano	3	5	2.5k	md	base	base	–	85.2	0.3	84.8	85.3	85.5	–	–	1.24	1836			
mini	mini	nano	3	5	5k	md	base	base	–	89.2	0.7	89.4	88.4	89.7	–	–	1.27	1836			
mini	mini	nano	3	5	10k	md	base	base	–	87.6	1.2	88.6	88.1	86.3	–	–	1.32	1836			
mini	mini	nano	3	8	2.5k	md	base	base	–	85.5	0.1	85.6	85.5	85.5	–	–	1.40	1836			
mini	mini	nano	3	8	5k	md	base	base	–	87.8	1.1	88.7	88.1	86.6	–	–	1.46	1836			
mini	mini	nano	3	8	10k	md	base	base	–	89.7	1.3	91.2	88.9	89.1	–	–	1.54	1836			
mini	mini	nano	5	2	2.5k	md	base	base	–	87.0	1.4	88.2	87.4	85.5	–	–	1.22	1836			
mini	mini	nano	5	2	5k	md	base	base	–	87.6	1.0	87.3	88.7	86.8	–	–	1.23	1836			
mini	mini	nano	5	2	10k	md	base	base	–	88.7	1.1	87.4	89.4	89.2	–	–	1.25	1836			
mini	mini	nano	5	5	2.5k	md	base	base	–	87.6	0.7	86.9	87.7	88.2	–	–	1.44	1836			
mini	mini	nano	5	5	5k	md	base	base	–	89.8	0.9	90.0	90.7	90.5	88.9	88.7	1.46	2971			
mini	mini	nano	5	5	10k	md	base	base	–	88.6	1.0	89.5	88.6	87.6	–	–	1.51	1836			
mini	mini	nano	5	8	2.5k	md	base	base	–	88.6	0.5	88.9	88.1	88.9	–	–	1.61	1836			
mini	mini	nano	5	8	5k	md	base	base	–	89.5	1.0	90.4	88.4	89.9	–	–	1.67	1836			
mini	mini	nano	5	8	10k	md	base	base	–	89.7	0.6	90.2	89.7	89.7	88.7	90.2	1.75	3048			
mini	mini	nano	8	2	2.5k	md	base	base	–	85.9	1.1	87.1	84.8	85.9	–	–	1.47	1836			
mini	mini	nano	8	2	5k	md	base	base	–	88.0	0.9	88.6	88.4	86.9	–	–	1.48	1836			
mini	mini	nano	8	2	10k	md	base	base	–	89.2	0.7	89.9	88.4	89.2	–	–	1.50	1836			
mini	mini	nano	8	5	2.5k	md	base	base	–	89.4	2.6	90.2	86.4	91.5	–	–	1.69	1836			
mini	mini	nano	8	5	5k	md	base	base	–	90.7	0.1	90.7	90.8	90.7	–	–	1.75	1836			
mini	mini	nano	8	5	10k	md	base	base	–	89.5	0.7	90.2	88.9	89.5	–	–	1.77	1836			
mini	mini	nano	8	8	2.5k	md	base	base	–	89.4	0.9	90.2	88.4	89.7	–	–	1.89	1836			
mini	mini	nano	8	8	5k	md	base	base	–	90.4	1.6	90.4	92.0	88.7	–	–	1.95	1836			
mini	mini	nano	8	8	10k	md	base	base	–	91.9	0.8	91.3	92.8	91.7	–	–	2.03	1836			
Commercial providers																					
–	–	–	–	–	–	–	–	–	–	86.9	–	86.9	–	–	–	–	0.50	612	commercial provider: Exa		
–	–	–	–	–	–	–	–	–	–	95.5	1.0	96.6	95.1	94.8	–	–	1.20	1836	/answer (see Table 2) commercial provider: Exa agent, minimal effort (see Table 2)		

continued on the next page

Table 12: Exploratory ask-web variants (continued).

Models			Hyperparams				Prompts			Results										n	Note
qexp	urlr	synt	n-q	n-cr	trim	Crawl	qexp	synt	Tol.	Acc.	Std	T1	T2	T3	T4	T5	Cost/q (€)				
-	-	-	-	-	-	-	-	-	-	89.9	-	89.9	-	-	-	-	1.00	612	commercial provider: Parallel Task API, base processor (see Table 2)		

D Tables

Table 13: Full prompt text for each prompt code used in `company-eval` (Table 14).

Prompt code	Full prompt
<code>pe-owned</code>	Determine whether this company (account) was owned or controlled by a private equity (PE) firm in April 2026.
<code>ma-activity</code>	Determine whether the target company completed, announced, or definitively agreed to a merger or acquisition from May 1, 2025 through April 30, 2026.
<code>recent-funding</code>	Determine whether the company raised funding, took on debt financing, or announced a recapitalization from January 31, 2026 through April 30, 2026.
<code>geographic-expansion</code>	Determine whether the company expanded into a new country or materially expanded its footprint in an existing country from May 1, 2025 through April 30, 2026.
<code>c-suite-change</code>	Determine whether the company publicly announced a new CEO, CFO, COO, President, or other chief officer role from January 31, 2026 through April 30, 2026.
<code>headcount-growth</code>	Determine whether the company’s employee headcount grew more than 20% year-over-year by the most recent public filing as of April 30, 2026.
<code>ontario-operations</code>	Determine whether the company currently has operations in Ontario, Canada, as of April 30, 2026.
<code>public-sector-sales</code>	Determine whether the company sells products or services to the public sector (state/local/federal government, education, or public agencies) as of April 30, 2026.
<code>ai-initiative</code>	Determine whether the company explicitly mentioned a new AI solution, AI feature, or AI product initiative in investor-facing or regulatory materials from November 1, 2024 through April 30, 2026.
<code>investment-announcement</code>	Check from February 1, 2026 through April 30, 2026 for a dated announcement of funding, private equity investment, acquisition, recapitalization, or strategic investment.
<code>canada-power-signals</code>	Determine whether public materials from January 31, 2026 through April 30, 2026 show any Canada-related hydroelectric, dam, or long-horizon power-capital program signals.
<code>divestiture-activity</code>	Determine whether the organization had any divestiture, merger, acquisition, or strategic investment activity as of April 30, 2026.
<code>new-product-launch</code>	Determine whether the company publicly launched, announced, or materially advanced a major new product, platform capability, or service offering from January 31, 2026 through April 30, 2026.
<code>plant-expansion</code>	Determine whether the company announced a plant expansion, plant modernization, or physical capacity project from January 31, 2026 through April 30, 2026.
<code>retail-expansion</code>	Determine whether the company announced a retailer expansion, new retail footprint, or broad store rollout from January 31, 2026 through April 30, 2026.
<code>company-acquired</code>	Determine whether the target company was acquired from May 1, 2025 through April 30, 2026.
<code>specific-pe-owner</code>	Determine whether the target company is owned by Vista Equity Partners, Evergreen Coast Capital, or a different private equity firm as of April 30, 2026.
<code>ipo-signals</code>	Determine whether the target company is preparing for or has signaled an IPO / S-1 / public listing path from May 1, 2025 through April 30, 2026.
<code>cao-eng-director-change</code>	Determine whether the target company had a leadership change involving a Chief Administrative Officer, an Engineering Director, or a Head of Engineering from January 31, 2026 through April 30, 2026.

Table 14: company-eval ground truth. T=True, F=False, I=Inconclusive.

company	pe-owned	ma-activity	recent-funding	geographic-expansion	c-suite-change	headcount-growth	ontario-operations	public-sector-sales	ai-initiative	investment-announcement	canada-power-signals	divestiture-activity	new-product-launch	plant-expansion	retail-expansion	company-acquired	specific-pe-owner	ipo-signals	cao-eng-director-change
OpenAI	F	T	T	T	T	I	I	T	T	T	I	T	T	I	I	F	F	T	T
Intuit	F	T	F	T	F	F	T	T	T	T	F	T	T	I	T	F	F	F	F
Snowflake	F	T	I	T	T	F	T	T	T	T	I	T	T	F	F	F	F	F	I
MongoDB	F	I	F	T	T	F	T	T	T	T	F	T	T	I	F	F	F	F	F
Visa	F	T	T	T	F	F	T	T	T	T	F	T	T	F	F	F	F	F	F
Meta	F	T	T	T	F	F	T	T	T	T	T	T	T	T	T	F	F	F	T
NVIDIA	F	T	F	T	T	F	T	T	T	T	I	T	T	T	F	F	F	F	T
Intel	F	T	T	T	T	F	T	T	T	T	F	T	T	T	F	F	F	I	F
TSMC	F	F	T	T	F	F	T	T	T	T	I	T	T	T	F	F	F	F	I
Altera	T	T	F	T	I	F	T	T	T	F	I	T	I	I	I	T	I	T	F
Thermo Fisher Scientific	F	T	T	T	F	F	T	T	T	T	I	T	T	T	F	F	F	F	F
Illumina	F	T	F	T	F	F	T	T	T	I	F	T	T	F	F	F	F	F	F
Danaher	F	T	T	T	I	F	T	T	T	T	F	T	T	T	F	F	F	F	F
Medline	T	F	I	T	F	F	T	T	T	T	F	T	T	T	F	F	T	T	I
Dell Technologies	I	T	I	T	T	F	T	T	T	I	I	T	T	T	F	F	F	F	F
Cisco	F	T	I	T	I	F	T	T	T	T	F	T	T	T	F	F	F	F	I
Hewlett Packard Enterprise	F	T	T	T	F	F	I	T	T	T	I	T	I	I	F	F	F	F	F
Juniper Networks	F	T	F	T	F	F	T	T	T	F	F	T	T	F	F	T	F	F	F
CVS Health	F	T	F	T	T	F	I	T	T	T	F	T	T	T	T	F	F	F	F
Walgreens Boots Alliance	T	T	F	T	F	F	F	T	I	F	F	T	T	T	T	T	T	F	F
Eli Lilly	F	T	I	T	I	F	T	T	T	T	F	T	I	T	I	F	F	F	F
Pfizer	F	T	F	T	F	F	T	T	T	T	F	T	T	I	F	F	F	F	F
Honeywell	F	T	T	T	I	F	T	T	T	T	I	T	T	T	F	F	F	T	F
Caterpillar	F	T	T	T	T	F	T	T	T	T	I	T	T	T	F	F	F	F	F
GE Vernova	F	T	T	T	F	F	T	T	T	T	T	T	T	T	F	F	F	F	F
Siemens	F	T	I	T	T	F	T	T	T	T	I	T	T	T	T	F	F	I	F
JPMorgan Chase	F	T	T	T	I	F	T	T	T	T	I	T	T	I	T	F	F	F	I
Goldman Sachs	F	T	T	T	F	F	T	T	T	T	I	T	T	F	F	F	F	F	I
Capital One	F	T	T	T	F	T	T	T	T	T	F	T	T	I	T	F	F	F	I
Discover Financial Services	F	T	F	T	F	F	T	T	I	F	F	T	T	F	I	T	F	F	F
ServiceNow	F	T	T	T	T	F	T	T	T	T	I	T	T	I	F	F	F	F	I
Workday	F	T	F	T	T	F	T	T	T	T	F	T	T	I	F	F	F	F	F
HubSpot	F	T	T	T	F	F	T	T	T	T	F	T	T	I	F	F	F	F	T
Figma	F	T	F	T	F	F	T	T	T	I	F	T	T	I	F	F	F	T	I
Smartsheet	T	F	F	T	T	F	F	T	T	F	F	T	T	I	F	F	T	F	I
Informatica	F	T	F	T	I	F	T	T	T	F	F	T	T	I	F	T	F	F	F

Table 15: url1-eval full benchmark: retrieved URLs per trial for each query. URLs are identical across providers, so one set is shown per trial. Ground truth shaded True, False, Inconclusive.

Q1. ai-initiative Cisco GT: True	Q2. ai-initiative Danaher GT: True	Q3. ai-initiative Intuit GT: True
T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5
Q4. c-suite-change Discover Financial Services GT: False	Q5. c-suite-change Figma GT: False	Q6. c-suite-change NVIDIA GT: True
T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4	T2 url1, url2, url3, url4, url5
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4	T8 url1, url2, url3, url4, url5
Q7. canada-power-signals Cisco GT: False	Q8. canada-power-signals Discover Financial Services GT: False	Q9. canada-power-signals Eli Lilly GT: False
T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5	T2 url1, url2, url3
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1	T5 url1	T5 url1, url2, url3, url4
T6 url1, url2, url3, url4, url5	T6 url1	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5
Q10. cao-eng-director-change Honeywell GT: False	Q11. cao-eng-director-change Pfizer GT: False	Q12. divestiture-activity Altera GT: True
T1 url1, url2, url3, url4, url5	T1 url1	T1 url1, url2, url3, url4, url5
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4	T8 url1	T8 url1, url2, url3, url4, url5

Q13. divestiture-activity Workday GT: True	Q14. geographic-expansion Dell Technologies GT: True	Q15. geographic-expansion Visa GT: True
T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5	T2 url1, url2, url3
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5
Q16. ipo-signals Caterpillar GT: False	Q17. ipo-signals NVIDIA GT: False	Q18. ma-activity Thermo Fisher Scientific GT: True
T1 url1, url2, url3, url4	T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4	T2 url1, url2, url3, url4, url5
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5
Q19. ontario-operations Capital One GT: True	Q20. ontario-operations Caterpillar GT: True	Q21. ontario-operations JPMorgan Chase GT: True
T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5
Q22. ontario-operations Visa GT: True	Q23. ontario-operations Workday GT: True	Q24. pe-owned Capital One GT: False
T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5

Q25. pe-owned Figma GT: False	Q26. pe-owned JPMorgan Chase GT: False	Q27. pe-owned NVIDIA GT: False
T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5
Q28. pe-owned Workday GT: False	Q29. plant-expansion Meta GT: True	Q30. plant-expansion TSMC GT: True
T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4 T3 url1, url2, url3, url4 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5
Q31. public-sector-sales Juniper Networks GT: True	Q32. public-sector-sales Meta GT: True	Q33. recent-funding GE Vernova GT: True
T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5
Q34. recent-funding Informatica GT: False	Q35. recent-funding ServiceNow GT: True	Q36. recent-funding Workday GT: False
T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4 T2 url1, url2, url3, url4 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5

Q37. retail-expansion Danaher GT: False	Q38. retail-expansion GE Vernova GT: False	Q39. specific-pe-owner Figma GT: False
T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3
Q40. specific-pe-owner OpenAI GT: False		
T1 url1, url2, url3, url4, url5 T2 url1, url2, url3, url4, url5 T3 url1, url2, url3, url4, url5 T4 url1, url2, url3, url4, url5 T5 url1, url2, url3, url4, url5 T6 url1, url2, url3, url4, url5 T7 url1, url2, url3, url4, url5 T8 url1, url2, url3, url4, url5		

Table 16: url1-eval full benchmark: mean synthesis accuracy (%) per (query, target) cell (Q1–Q40), over the 8 trials. Rows are synthesizer providers; **ovr** = mean over the 40 cells.

provider	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
nano	100	65	100	92	100	88	92	100	98	100	100	100	100	70	95	100	92	100	100	100
xhigh	100	65	100	88	100	88	100	100	100	100	100	100	100	62	92	100	100	100	100	100
oss	100	22	100	100	100	88	98	100	100	100	100	100	100	57	82	100	100	100	100	100
qwen	100	48	100	100	100	84	100	100	100	100	100	100	100	57	78	100	100	100	100	100
qwen(max)	100	40	100	98	100	88	100	100	100	100	100	100	100	30	75	100	100	100	100	94

provider	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37	Q38	Q39	Q40	ovr
nano	100	100	98	100	100	100	98	100	100	100	100	88	88	78	98	85	100	100	85	62	94
xhigh	100	100	100	100	100	100	100	100	100	100	100	100	88	100	100	100	100	100	100	100	97
oss	100	100	82	100	100	100	100	100	100	100	100	95	88	82	100	57	100	100	100	88	94
qwen	100	100	100	100	100	100	100	100	100	78	100	98	88	100	100	100	100	100	100	94	96
qwen(max)	100	100	95	100	100	100	100	100	100	90	100	92	88	100	100	100	100	100	100	100	95

Legend: Q1=ai-initiative/Cisco; Q2=ai-initiative/Danaher; Q3=ai-initiative/Intuit; Q4=c-suite-change/Discover Financial Services; Q5=c-suite-change/Figma; Q6=c-suite-change/NVIDIA;
Q7=canada-power-signals/Cisco; Q8=canada-power-signals/Discover Financial Services; Q9=canada-power-signals/Eli Lilly; Q10=cao-eng-director-change/Honeywell;
Q11=cao-eng-director-change/Pfizer; Q12=divestiture-activity/Altera; Q13=divestiture-activity/Workday; Q14=geographic-expansion/Dell Technologies; Q15=geographic-expansion/Visa;
Q16=ipo-signals/Caterpillar; Q17=ipo-signals/NVIDIA; Q18=ma-activity/Thermo Fisher Scientific; Q19=ontario-operations/Capital One; Q20=ontario-operations/Caterpillar;
Q21=ontario-operations/JPMorgan Chase; Q22=ontario-operations/Visa; Q23=ontario-operations/Workday; Q24=pe-owned/Capital One; Q25=pe-owned/Figma; Q26=pe-owned/JPMorgan Chase;
Q27=pe-owned/NVIDIA; Q28=pe-owned/Workday; Q29=plant-expansion/Meta; Q30=plant-expansion/TSMC; Q31=public-sector-sales/Juniper Networks; Q32=public-sector-sales/Meta;
Q33=recent-funding/GE Vernova; Q34=recent-funding/Informatica; Q35=recent-funding/ServiceNow; Q36=recent-funding/Workday; Q37=retail-expansion/Danaher; Q38=retail-expansion/GE Vernova; Q39=specific-pe-owner/Figma; Q40=specific-pe-owner/OpenAI.

Table 17: url-eval full benchmark: per-trial synthesis accuracy (%). Each block is one problem (Q1–Q40, legend in Table 16); rows are its 8 trials. Columns, in order: trial, nano, xhigh, oss, qwen, qwen(max), and the mean across providers. E = all samples errored.

Q1. ai-initiative Cisco							Q2. ai-initiative Danaher							Q3. ai-initiative Intuit						
1	100	100	100	100	100	100	1	20	0	0	0	0	4	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	40	100	100	88	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	0	60	40	60	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	20	100	0	0	0	24	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	60	80	0	20	0	32	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	100	40	40	100	80	72	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	20	0	0	0	0	4	8	100	100	100	100	100	100
avg	100	100	100	100	100	100	avg	65	65	22	48	40		avg	100	100	100	100	100	100
Q4. c-suite-change Discover Financial Services							Q5. c-suite-change Figma							Q6. c-suite-change NVIDIA						
1	100	100	100	100	100	100	1	100	100	100	100	100	100	1	100	100	100	75	100	95
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	100	100	100	100	3	100	100	100	100	100	100
4	80	60	100	100	80	84	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	60	40	100	100	100	80	5	100	100	100	100	100	100	5	0	0	0	0	0	0
6	100	100	100	100	100	100	6	100	100	100	100	100	100	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	100	100	100	100	100	100	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	100	100	100	100	100	100	8	100	100	100	100	100	100
avg	92	88	100	100	98		avg	100	100	100	100	100	100	avg	88	88	88	84	88	
Q7. canada-power-signals Cisco							Q8. canada-power-signals Discover Financial Services							Q9. canada-power-signals Eli Lilly						
1	40	100	80	100	100	84	1	100	100	100	100	100	100	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	100	100	100	100	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	100	100	100	100	100	100	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	100	100	100	100	100	100	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	100	100	100	100	100	100	8	80	100	100	100	100	96
avg	92	100	98	100	100		avg	100	100	100	100	100	100	avg	98	100	100	100	100	
Q10. cao-eng-director-change Honeywell							Q11. cao-eng-director-change Pfizer							Q12. divestiture-activity Altera						
1	100	100	100	100	100	100	1	100	100	100	100	100	100	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	100	100	100	100	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	100	100	100	100	100	100	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	100	100	100	100	100	100	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	100	100	100	100	100	100	8	100	100	100	100	100	100
avg	100	100	100	100	100	100	avg	100	100	100	100	100	100	avg	100	100	100	100	100	100
Q13. divestiture-activity Workday							Q14. geographic-expansion Dell Technologies							Q15. geographic-expansion Visa						
1	100	100	100	100	100	100	1	0	0	0	0	0	0	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	80	100	20	80	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	60	60	40	72	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	60	92	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	100	100	100	100	20	84	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	60	0	20	0	0	16	7	60	40	0	0	0	20
8	100	100	100	100	100	100	8	0	0	0	0	0	0	8	100	100	60	20	0	56
avg	100	100	100	100	100	100	avg	70	62	57	57	30		avg	95	92	82	78	75	

Q16. ipo-signals Caterpillar							Q17. ipo-signals NVIDIA							Q18. ma-activity Thermo Fisher Scientific						
1	100	100	100	100	100	100	1	100	100	100	100	100	100	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	100	100	100	100	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	80	100	100	100	100	96	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	100	100	100	100	100	100	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	60	100	100	100	100	92	8	100	100	100	100	100	100
avg	100	100	100	100	100	100	avg	92	100	100	100	100		avg	100	100	100	100	100	100
Q19. ontario-operations Capital One							Q20. ontario-operations Caterpillar							Q21. ontario-operations JPMorgan Chase						
1	100	100	100	100	100	100	1	100	100	100	100	100	100	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	100	100	80	96	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	100	100	100	100	E	100	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	100	100	100	100	80	96	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	100	100	100	100	100	100	8	100	100	100	100	100	100
avg	100	100	100	100	100	100	avg	100	100	100	100	94		avg	100	100	100	100	100	100
Q22. ontario-operations Visa							Q23. ontario-operations Workday							Q24. pe-owned Capital One						
1	100	100	100	100	100	100	1	100	100	60	100	60	84	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	100	100	100	100	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	100	100	100	100	100	100	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	80	100	0	100	100	76	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	100	100	100	100	100	100	8	100	100	100	100	100	100
avg	100	100	100	100	100	100	avg	98	100	82	100	95		avg	100	100	100	100	100	100
Q25. pe-owned Figma							Q26. pe-owned JPMorgan Chase							Q27. pe-owned NVIDIA						
1	100	100	100	100	100	100	1	100	100	100	100	100	100	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	100	100	100	100	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	100	100	100	100	100	100	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	100	100	100	100	100	100	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	100	100	100	100	100	100	8	80	100	100	100	100	96
avg	100	100	100	100	100	100	avg	100	100	100	100	100		avg	98	100	100	100	100	100
Q28. pe-owned Workday							Q29. plant-expansion Meta							Q30. plant-expansion TSMC						
1	100	100	100	100	100	100	1	100	100	100	100	100	100	1	100	100	100	0	60	72
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	100	100	100	100	100	100	3	100	100	100	100	100	100	3	100	100	100	100	100	100
4	100	100	100	100	100	100	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	100	100	100	100	100	100	6	100	100	100	100	100	100
7	100	100	100	100	100	100	7	100	100	100	100	100	100	7	100	100	100	100	100	100
8	100	100	100	100	100	100	8	100	100	100	100	100	100	8	100	100	100	20	60	76
avg	100	100	100	100	100	100	avg	100	100	100	100	100		avg	100	100	100	78	90	

Q31. public-sector-sales Juniper Networks							
1	100	100	100	100	100	100	100
2	100	100	100	100	100	100	100
3	100	100	100	100	100	100	100
4	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100
avg	100	100	100	100	100	100	100
Q32. public-sector-sales Meta							
1	100	100	100	100	100	100	100
2	100	100	100	100	100	100	100
3	100	100	100	100	100	100	100
4	0	100	60	80	40	56	100
5	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100
avg	88	100	95	98	92	100	100
Q33. recent-funding GE Vernova							
1	100	100	100	100	100	100	100
2	100	100	100	100	100	100	100
3	100	100	100	100	100	100	100
4	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	0	0	0	0	0	0	0
8	100	100	100	100	100	100	100
avg	88	88	88	88	88	88	88
Q34. recent-funding Informatica							
1	40	100	60	100	100	100	80
2	40	100	60	100	100	100	80
3	40	100	80	100	100	100	84
4	100	100	100	100	100	100	100
5	100	100	80	100	100	100	96
6	100	100	80	100	100	100	96
7	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100
avg	78	100	82	100	100	100	100
Q35. recent-funding ServiceNow							
1	80	100	100	100	100	100	96
2	100	100	100	100	100	100	100
3	100	100	100	100	100	100	100
4	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100
avg	98	100	100	100	100	100	100
Q36. recent-funding Workday							
1	100	100	100	100	100	100	100
2	100	100	60	100	100	100	92
3	100	100	0	100	100	100	80
4	80	100	20	100	100	100	80
5	80	100	80	100	100	100	92
6	100	100	100	100	100	100	100
7	20	100	0	100	100	100	64
8	100	100	100	100	100	100	100
avg	85	100	57	100	100	100	100
Q37. retail-expansion Danaher							
1	100	100	100	100	100	100	100
2	100	100	100	100	100	100	100
3	100	100	100	100	100	100	100
4	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100
avg	100	100	100	100	100	100	100
Q38. retail-expansion GE Vernova							
1	100	100	100	100	100	100	100
2	100	100	100	100	100	100	100
3	100	100	100	100	100	100	100
4	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100
avg	100	100	100	100	100	100	100
Q39. specific-pe-owner Figma							
1	20	100	100	100	100	100	84
2	100	100	100	100	100	100	100
3	60	100	100	100	100	100	92
4	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100
avg	85	100	100	100	100	100	100
Q40. specific-pe-owner OpenAI							
1	100	100	100	75	100	100	95
2	60	100	100	100	100	100	92
3	20	100	40	100	100	100	72
4	100	100	100	100	100	100	100
5	20	100	100	100	100	100	84
6	20	100	60	80	100	100	72
7	80	100	100	100	100	100	96
8	100	100	100	100	100	100	100
avg	62	100	88	94	100	100	100

Table 18: url-eval mini benchmark: retrieved URLs per trial for each query. URLs are identical across providers, so one set is shown per trial. Ground truth shaded True, False, Inconclusive.

Q1. c-suite-change Smartsheet GT: True	Q2. c-suite-change GE Vernova GT: False	Q3. canada-power-signals Medline GT: False
T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4, url5
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5
T9 url1, url2, url3, url4, url5	T9 url1, url2, url3, url4, url5	T9 url1, url2, url3, url4, url5
T10 url1, url2, url3, url4, url5	T10 url1, url2, url3, url4, url5	T10 url1, url2, url3, url4, url5
T11 url1, url2, url3, url4, url5	T11 url1, url2, url3, url4, url5	T11 url1, url2, url3, url4, url5
T12 url1, url2, url3, url4, url5	T12 url1, url2, url3, url4, url5	T12 url1, url2, url3, url4, url5
T13 url1, url2, url3, url4, url5	T13 url1, url2, url3, url4, url5	T13 url1, url2, url3, url4, url5
T14 url1, url2, url3, url4, url5	T14 url1, url2, url3, url4, url5	T14 url1, url2, url3, url4, url5
T15 url1, url2, url3, url4	T15 url1, url2, url3, url4, url5	T15 url1, url2, url3, url4, url5
T16 url1, url2, url3, url4, url5	T16 url1, url2, url3, url4, url5	T16 url1, url2, url3, url4, url5
Q4. geographic-expansion Hewlett Packard Enterprise GT: True	Q5. investment-announcement Discover Financial Services GT: False	Q6. ma-activity Intel GT: True
T1 url1, url2, url3, url4, url5	T1 url1, url2, url3, url4	T1 url1, url2, url3, url4, url5
T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5	T2 url1, url2, url3, url4, url5
T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5	T3 url1, url2, url3, url4, url5
T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5	T4 url1, url2, url3, url4, url5
T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5	T5 url1, url2, url3, url4, url5
T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5	T6 url1, url2, url3, url4, url5
T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5	T7 url1, url2, url3, url4, url5
T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5	T8 url1, url2, url3, url4, url5
T9 url1, url2, url3, url4, url5	T9 url1, url2, url3, url4, url5	T9 url1, url2, url3, url4
T10 url1, url2, url3, url4, url5	T10 url1, url2, url3, url4, url5	T10 url1, url2, url3, url4, url5
T11 url1, url2, url3, url4, url5	T11 url1, url2, url3, url4, url5	T11 url1, url2, url3, url4, url5
T12 url1, url2, url3, url4, url5	T12 url1, url2, url3, url4, url5	T12 url1, url2, url3, url4, url5
T13 url1, url2, url3, url4, url5	T13 url1, url2, url3, url4, url5	T13 url1, url2, url3, url4, url5
T14 url1, url2, url3, url4, url5	T14 url1, url2, url3, url4, url5	T14 url1, url2, url3, url4, url5
T15 url1, url2, url3, url4, url5	T15 url1, url2, url3, url4, url5	T15 url1, url2, url3, url4, url5
T16 url1, url2, url3, url4, url5	T16 url1, url2, url3, url4	T16 url1, url2, url3, url4, url5

Table 19: url-eval mini benchmark: per-trial synthesis accuracy (%). Each block is one problem (Q1–Q6, legend in Table 11); rows are its 16 trials. Columns, in order: trial, nano, xhigh, oss, qwen, qwen(max), and the mean across providers. E = all samples errored.

Q1. c-suite-change Smartsheet							Q2. c-suite-change GE Vernova							Q3. canada-power-signals Medline						
1	100	100	100	100	100	100	1	0	80	40	100	100	64	1	100	100	100	100	100	100
2	100	100	100	100	100	100	2	100	100	100	100	100	100	2	0	80	60	100	80	64
3	100	100	100	100	100	100	3	100	100	100	100	100	100	3	100	100	100	100	100	100
4	0	80	0	0	0	16	4	0	100	40	100	100	68	4	0	100	60	100	100	72
5	100	100	100	100	100	100	5	100	100	100	100	100	100	5	100	100	100	100	100	100
6	100	100	100	100	100	100	6	100	100	100	100	100	100	6	0	100	60	80	100	68
7	100	100	100	100	100	100	7	100	100	100	100	100	100	7	0	80	60	60	100	60
8	0	0	20	0	40	12	8	100	100	100	100	100	100	8	0	100	0	80	80	52
9	0	100	0	0	0	20	9	100	100	100	100	100	100	9	0	100	0	60	100	52
10	0	100	0	0	0	20	10	80	100	100	100	100	96	10	0	100	40	60	100	60
11	0	80	0	0	0	16	11	100	100	100	100	100	100	11	100	100	100	100	E	100
12	0	100	0	0	0	20	12	100	100	100	100	100	100	12	100	100	100	100	100	100
13	0	100	0	0	0	20	13	100	100	100	100	100	100	13	100	100	100	100	100	100
14	100	100	100	100	100	100	14	100	100	100	100	100	100	14	100	100	100	100	100	100
15	100	100	100	100	100	100	15	100	100	100	100	100	100	15	0	100	20	0	100	44
16	20	100	0	20	0	28	16	100	100	100	100	100	100	16	0	100	20	60	100	56
avg	51	91	51	51	52		avg	86	99	92	100	100		avg	44	98	64	81	97	

Q4. geographic-expansion Hewlett Packard Enterprise							Q5. investment-announcement Discover Financial Services							Q6. ma-activity Intel						
1	20	100	0	60	80	52	1	0	0	0	0	0	0	1	40	100	100	100	80	84
2	60	100	20	20	0	40	2	100	100	100	100	100	100	2	100	100	100	100	100	100
3	20	100	60	80	80	68	3	0	100	40	80	100	64	3	80	100	100	100	100	96
4	100	100	100	80	20	80	4	100	100	100	100	100	100	4	100	100	100	100	100	100
5	20	80	0	0	0	20	5	0	0	0	0	0	0	5	60	100	80	100	100	88
6	40	80	0	0	0	24	6	0	0	0	0	0	0	6	20	100	40	80	80	64
7	0	40	0	0	0	8	7	0	0	0	0	20	4	7	40	100	100	100	100	88
8	0	100	0	20	40	32	8	100	100	100	100	100	100	8	100	100	100	100	100	100
9	100	0	80	0	0	36	9	0	60	0	20	60	28	9	100	100	100	100	100	100
10	0	80	0	0	0	16	10	80	100	80	100	100	92	10	0	100	0	100	40	48
11	40	100	0	0	0	28	11	0	0	0	0	0	0	11	40	100	60	80	100	76
12	100	100	100	100	0	80	12	0	0	0	0	0	0	12	20	100	100	60	0	56
13	40	100	20	20	0	36	13	100	100	100	100	100	100	13	60	100	100	100	100	92
14	0	60	0	0	0	12	14	100	100	100	100	100	100	14	80	100	100	100	100	96
15	0	100	0	20	20	28	15	0	0	0	0	60	12	15	100	100	100	100	100	100
16	0	80	0	0	0	16	16	0	0	0	0	0	0	16	80	100	100	80	100	92
avg	34	82	24	25	15		avg	36	48	39	44	52		avg	64	100	86	94	88	

E Trace Analysis

This section shows two traces for the same query under variants of the **ask-web** architecture. Both variants use 5 subqueries, crawl at most 5 URLs, trim each crawl result to 10,000 tokens, and use GPT 5.0 medium (medium reasoning effort) for query expansion and URL ranking. They differ in the synthesizer. The first uses GPT 5.0 medium with minimal reasoning effort and answers incorrectly; the second uses GPT 5.0 nano with high reasoning effort and answers correctly.

The listings below show the input and output of each LLM stage — query expansion, URL ranking, and synthesis — for the following (prompt, company) pair.

prompt: “Determine whether the company publicly announced a new CEO, CFO, COO, President, or other chief officer role from January 31, 2026 through April 30, 2026.” *company*: HubSpot.

None of the links received by the first synthesizer (Listing 5) provides evidence for a new chief-officer appointment within the January 31–April 30, 2026 window, so the ground-truth answer is FALSE. Nevertheless, this **ask-web** variant answers TRUE.

Listing 1: Query expansion — input.

```
Research question: Determine whether the company publicly announced a new CEO, CFO,
COO, President, or other chief officer role from January 31, 2026 through
April 30, 2026.

Company: HubSpot
```

Listing 2: Query expansion — output (five Google search queries).

```
1. "HubSpot" AND ("press release" OR "press-release") AND (CEO OR CFO OR COO OR
  President OR "chief officer") AND (January OR February OR March OR April 2026)
2. "HubSpot" AND ("Form 8-K" OR "8-K" OR "SEC filing") AND (CEO OR CFO OR COO OR
  President OR "chief officer") AND (January OR February OR March OR April 2026)
3. HubSpot AND (announce OR announces OR announced OR appoint OR appoints OR
  appointed OR named) AND (CEO OR CFO OR COO OR President OR "chief officer")
  AND (January OR February OR March OR April 2026)
4. "HubSpot" AND ("leadership change" OR "executive change" OR "management change"
  OR "leadership transition") AND (CEO OR CFO OR COO OR President OR
  "chief officer") AND (January OR February OR March OR April 2026)
5. "HubSpot" AND (LinkedIn OR "social media" OR "company blog" OR "press briefing")
  AND (CEO OR CFO OR COO OR President OR "chief officer") AND (January OR
  February OR March OR April 2026)
```

Listing 3: URL ranking — input (original question, company, and the web search results grouped by search query; each result has Source, Title, and Content snippet).

```
Original question: Determine whether the company publicly announced a new CEO, CFO, COO, President, or
other chief officer role from January 31, 2026 through April 30, 2026.

Company: HubSpot

Web search results:
Search query: "HubSpot" AND ("press release" OR "press-release") AND (CEO OR CFO OR COO OR President OR
"chief officer") AND (January OR February OR March OR April 2026)

- Source: https://ir.hubspot.com/node/15261/html
; Title: SEC Filing | HubSpot
; Content: On April 1, 2026 the Company issued a press release announcing Mr. ... CEO of HubSpot. "His
vast experience as a CFO at scaling companies, combined ...Read more
- Source: https://www.hubspot.com/media-kit
; Title: Press Center | Spring 2026 Spotlight
; Content: ... 2026 Spotlight press release Read the HubSpot AEO press release. Featured ... President,
Digital Marketing Manager. Mercantile Bank. Read more about ...Read more
- Source: https://finance.yahoo.com/markets/stocks/articles/hubspot-announces-mike-berry-joins
-201000786.html
; Title: HubSpot Announces Mike Berry Joins Board of Directors
; Content: Apr 1, 2026 -- Contact the press release distributor directly with any inquiries. Business
Wire. HubSpot Announces Mike Berry Joins Board of Directors.Read more
- Source: https://ir.hubspot.com/static-files/a4c0bbe9-83ff-47b3-89f6-259fecf58a2e
; Title: HubSpot, Inc. (HUBS)
```

; Content: May 7, 2026 -- Chief Executive Officer & Director, HubSpot, Inc. Thank you, Chuck, and welcome to everyone joining us today. I'll start with our Q1 2026 ...Read more
- Source: <https://www.venasolutions.com/newsroom/vena-taps-saas-growth-veteran-and-former-hubspot-chief-sales-officer-as-new-ceo>
; Title: Vena Taps Former HubSpot Chief Sales Officer as CEO
; Content: Vena Taps SaaS Growth Veteran and Former HubSpot Chief Sales Officer as New CEO ... Press Release. New Vena Research Finds FP&A Teams Are Hitting a Maturity ...Read more
- Source: <https://www.customerexperiencedive.com/news/hubspot-chief-customer-officer-departure-customer-success/713765/>
; Title: HubSpot nixes chief customer officer role following ...
; Content: Apr 19, 2024 -- HubSpot nixes chief customer officer role following executive departure. HubSpot will not fill the space left by the departure of Rob Giglio ...Read more
- Source: <https://en.wikipedia.org/wiki/HubSpot>
; Title: HubSpot
; Content: In February 2026, HubSpot acquired the creator-led entrepreneurship publication Starter Story "The Deadhead who became a \$38 billion CEO: What HubSpot ...Read more
- Source: <https://www.linkedin.com/in/brianhalligan>
; Title: Brian Halligan - Long Strange Trip: CEO to ...
; Content: These days I spend my time helping startup founders turn into scale up CEOs through... Articles by Brian. How HubSpot "Fixed" Retention. Nov 4, ...Read more
- Source: <https://www.syntaxdata.com/press>
; Title: Media & Press | Syntax
; Content: Press Release. Partnerships. November 4, 2025. Bruce Traan Joins Syntax Data as President and COO of Syntax's Index Division. Syntax Data, a financial data and ...Read more

Search query: "HubSpot" AND ("Form 8-K" OR "8-K" OR "SEC filing") AND (CEO OR CFO OR COO OR President OR "chief officer") AND (January OR February OR March OR April 2026)

- Source: <https://ir.hubspot.com/static-files/9a43d64a-6547-4a58-a687-c05ca460f087>
; Title: Form 8-K for Hubspot INC filed 02/11/2026
; Content: Feb 11, 2026 -- Check the appropriate box below if the Form 8-K filing is ... said Yamini Rangan, Chief Executive Officer at HubSpot. "AI adoption ...Read more
- Source: <https://www.stocktitan.net/sec-filings/HUBS/page-5.html>
; Title: HUBS SEC Filings - Hubspot Inc 10-K, 10-Q, 8-K Forms
; Content: ... 8-K material events, and insider trading forms. ... HubSpot CEO Yamini Rangan reported equity ... The most recent SEC filing for Hubspot (HUBS) was filed on ...
- Source: <https://www.sec.gov/Archives/edgar/data/1889956/000119312525305275/ck0001889956-20251125.htm>
; Title: FORM 8-K
; Content: Nov 25, 2025 -- ... Form 8-K filing is intended to simultaneously satisfy the filing obligation of the registrant under any of the following provisions ...Read more
- Source: <https://ir.hubspot.com/node/15186/html>
; Title: SEC Filing
; Content: C/O HUBSPOT, INC. 2 CANAL PARK. (Street). CAMBRIDGE, MA, 02141 ...Read more
- Source: <https://ir.hubspot.com/node/12641/html>
; Title: SEC Filing | HubSpot
; Content: ... Form 8-K and will provide the final results in an amendment to the Form 8-K ... HubSpot's sustainability programs in collaboration with our executive leadership ...Read more
- Source: <https://finance.yahoo.com/quote/HUBS/profile/>
; Title: HubSpot, Inc. (HUBS) Company Profile & Facts
; Content: CEO, President and Director, 1.45M, --, 1974. Mr. Dharmesh Shah, Co-Founder, CTO ... 8-K: Corporate Changes & Voting Matters. See Full Filing. May 13, 2026 at ...Read more
- Source: <https://community.hubspot.com/t5/Releases-and-Updates/Top-Product-Updates-for-January-2026/ba-p/1247546>
; Title: Top Product Updates for January 2026
; Content: Feb 2, 2026 -- Top Product Updates of January 2026. January brought a wave of product updates to HubSpot, and while some grabbed headlines, others quietly ...Read more
- Source: <https://www.marketbeat.com/stocks/NYSE/HUBS/sec-filings/>
; Title: HubSpot (HUBS) 10K Form and Latest SEC Filings 2026
; Content: MarketBeat offers HubSpot's complete SEC filing history through 2026 ... Form 8-K Current report pursuant to Section 13 or 15(d). 03/06/2026 11:58 AM.Read more
- Source: <https://stocklight.com/stocks/us/nyse-hubs/hubspot/annual-reports/nyse-hubs-2026-10K-26621141.pdf>
; Title: HubSpot Annual Report 2026 Form 10-K (NYSE:HUBS)
; Content: HubSpot Annual Report 2026. Form 10-K (NYSE:HUBS) ... Form 8-K, and amendments to these reports filed or ... Executive Severance Plan, effective April 10, 2025.

Search query: HubSpot AND (announce OR announces OR announced OR appoint OR appoints OR appointed OR named) AND (CEO OR CFO OR COO OR President OR "chief officer") AND (January OR February OR March OR April 2026)

- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312526138021/hubs-ex99_1.htm
; Title: EX-99.1
; Content: Mike Berry, Chief Financial Officer of MongoDB, has been appointed to the company's Board of

Directors, effective April 1, 2026. Yamini Rangan, CEO of HubSpot.

- Source: <https://community.hubspot.com/t5/Advocates-Blog/Meet-the-Champion-of-the-Month-March-2026/ba-p/1254364>

; Title: Meet the Champion of the Month: March 2026

; Content: Mar 2, 2026 -- Hello Community! I am thrilled to announce that our Champion of the Month for March 2026 is: Tim Beukman What makes Tim our Community ...

- Source: <https://www.hubspot.com/company-news/hubspot-appoints-kate-bueker-as-chief-financial-officer>

; Title: HubSpot Appoints Kate Bueker as Chief Financial Officer

; Content: HubSpot today announced that Kate Bueker will become Chief Financial Officer of HubSpot, effective June 18, 2018. Bueker brings almost 25 years of financial ...Read more

- Source: <https://simplywall.st/stocks/de/software/fra-096/hubspot-shares/management>

; Title: HubSpot, Inc. (096) Leadership & Management Team ...

; Content: HubSpot, Inc. announced that Mike Berry, Chief Financial Officer of MongoDB, has been appointed to the company's Board of Directors, effective April 1, 2026.Read more

- Source: <https://www.instagram.com/reel/DZdH00jk9Vy/>

; Title: HubSpot Life | we turned twenty. thank you to every ...

; Content: Hello, world. Turns one! Benjamin Faes RWS CEO Matthew Hardy RWS SVP Product The Global Dialogue 2026 Bengaluru 25th March 2026 Riding the AI ...

- Source: <https://unbound.hubspot.com/>

; Title: UNBOUND 2026 | HubSpot's Annual Conference

; Content: HubSpot's flagship event is now UNBOUND. Join go-to-market leaders in Boston, Sept 16-18, 2026, for AI-driven insights and unstoppable growth.

Search query: "HubSpot" AND ("leadership change" OR "executive change" OR "management change" OR "leadership transition") AND (CEO OR CFO OR COO OR President OR "chief officer") AND (January OR February OR March OR April 2026)

- Source: <https://community.hubspot.com/t5/Releases-and-Updates/Top-Product-Updates-for-February-2026/ba-p/1255291>

; Title: Top Product Updates for February 2026

; Content: HubSpot's AI connectors got a major upgrade in February. Both the Claude and ChatGPT connectors can now create and update CRM records, log activities, and ...Read more

- Source: <https://community.hubspot.com/t5/Releases-and-Updates/The-January-2026-Industry-Edit-Essential-HubSpot-Updates/ba-p/1245652>

; Title: The January 2026 Industry Edit: Essential HubSpot Updates

; Content: Jan 28, 2026 -- Over the last month, HubSpot has released a set of updates that make it easier to support customers, run smarter outreach, and bring AI into ...

- Source: <https://www.youtube.com/watch?v=Kbo7Q14OnzU>

; Title: 8 HubSpot Updates Admins Shouldn't Miss (Jan 2026)

; Content: Did HubSpot ship 47 updates in January, or did it just feel like it? ... 8 HubSpot Updates Admins Shouldn't Miss (Jan 2026). 409 views - 3 ...

- Source: <https://www.martech.org/15-hubspot-updates-from-march-2026-managers-and-admins-need-to-know/>

; Title: 15 HubSpot updates from March 2026 managers and ...

; Content: Apr 24, 2026 -- The March 2026 HubSpot updates are all about flexibility, especially in how your automation and data actually work.Read more

- Source: <https://community.hubspot.com/t5/Releases-and-Updates/The-March-2026-Industry-Edit-Essential-HubSpot-Updates/ba-p/1261823>

; Title: The March 2026 Industry Edit: Essential HubSpot Updates

; Content: Mar 25, 2026 -- Automatically generate contracts from quotes and manage mid-term changes like upgrades or pricing adjustments directly in HubSpot. Manage ...Read more

- Source: <https://community.hubspot.com/t5/Releases-and-Updates/Top-Product-Updates-for-March-2026/ba-p/1263963>

; Title: Top Product Updates for March 2026

; Content: Apr 1, 2026 -- Top HubSpot Product Updates for March 2026 - 1. Workflow Improvements - 2. Customer Agent Improvements - 3. Breeze AI Reporting - 4. CRM and Admin ...Read more

- Source: <https://www.smartbugmedia.com/inbound-marketing-agency/news-releases/topic/news-release>

; Title: SmartBug News and Media Coverage | News Release

; Content: SmartBug Media(R) announced a strategic leadership transition, promoting Adam Bleibtreu from his current role as president of the company to chief executive ...Read more

- Source: <https://www.ainvest.com/news/hubspot-2026-q1-earnings-leadership-change-won-slow-growth-2605/>

; Title: Hubspot's 2026 Q1 Earnings: Why the Leadership Change Won't ...

; Content: Hubspot's 2026 Q1 Earnings: Why the Leadership Change Won't Slow Growth. ... In March 2026, CEO Brian Halligan announced plans to step down by the end of 2026, ...

- Source: <https://www.youtube.com/watch?v=XIROrLkp63I>

; Title: What's New In HubSpot - January 2026

; Content: What's New In HubSpot - January 2026. 36 views - 4 months ago ...more. BBD Boom. 106. Subscribe. 0. Share. Save. Report. Comments.

- Source: <https://www.bbdboom.com/blog/hubspot-product-update-april-2026>

; Title: HubSpot Product Update: April 2026

; Content: Apr 29, 2026 -- Welcome back to our HubSpot Product Spotlight! In this blog, we'll explore some of our favourite HubSpot updates from the last month.

Search query: "HubSpot" AND (LinkedIn OR "social media" OR "company blog" OR "press briefing") AND (CEO

OR CFO OR COO OR President OR "chief officer") AND (January OR February OR March OR April 2026)

- Source: https://www.linkedin.com/posts/artur-kiulian_i-started-2026-with-an-obsession-to-understand-activity-7417408511904620544-Rblq
- ; Title: HubSpot's Strategy Secrets: 5 Key Takeaways from Daby O.
- ; Content: Around 53% of the so-called "experts" on LinkedIn haven't touched a HubSpot property in more than 9 months. Around 99% of people reading this ...Read more
- Source: https://www.linkedin.com/posts/lexkeaton_lex-for-2026-presidents-club-i-activity-7416609074789662720--iWd
- ; Title: Lex for 2026 President's Club I joined HubSpot in mid ...
- ; Content: Lex for 2026 President's Club I joined HubSpot in mid-2025, which made qualifying for 2026 President's Club a stretch, but there is still ...Read more
- Source: <https://vantagepoint.io/blog/sf/blog/hs/hubspot-sales-workspace-update-april-2026-preparation-guide>
- ; Title: HubSpot Sales Workspace Update April 2026
- ; Content: HubSpot is forcing all Sales Hub Professional and Enterprise users onto a redesigned Sales Workspace on April 27, 2026.
- Source: https://www.linkedin.com/posts/hubspot-academy_2026-state-of-marketing-report-insights-activity-7438320357628964864-8p6z
- ; Title: HubSpot's 2026 State of Marketing Report: Brand POV ...
- ; Content: Crystal King coming to you with the freshest stats & insights from HubSpot's 2026 State of Marketing Report Learn the importance of brand ...
- Source: <https://www.clay.com/dossier/hubspot-ceo>
- ; Title: Who is the CEO of HubSpot in 2026? Yamini Rangan's Bio
- ; Content: HubSpot - Chief Executive Officer: 2021 - Present ; HubSpot - Chief Customer Officer: 2020 - 2021 ; Dropbox - Chief Customer Officer: 2018 - 2020 ; Dropbox - Vice ...Read more
- Source: <https://cxfoundation.com/news/hubspot-spring-spotlight-2026>
- ; Title: HubSpot Spring Spotlight 2026: 5 Big Talking Points
- ; Content: Apr 14, 2026 -- HubSpot has announced its Spring Spotlight 2026 release, focused on the idea of building the "agentic customer platform".Read more
- Source: <https://southbound.revpartners.io/>
- ; Title: Southbound '26 | The GTM Conference for Revenue Leaders
- ; Content: Chief Executive Officer HubSpot. Rachel Elsts Downey. CEO & Founder Share Your Genius. Jon Acuff. NYT Bestselling Author, CEO

Listing 4: URL ranking — output (selected URLs to crawl and feedback on the generated queries).

urls:

1. <https://ir.hubspot.com/node/15261/html>
2. https://www.sec.gov/Archives/edgar/data/1404655/000119312526138021/hubs-ex99_1.htm
3. <https://finance.yahoo.com/markets/stocks/articles/hubspot-announces-mike-berry-joins-201000786.html>
4. <https://ir.hubspot.com/static-files/9a43d64a-6547-4a58-a687-c05ca460f087>
5. <https://ir.hubspot.com/node/15186/html>

query_expansion_feedback:

The generated queries are on the right track by targeting press releases and SEC filings (good sources for official executive appointments). A few suggestions to improve precision and reduce noise:

- Use stronger action verbs and appointment language as keywords: appoint/appointed, named, joins, becomes, promoted, steps down, resigns, interim, effective.
- Target specific C-suite titles and variants beyond CEO/CFO/COO/President: include "Chief [Customer/Product/Revenue/Sales/Marketing] Officer", "Chief [X] Officer", and abbreviations (CRO, CCO, CTO) to catch other chief roles.
- Focus on the company's official channels and authoritative distributors: hubspot.com/company-news (newsroom), ir.hubspot.com (investor relations / 8-K exhibits), sec.gov EDGAR (Form 8-K / EX-99.1), and press-distribution sites (Business Wire, PR Newswire). Also check major business outlets (Reuters, Bloomberg, Yahoo Finance) and the company's leadership page/LinkedIn.
- If many results refer to board appointments (directors) rather than executive officers, explicitly exclude board/director terms (e.g. -board -director).
- Expand to filing-specific keywords: Form 8-K + Item numbers typically used for officer appointments/departures (e.g. Item 5.02) and EX-99.1 exhibits, which often contain the actual press-release text.
- If results are returning product updates or community posts, add site restrictions to prioritize corporate / investor pages and newswire domains.

Listing 5: Synthesis — input. Same grouped format as the URL ranking input; each result shows Source and Title, and every URL selected for crawling additionally carries Full crawled content, whose page text is replaced here by the crawled URL in braces. Non-selected results show Source and Title only.

```

Question: Determine whether the company publicly announced a new CEO, CFO, COO, President, or other
chief officer role from January 31, 2026 through April 30, 2026.

Company: HubSpot

Web search results:
Search query: "HubSpot" AND ("press release" OR "press-release") AND (CEO OR CFO OR COO OR President OR
"chief officer") AND (January OR February OR March OR April 2026)

- Source: https://ir.hubspot.com/node/15261/html
; Title: SEC Filing | HubSpot
; Full crawled content: {https://ir.hubspot.com/node/15261/html}
- Source: https://www.hubspot.com/media-kit
; Title: Press Center | Spring 2026 Spotlight
- Source: https://finance.yahoo.com/markets/stocks/articles/hubspot-announces-mike-berry-joins
-201000786.html
; Title: HubSpot Announces Mike Berry Joins Board of Directors
; Full crawled content: {https://finance.yahoo.com/markets/stocks/articles/hubspot-announces-mike-berry-
joins-201000786.html}
- Source: https://ir.hubspot.com/static-files/a4c0bbe9-83ff-47b3-89f6-259fecf58a2e
; Title: HubSpot, Inc. (HUBS)
- Source: https://www.venasolutions.com/newsroom/vena-taps-saas-growth-veteran-and-former-hubspot-chief-
sales-officer-as-new-ceo
; Title: Vena Taps Former HubSpot Chief Sales Officer as CEO
- Source: https://www.customereperiencedive.com/news/hubspot-chief-customer-officer-departure-customer-
success/713765/
; Title: HubSpot nixes chief customer officer role following ...
- Source: https://en.wikipedia.org/wiki/HubSpot
; Title: HubSpot
- Source: https://www.linkedin.com/in/brianhalligan
; Title: Brian Halligan - Long Strange Trip: CEO to ...
- Source: https://www.syntaxdata.com/press
; Title: Media & Press | Syntax
---
Search query: "HubSpot" AND ("Form 8-K" OR "8-K" OR "SEC filing") AND (CEO OR CFO OR COO OR President OR
"chief officer") AND (January OR February OR March OR April 2026)

- Source: https://ir.hubspot.com/static-files/9a43d64a-6547-4a58-a687-c05ca460f087
; Title: Form 8-K for Hubspot INC filed 02/11/2026
; Full crawled content: {https://ir.hubspot.com/static-files/9a43d64a-6547-4a58-a687-c05ca460f087}
- Source: https://www.stocktitan.net/sec-filings/HUBS/page-5.html
; Title: HUBS SEC Filings - Hubspot Inc 10-K, 10-Q, 8-K Forms
- Source: https://www.sec.gov/Archives/edgar/data/1889956/000119312525305275/ck0001889956-20251125.htm
; Title: FORM 8-K
- Source: https://ir.hubspot.com/node/15186/html
; Title: SEC Filing
; Full crawled content: {https://ir.hubspot.com/node/15186/html}
- Source: https://ir.hubspot.com/node/12641/html
; Title: SEC Filing | HubSpot
- Source: https://finance.yahoo.com/quote/HUBS/profile/
; Title: HubSpot, Inc. (HUBS) Company Profile & Facts
- Source: https://community.hubspot.com/t5/Releases-and-Updates/Top-Product-Updates-for-January-2026/ba-
p/1247546
; Title: Top Product Updates for January 2026
- Source: https://www.marketbeat.com/stocks/NYSE/HUBS/sec-filings/
; Title: HubSpot (HUBS) 10K Form and Latest SEC Filings 2026
- Source: https://stocklight.com/stocks/us/nyse-hubs/hubspot/annual-reports/nyse-hubs-2026-10K-26621141.
pdf
; Title: HubSpot Annual Report 2026 Form 10-K (NYSE:HUBS)
---
Search query: HubSpot AND (announce OR announces OR announced OR appoint OR appoints OR appointed OR
named) AND (CEO OR CFO OR COO OR President OR "chief officer") AND (January OR February OR March OR
April 2026)

- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312526138021/hubs-ex99_1.htm
; Title: EX-99.1
; Full crawled content: {https://www.sec.gov/Archives/edgar/data/1404655/000119312526138021/hubs-ex99_1.
html}

```



```

- Source: https://community.hubspot.com/t5/Advocates-Blog/Meet-the-Champion-of-the-Month-March-2026/ba-p/1254364
; Title: Meet the Champion of the Month: March 2026
- Source: https://www.hubspot.com/company-news/hubspot-appoints-kate-bueker-as-chief-financial-officer
; Title: HubSpot Appoints Kate Bueker as Chief Financial Officer
- Source: https://simplywall.st/stocks/de/software/fra-096/hubspot-shares/management
; Title: HubSpot, Inc. (096) Leadership & Management Team ...
- Source: https://www.instagram.com/reel/DZdH00jk9Vy/
; Title: HubSpot Life | we turned twenty. thank you to every ...
- Source: https://unbound.hubspot.com/
; Title: UNBOUND 2026 | HubSpot's Annual Conference
---
Search query: "HubSpot" AND ("leadership change" OR "executive change" OR "management change" OR "
leadership transition") AND (CEO OR CFO OR COO OR President OR "chief officer") AND (January OR February
OR March OR April 2026)

- Source: https://community.hubspot.com/t5/Releases-and-Updates/Top-Product-Updates-for-February-2026/ba-p/1255291
; Title: Top Product Updates for February 2026
- Source: https://community.hubspot.com/t5/Releases-and-Updates/The-January-2026-Industry-Edit-Essential-HubSpot-Updates/ba-p/1245652
; Title: The January 2026 Industry Edit: Essential HubSpot Updates
- Source: https://www.youtube.com/watch?v=Kbo7Q140nzU
; Title: 8 HubSpot Updates Admins Shouldn't Miss (Jan 2026)
- Source: https://www.martech.org/15-hubspot-updates-from-march-2026-managers-and-admins-need-to-know/
; Title: 15 HubSpot updates from March 2026 managers and ...
- Source: https://community.hubspot.com/t5/Releases-and-Updates/The-March-2026-Industry-Edit-Essential-HubSpot-Updates/ba-p/1261823
; Title: The March 2026 Industry Edit: Essential HubSpot Updates
- Source: https://community.hubspot.com/t5/Releases-and-Updates/Top-Product-Updates-for-March-2026/ba-p/1263963
; Title: Top Product Updates for March 2026
- Source: https://www.smartbugmedia.com/inbound-marketing-agency/news-releases/topic/news-release
; Title: SmartBug News and Media Coverage | News Release
- Source: https://www.ainvest.com/news/hubspot-2026-q1-earnings-leadership-change-won-slow-growth-2605/
; Title: Hubspot's 2026 Q1 Earnings: Why the Leadership Change Won't ...
- Source: https://www.youtube.com/watch?v=XIR0rLkp63I
; Title: What's New In HubSpot - January 2026
- Source: https://www.bdbboom.com/blog/hubspot-product-update-april-2026
; Title: HubSpot Product Update: April 2026
---
Search query: "HubSpot" AND (LinkedIn OR "social media" OR "company blog" OR "press briefing") AND (CEO
OR CFO OR COO OR President OR "chief officer") AND (January OR February OR March OR April 2026)

- Source: https://www.linkedin.com/posts/artur-kiulian_i-started-2026-with-an-obsession-to-understand-activity-7417408511904620544-Rblq
; Title: HubSpot's Strategy Secrets: 5 Key Takeaways from Daby O.
- Source: https://www.linkedin.com/posts/lexkeaton_lex-for-2026-presidents-club-i-activity-7416609074789662720--iWd
; Title: Lex for 2026 President's Club I joined HubSpot in mid ...
- Source: https://vantagepoint.io/blog/sf/blog/hs/hubspot-sales-workspace-update-april-2026-preparation-guide
; Title: HubSpot Sales Workspace Update April 2026
- Source: https://www.linkedin.com/posts/hubspot-academy_2026-state-of-marketing-report-insights-activity-7438320357628964864-8p6z
; Title: HubSpot's 2026 State of Marketing Report: Brand POV ...
- Source: https://www.clay.com/dossier/hubspot-ceo
; Title: Who is the CEO of HubSpot in 2026? Yamini Rangan's Bio
- Source: https://cxfoundation.com/news/hubspot-spring-spotlight-2026
; Title: HubSpot Spring Spotlight 2026: 5 Big Talking Points
- Source: https://southbound.revpartners.io/
; Title: Southbound '26 | The GTM Conference for Revenue Leaders
---

```

Listing 6: Synthesis — output. Ground truth is False; the synthesizer's chain of thought was not captured in the trace.

True

Listing 7: Query expansion — input.

Research question: Determine whether the company publicly announced a new CEO, CFO, COO, President, or other chief officer role from January 31, 2026 through April 30, 2026.

Company: HubSpot

Listing 8: Query expansion — output (five Google search queries).

```
1. site:investor.hubspot.com "HubSpot" (announced OR appointed OR "named") (CEO OR CFO OR COO OR President OR "chief") 2026
2. site:sec.gov "HubSpot" "Form 8-K" (CEO OR CFO OR COO OR President OR chief)
3. "HubSpot" (appointed OR "named" OR "hired") (CEO OR CFO OR COO OR President OR "chief officer") Q1 2026
4. site:hubspot.com "HubSpot" "leadership" "executive team" 2026
5. "HubSpot" (resigned OR "will step down" OR "stepping down" OR "resignation") (CEO OR CFO OR COO OR President OR chief) Feb Mar Apr 2026
```

Listing 9: URL ranking — input (original question, company, and the web search results grouped by search query; each result has Source, Title, and Content snippet).

Original question: Determine whether the company publicly announced a new CEO, CFO, COO, President, or other chief officer role from January 31, 2026 through April 30, 2026.

Company: HubSpot

Web search results:

Search query: site:investor.hubspot.com "HubSpot" (announced OR appointed OR "named") (CEO OR CFO OR COO OR President OR "chief") 2026

```
- Source: https://www.hubspot.com/
; Title: HubSpot | Software & Tools for your Business - Homepage
; Content: HubSpot's customer platform includes all the marketing, sales, customer service, and CRM software you need to grow your business.
- Source: https://app.hubspot.com/login
; Title: HubSpot Login and Sign in
; Content: Sign into your HubSpot account through HubSpot's login page. No login? Signup for a free account ... See how HubSpot can help you achieve your growth goals.Read more
- Source: https://www.hubspot.com/careers
; Title: HubSpot Careers
; Content: Ready to do your best work? Explore career opportunities globally and learn more about HubSpot culture.
- Source: https://www.hubspot.com/products/crm
; Title: Streamline Your Entire Business With a Free CRM
; Content: HubSpot's free CRM unifies all your customer data on one platform, with AI that makes it easy to understand. And unlike it's 100% free with no expiration date. ...
- Source: https://academy.hubspot.com/
; Title: HubSpot Academy - Homepage
; Content: HubSpot Academy is the worldwide leader in inbound marketing, sales, and customer service/support training.
- Source: https://blog.hubspot.com/
; Title: HubSpot Blog | Marketing, Sales, Agency, and Customer ...
; Content: HubSpot's Marketing Blog covers everything you need to know to master inbound marketing.
- Source: https://www.linkedin.com/company/hubspot
; Title: HubSpot
; Content: HubSpot is a leading agentic customer platform that provides software and support to help businesses grow better.Read more
- Source: https://www.instagram.com/hubspot/?hl=en
; Title: HubSpot (@hubspot) - Cambridge, MA
; Content: The agentic customer platform to scale your business. quotes, contracts, billing, and payments, built right into HubSpot.
- Source: https://knowledge.hubspot.com/account-management/log-in-to-hubspot
; Title: Log in to HubSpot
; Content: Jan 15, 2026 -- You can log in to HubSpot using several authentication methods, including a password, passkeys, Google, Microsoft, Apple, or single sign-on, ...Read more
---
```

Search query: site:sec.gov "HubSpot" "Form 8-K" (CEO OR CFO OR COO OR President OR chief)

```
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000095017025066998/hubs-20250508.htm
; Title: 8-K
; Content: HUBSPOT, INC. (Exact name of Registrant as Specified in Its Charter) ... Form 8-K filing is intended to simultaneously satisfy the filing obligation of ...Read more
```

- Source: <https://www.sec.gov/Archives/edgar/data/1404655/000119312518191656/d728936d8k.htm>
; Title: 8-K
; Content: Jun 13, 2018 -- On June 13, 2018, HubSpot, Inc. (the "Company") announced that Kathryn Bueker will become the Company's Chief Financial Officer, effective June ...Read more

- Source: https://www.sec.gov/Archives/edgar/data/1404655/000156459017007153/hubs-def14a_20170607.htm
; Title: hubs-def14a_20170607.htm
; Content: We will report final results by filing a Form 8-K within four business days after the Annual Meeting. ... Bohn served as the president, chief ... Prior to HubSpot, ...

- Source: <https://www.sec.gov/Archives/edgar/data/1404655/000095017025018873/hubs-20241231.htm>
; Title: 10-K
; Content: A copy of our Culture Code can be found at: <https://network.hubspot.com/slides/the-hubspot-culture-code>. ... Form 8-K, and amendments to these reports ...Read more

- Source: <https://www.sec.gov/Archives/edgar/data/1404655/000119312516457596/d137374d8k.htm>
; Title: Form 8-K
; Content: Form 8-K. UNITED STATES. SECURITIES AND EXCHANGE COMMISSION. WASHINGTON, D.C. ... HUBSPOT, INC. (Exact name of registrant as specified in its charter) ...Read more

- Source: <https://www.sec.gov/Archives/edgar/data/1404655/000119312516519796/d155967d8k.htm>
; Title: 8-K
; Content: HUBSPOT, INC. (Exact name of registrant as specified in its charter) ... Check the appropriate box below if the Form 8-K filing is intended to ...Read more

- Source: https://www.sec.gov/Archives/edgar/data/1404655/000156459020004381/hubs-10k_20191231.htm
; Title: hubs-10k_20191231.htm
; Content: HubSpot CMS can be purchased as a standalone product and/or with any version of Marketing Hub, Sales Hub, or Service Hub. Features include: website pages, ...Read more

- Source: <https://www.sec.gov/ix?doc=/Archives/edgar/data/0001404655/000095017022021289/hubs-20220930.htm>
; Title: XBRL Viewer
; Content: HubSpot, Inc. (Exact name of registrant as specified in its charter) ... Form 8-K filed on April 14, 2022. (4) Incorporated by reference to Exhibit 4.1 ...Read more

- Source: <https://www.sec.gov/Archives/edgar/data/1404655/000119312525267023/hubs-20251103.htm>
; Title: 8-K
; Content: Nov 3, 2025 -- HUBSPOT, INC. (Exact name of Registrant as Specified in Its Charter) ... Form 8-K filing is intended to simultaneously satisfy the ...Read more

- Source: <https://www.sec.gov/Archives/edgar/data/1404655/000095017021000794/hubs-20210804.htm>
; Title: 8-K
; Content: ... Form 8-K filing is intended to simultaneously ... Chief Executive Officer and President, effective September 7, 2021 and upon Mr. ... HubSpot, Inc. dated August 4, ...

Search query: "HubSpot" (appointed OR "named" OR "hired") (CEO OR CFO OR COO OR President OR "chief officer") Q1 2026

- Source: <https://simplywall.st/stocks/de/software/fra-096/hubspot-shares/management>
; Title: HubSpot, Inc. (096) Leadership & Management Team ...
; Content: HubSpot's CEO is Yamini Rangan, appointed in Sep 2021, has a tenure of 4.75 ... named executive officers; to approve the HubSpot, Inc. 2024 Stock ...Read more

- Source: <https://cust.co/companies/hubspot/>
; Title: HubSpot NRR: 103% - FY2026-Q1 | Cust
; Content: As of its most recent disclosure (FY2026-Q1), HubSpot ... CEO. CCO comp tied to retention: Yes - 70% ARR ... Named CS initiatives across recent disclosures (newest ...

- Source: <https://www.stocktitan.net/news/HUBS/hub-spot-announces-clara-shih-joins-board-of-2sxnaytlr6h.html>
; Title: HubSpot Announces Clara Shih Joins Board of Directors
; Content: Clara Shih joins HubSpot's Board effective Nov 3, 2025. Meta Business AI Group head and former Salesforce AI/Service Cloud CEO to support HubSpot's AI ...

- Source: <https://www.instagram.com/p/C0uMYXmuiQZ/?hl=en>
; Title: HubSpot Life | We're excited to share that our incredible ...
; Content: We're excited to share that our incredible leader, Yamini Rangan, has been named a 2023 Best CEO by Comparably! ... April 9. OCR. HubSpot Q1 2026 ...Read more

- Source: <https://www.aol.com/articles/hubspot-hubs-q1-2026-earnings-224859000.html>
; Title: HubSpot (HUBS) Q1 2026 Earnings Transcript
; Content: May 7, 2026 -- Now it is my pleasure to turn the call over to HubSpot, Inc.'s Chief Executive Officer, Yamini Rangan. Yamini? Yamini Rangan: Thank you, Chuck, ...Read more

- Source: <https://www.morningstar.com/news/business-wire/20260401561380/hubspot-announces-mike-berry-joins-board-of-directors>
; Title: HubSpot Announces Mike Berry Joins Board of Directors
; Content: Apr 1, 2026 -- Berry joins HubSpot's board as the company continues to scale its operations and deepen its financial discipline across global markets. As CFO ...Read more

- Source: <https://www.instagram.com/reel/DYS0GK0J2jM/>
; Title: HubSpot just crashed after earnings, but the CEO stepped in ...
; Content: HubSpot just crashed after earnings, but the CEO stepped in and bought over \$500,000 of stock for the first time ever.

- Source: <https://seekingalpha.com/article/4890988-hubspot-inc-hubs-discusses-2026-strategy-and-ai-driven-innovations-for-growth-companies>

; Title: HubSpot, Inc. (HUBS) Discusses 2026 Strategy and AI ...
; Content: Apr 15, 2026 -- Now it's my pleasure to turn it over to HubSpot's Chief Executive Officer, Yamini Rangan. Yamini Rangan CEO, President and Director. Thank ...Read more
- Source: <https://sherwood.news/markets/cloudflare-hubspot-both-beat-q1-earnings-ai-threats-looming-not-enough-to-satisfy-investors/>
; Title: Cloudflare and HubSpot both beat on Q1 earnings
; Content: May 8, 2026 -- Cloudflare and HubSpot both beat on Q1 earnings -- but with AI threats looming, it wasn't enough to satisfy investors - Revenue of \$640 million, ...Read more

Search query: site:hubspot.com "HubSpot" "leadership" "executive team" 2026

- Source: <https://ir.hubspot.com/governance/management>
; Title: Management
; Content: She first joined Atlassian in 2016 and held a range of leadership roles across commercial and product counsel, as well as head of privacy. ... executive team to ...Read more
- Source: <https://community.hubspot.com/t5/Email-Marketing-Tool/Seed-list-executive-leadership-list/m-p/1278770>
; Title: Seed list/executive leadership list
; Content: May 26, 2026 -- What is the best way to send to a seed list or an executive leadership list who may also be in any lists we include in our "don't send to list" because ...
- Source: <https://www.hubspot.com/company-news/katie-burke-becomes-hubspots-chief-people-officer>
; Title: Katie Burke Becomes HubSpot's Chief People Officer
; Content: As Chief People Officer, Katie will be the first female member of HubSpot's c-level executive team ... leadership, ethnicity, and age. And while there is ...Read more
- Source: <https://offers.hubspot.com/-tx-hubspot-house-mar2026>
; Title: RSVP Today | HubSpot House at SXSW
; Content: Mar 16, 2026 -- Featured Speakers: Kipp Bodnar (CMO, HubSpot), Chloe Tambe (Director, AI Transformation, HubSpot) and more! This event took place on March 16, ...Read more
- Source: <https://ir.hubspot.com/node/12641/html>
; Title: SEC Filing | HubSpot
; Content: We are proud to have a range of diverse perspectives and backgrounds that contribute to the innovation and thought leadership at HubSpot. ... executive team ...Read more
- Source: <https://community.hubspot.com/t5/Data-Management-Reporting/How-do-you-communicate-with-company-leaders/m-p/1262749>
; Title: Re: How do you communicate with company leaders?
; Content: Mar 27, 2026 -- In order to communicate to leadership, it's helpful to identify advocates for you that sit close to leadership. ... hubspot Copyright (c) 2026 ...Read more
- Source: <https://www.hubspot.com/careers-blog/mba-fellows-program>
; Title: How Employees Can Get an MBA at the Office
; Content: Cases help participants understand HubSpot's development, corporate finance, competitive strategy, and leadership. ... The executive team is incredibly ...Read more
- Source: <https://www.hubspot.com/web-guide/apac/resources/sales-in-asia/talent>
; Title: Chapter 5 | Building and Scaling Effective Sales Teams for ...
; Content: Leadership sets the tone -- make sure your leaders and executive team are living the values you want your team to embody. Hire team players and invest in ...Read more
- Source: <https://www.hubspot.com/company-news/tech-veteran-john-kinzer-joins-hubspot-as-cfo>
; Title: Tech Veteran John Kinzer Joins HubSpot As CFO
; Content: John is an outstanding addition to our executive team, and he's ... leadership team. From the mission to the market opportunity to the financial ...Read more
- Source: <https://offers.hubspot.com/the-smarketing-shift-webinar>
; Title: The Smarketing Shift: The Executive Conversation No One Is ...
; Content: The conversation focuses on leadership decision-making, change management, and the real challenges of Sales and Marketing collaboration. HubSpot tools may be ...

Search query: "HubSpot" (resigned OR "will step down" OR "stepping down" OR "resignation") (CEO OR CFO OR COO OR President OR chief) Feb Mar Apr 2026

- Source: <https://ir.hubspot.com/node/15261/html>
; Title: SEC Filing | HubSpot
; Content: On March 27, 2026, Ron Gill notified the Board of Directors (the "Board") of HubSpot, Inc. (the "Company"), that he will resign as a director effective June 30, ...Read more
- Source: <https://www.mediapost.com/publications/article/365746/hubspot-ceo-halligan-steps-down-after-15-years-re.html>
; Title: HubSpot CEO Halligan Steps Down After 15 Years ...
; Content: Aug 5, 2021 -- HubSpot CEO Brian Halligan is stepping down to take on the role of executive chairman. He will be replaced by Yamini Rangan, now the technology firm's chief ...Read more
- Source: <https://simplywall.st/stocks/gb/software/lse-0tzz/hubspot-shares/management>
; Title: HubSpot, Inc. (OTZZ) Leadership & Management Team ...
; Content: Gill's resignation from the Board on June 30, 2026. On March 27, 2026, Ron Gill notified the Board of Directors, that he will resign as a director effective ...
- Source: <https://www.stockinsights.ai/us/HUBS/8-K/management-changes-20260401-cab>
; Title: HUBS HUBSPOT INC Management Changes 8-K Filing
; Content: On March 27, 2026, Ron Gill notified the Board of Directors (the "Board") of HubSpot, Inc. (

```

the "Company"), that he will resign as a director effective June 30, ...
- Source: https://ir.hubspot.com/node/15261/ixbri-viewer
; Title: 0001193125-26-138021 | 8-K | iXBRL Viewer | Hubspot INC
; Content: On March 27, 2026, Ron Gill notified the Board of Directors (the "Board") of HubSpot, Inc. (
the "Company"), that he will resign as a director effective June 30, ...
- Source: https://www.cfo.com/news/cfos-on-the-move-dec-5-Jessica-Ross-GitLab-Bill-Koefoed-OneStream-
Mike-Lenihan-Texas-Roadhouse/807088/
; Title: CFOs On the Move: Week ending Dec. 5
; Content: Dec 5, 2025 -- ... CFO is stepping down and Texas Roadhouse appoints a new chief financial
officer ... chief financial officer of HubSpot and Blackboard, has ...Read more
- Source: https://libraryofminds.com/brian-halligan
; Title: Brian Halligan - Episode 8
; Content: HubSpot's co-founder who literally wrote the book on inbound marketing ... After stepping
down as President of Lyft in 2023 and transitioning to Vice ...Read more
- Source: https://www.bizjournals.com/boston/news/2021/08/04/hubspot-appoints-yamini-rangan-as-ceo.html
; Title: Yamini Rangan to succeed Brian Halligan as HubSpot CEO
; Content: Aug 4, 2021 -- Brian Halligan is stepping down from his position as founding CEO of HubSpot
Inc. after 15 years in the role, and four months after a ...
- Source: https://www.linkedin.com/pulse/undisrupted-hubspots-reflections-disrupted-dan-lyons-dharmesh-
shah
; Title: Undisrupted: HubSpot's Reflections on "Disrupted"
; Content: So, we offered Dan a job at HubSpot, and he accepted. About 20 months later, Dan resigned
from HubSpot and then went on to write "Disrupted".Read more
---
```

Listing 10: URL ranking — output (selected URLs to crawl and feedback on the generated queries).

```

urls:
1. https://ir.hubspot.com/node/15261/ixbri-viewer
2. https://www.morningstar.com/news/business-wire/20260401561380/hubspot-announces-mike-berry-joins-
board-of-directors
3. https://www.hubspot.com/company-news/katie-burke-becomes-hubspots-chief-people-officer
4. https://www.hubspot.com/company-news/tech-veteran-john-kinzer-joins-hubspot-as-cfo
5. https://ir.hubspot.com/governance/management

query_expansion_feedback:
Overall the generated queries are on the right track: searching the company's investor site and the SEC
for Form 8-K filings is exactly where announcements of officer appointments or resignations are usually
recorded. However, the queries were a bit narrow and missed some common sources and wording patterns
that would capture public announcements more completely.

What worked well
- Targeting investor.hubspot.com and sec.gov -- these are the primary, authoritative places for 8-K
disclosures and investor press releases.
- Including action verbs like announced/appointed/named -- good for capturing appointment language.

What to improve / suggestions
- Expand target sites: include hubspot.com/company-news (newsroom/press releases) and third-party press
wire sites (BusinessWire, PR Newswire), major business outlets (Reuters, Bloomberg, Morningstar, Boston
Business Journal, Seeking Alpha). These often host or republish company announcements.
- Broaden title wording: search explicitly for job-title variants and abbreviations (CEO, CFO, COO,
President, CPO, CTO, CMO, CRO, "Chief People Officer", "Chief Product Officer", etc.) and include
synonyms like "joins as", "named", "becomes", "appointed", "is hired as", "assumes role".
- Use SEC/filing language too: include "Form 8-K", "8-K", and optionally "Item 5.02" (or other 8-K items
about officer/director changes) to find regulatory filings about officer departures/appointments.
- Avoid noise from board-only items if you only want C-suite hires: either add "chief" or title keywords
, or explicitly exclude terms like "board" if you want to filter out director appointments.
- Include press-wire and news aggregator names as sources in queries (e.g., BusinessWire, PR Newswire,
Morningstar, Reuters) -- many company press releases are distributed there and will show the
announcement date clearly.

Keywords and site directions to include in follow-up searches (not exact queries):
- Keywords: appointed, named, joins as, becomes, hired as, assumes role, chief, CEO, CFO, COO, President
, Chief People Officer, CPO, CTO, CMO, CRO, "executive team", "executive appointment", "management
change", "Form 8-K", "8-K", "Item 5.02"
- Sites to target: hubspot.com/company-news (newsroom), ir.hubspot.com (press releases & filings), sec.
gov (EDGAR / Form 8-K), BusinessWire, PR Newswire, Morningstar, Reuters, Bloomberg, Seeking Alpha,
Boston Business Journal.

Using those keywords and sites will reduce false positives (board-only items) and make it easier to find
any public announcement of a new CEO/CFO/COO/President or other chief officer in the specified
timeframe.
```

Listing 11: Synthesis — input. Same grouped format as the URL ranking input; each result shows Source and Title, and every URL selected for crawling additionally carries Full crawled content, whose page text is replaced here by the crawled URL in braces. Non-selected results show Source and Title only.

```

Question: Determine whether the company publicly announced a new CEO, CFO, COO, President, or other
chief officer role from January 31, 2026 through April 30, 2026.

Company: HubSpot

Web search results:
Search query: site:investor.hubspot.com "HubSpot" (announced OR appointed OR "named") (CEO OR CFO OR COO
OR President OR "chief") 2026

- Source: https://www.hubspot.com/
; Title: HubSpot | Software & Tools for your Business - Homepage
- Source: https://app.hubspot.com/login
; Title: HubSpot Login and Sign in
- Source: https://www.hubspot.com/careers
; Title: HubSpot Careers
- Source: https://www.hubspot.com/products/crm
; Title: Streamline Your Entire Business With a Free CRM
- Source: https://academy.hubspot.com/
; Title: HubSpot Academy - Homepage
- Source: https://blog.hubspot.com/
; Title: HubSpot Blog | Marketing, Sales, Agency, and Customer ...
- Source: https://www.linkedin.com/company/hubspot
; Title: HubSpot
- Source: https://www.instagram.com/hubspot/?hl=en
; Title: HubSpot (@hubspot) - Cambridge, MA
- Source: https://knowledge.hubspot.com/account-management/log-in-to-hubspot
; Title: Log in to HubSpot
---
Search query: site:sec.gov "HubSpot" "Form 8-K" (CEO OR CFO OR COO OR President OR chief)

- Source: https://www.sec.gov/Archives/edgar/data/1404655/000095017025066998/hubs-20250508.htm
; Title: 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312518191656/d728936d8k.htm
; Title: 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000156459017007153/hubs-def14a_20170607.htm
; Title: hubs-def14a_20170607.htm
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000095017025018873/hubs-20241231.htm
; Title: 10-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312516457596/d137374d8k.htm
; Title: Form 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312516519796/d155967d8k.htm
; Title: 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000156459020004381/hubs-10k_20191231.htm
; Title: hubs-10k_20191231.htm
- Source: https://www.sec.gov/ix?doc=/Archives/edgar/data/0001404655/000095017022021289/hubs-20220930.
htm
; Title: XBRL Viewer
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312525267023/hubs-20251103.htm
; Title: 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000095017021000794/hubs-20210804.htm
; Title: 8-K
---
Search query: "HubSpot" (appointed OR "named" OR "hired") (CEO OR CFO OR COO OR President OR "chief
officer") Q1 2026

- Source: https://simplywall.st/stocks/de/software/fra-096/hubspot-shares/management
; Title: HubSpot, Inc. (096) Leadership & Management Team ...
- Source: https://cust.co/companies/hubspot/
; Title: HubSpot NRR: 103% - FY2026-Q1 | Cust
- Source: https://www.stocktitan.net/news/HUBS/hub-spot-announces-clara-shih-joins-board-of-2sxnaytlr6h
.html
; Title: HubSpot Announces Clara Shih Joins Board of Directors
- Source: https://www.instagram.com/p/COuMYXmuiQZ/?hl=en
; Title: HubSpot Life | We're excited to share that our incredible ...
- Source: https://www.aol.com/articles/hubspot-hubs-q1-2026-earnings-224859000.html
; Title: HubSpot (HUBS) Q1 2026 Earnings Transcript
- Source: https://www.morningstar.com/news/business-wire/20260401561380/hubspot-announces-mike-berry-
joins-board-of-directors

```

```

; Title: HubSpot Announces Mike Berry Joins Board of Directors
; Full crawled content: {https://www.morningstar.com/news/business-wire/20260401561380/hubspot-announces-mike-berry-joins-board-of-directors}
- Source: https://www.instagram.com/reel/DYS0GK0J2jM/
; Title: HubSpot just crashed after earnings, but the CEO stepped in ...
- Source: https://seekingalpha.com/article/4890988-hubspot-inc-hubs-discusses-2026-strategy-and-ai-driven-innovations-for-growth-companies
; Title: HubSpot, Inc. (HUBS) Discusses 2026 Strategy and AI ...
- Source: https://sherwood.news/markets/cloudflare-hubspot-both-beat-q1-earnings-ai-threats-looming-not-enough-to-satisfy-investors/
; Title: Cloudflare and HubSpot both beat on Q1 earnings
---
Search query: site:hubspot.com "HubSpot" "leadership" "executive team" 2026

- Source: https://ir.hubspot.com/governance/management
; Title: Management
; Full crawled content: {https://ir.hubspot.com/governance/management}
- Source: https://community.hubspot.com/t5/Email-Marketing-Tool/Seed-list-executive-leadership-list/m-p/1278770
; Title: Seed list/executive leadership list
- Source: https://www.hubspot.com/company-news/katie-burke-becomes-hubspots-chief-people-officer
; Title: Katie Burke Becomes HubSpot's Chief People Officer
; Full crawled content: {https://www.hubspot.com/company-news/katie-burke-becomes-hubspots-chief-people-officer}
- Source: https://offers.hubspot.com/-tx-hubspothouse-mar2026
; Title: RSVP Today | HubSpot House at SXSW
- Source: https://ir.hubspot.com/node/12641/html
; Title: SEC Filing | HubSpot
- Source: https://community.hubspot.com/t5/Data-Management-Reporting/How-do-you-communicate-with-company-leaders/m-p/1262749
; Title: Re: How do you communicate with company leaders?
- Source: https://www.hubspot.com/careers-blog/mba-fellows-program
; Title: How Employees Can Get an MBA at the Office
- Source: https://www.hubspot.com/web-guide/apac/resources/sales-in-asia/talent
; Title: Chapter 5 | Building and Scaling Effective Sales Teams for ...
- Source: https://www.hubspot.com/company-news/tech-veteran-john-kinzer-joins-hubspot-as-cfo
; Title: Tech Veteran John Kinzer Joins HubSpot As CFO
; Full crawled content: {https://www.hubspot.com/company-news/tech-veteran-john-kinzer-joins-hubspot-as-cfo}
- Source: https://offers.hubspot.com/the-smarketing-shift-webinar
; Title: The Smarketing Shift: The Executive Conversation No One Is ...
---
Search query: "HubSpot" (resigned OR "will step down" OR "stepping down" OR "resignation") (CEO OR CFO OR COO OR President OR chief) Feb Mar Apr 2026

- Source: https://ir.hubspot.com/node/15261/html
; Title: SEC Filing | HubSpot
- Source: https://www.mediapost.com/publications/article/365746/hubspot-ceo-halligan-steps-down-after-15-years-re.html
; Title: HubSpot CEO Halligan Steps Down After 15 Years ...
- Source: https://simplywall.st/stocks/gb/software/lse-0tzz/hubspot-shares/management
; Title: HubSpot, Inc. (OTZZ) Leadership & Management Team ...
- Source: https://www.stockinsights.ai/us/HUBS/8-K/management-changes-20260401-cab
; Title: HUBS HUBSPOT INC Management Changes 8-K Filing
- Source: https://ir.hubspot.com/node/15261/ixbrl-viewer
; Title: 0001193125-26-138021 | 8-K | iXBRL Viewer | Hubspot INC
; Full crawled content: {https://ir.hubspot.com/node/15261/ixbrl-viewer}
- Source: https://www.cfo.com/news/cfos-on-the-move-dec-5-Jessica-Ross-GitLab-Bill-Koefoed-OneStream-Mike-Lenihan-Texas-Roadhouse/807088/
; Title: CFOs On the Move: Week ending Dec. 5
- Source: https://libraryofminds.com/brian-halligan
; Title: Brian Halligan - Episode 8
- Source: https://www.bizjournals.com/boston/news/2021/08/04/hubspot-appoints-yamini-rangan-as-ceo.html
; Title: Yamini Rangan to succeed Brian Halligan as HubSpot CEO
- Source: https://www.linkedin.com/pulse/undisrupted-hubspots-reflections-disrupted-dan-lyons-dharmesh-shah
; Title: Undisrupted: HubSpot's Reflections on "Disrupted"
---

```

Listing 12: Synthesis — output. Ground truth is False; the synthesizer's chain of thought was not captured in the trace.

False

F Synthesizer Trace Examples

The following listings reproduce the synthesizer chain-of-thought and verdict for the example queries discussed in Section 4.2. Each listing shows the question, target company, ground truth, synthesizer model, and the model's verbatim reasoning. The crawl results themselves, including page text, images, and tracking URLs, are omitted for space.

Listing 13: Trace 1 — ai-initiative / DanaHER: synthesizer xhigh (GPT-5.5, xhigh reasoning effort). Ground truth True; verdict True (correct). xhigh deliberates at length and uses its prior that Innovaccer is a healthcare-AI company.

```
QUESTION: Determine whether the company explicitly mentioned a new AI solution, AI feature, or AI
product initiative in investor-facing or regulatory materials from November 1, 2024 through April 30,
2026.
COMPANY: DanaHER
GROUND TRUTH: True
SYNTHESIZER: GPT-5.5, xhigh
VERDICT: True (correct)

SYNTHESIZER REASONING (verbatim):
**Searching for AI initiatives**

I need to find out if the company mentioned any new AI solutions, features, or products in their
investor materials from November 1, 2024, to April 30, 2026. I found a PRNewswire article by DanaHER
from January 9, 2025, which seems relevant since it's within that timeframe. It discusses an investment
partnership with Innovaccer Inc. focused on healthcare AI, but I should check if it explicitly mentions
any new AI solution or product initiative.
**Evaluating AI mentions in reports**

I need to examine all the provided results regarding AI usage in different documents. In the 2025 10-K,
AI is noted only in the forward-looking risk summary. There's a risk factor discussing uncertainties
about AI development but no new solutions mentioned. The 2026 10-Q reflects similar findings, with no
significant mentions of AI. Risk factors appear unchanged too. Finally, in the earnings presentations
and PRNewswire, AI is mentioned in risk statements, lacking new information.
**Investigating investor-facing news**

I need to determine if this content is more investor-facing rather than regulatory. It looks like it's a
press release related to public company news. I want to check if it mentions anything about AI
initiatives. The title refers to an investment partnership in Innovaccer Inc. and is provided by DanaHER
Corporation. The date is January 9, 2025. I'll need to retrieve the details from memory or reconstruct
based on what I have.
**Evaluating AI risk and content**

I'm examining if the risk factor regarding AI use in products only considers risks or if it addresses a
new AI solution, feature, or product initiative. If it's just about risk, my answer should be false. But
it seems PR might be valid. I wonder if the PRNewswire source might not show all content beyond the
title. The full crawled content appears to be truncated after the title, possibly including a base64
image, which complicates things.
**Analyzing AI product announcements**

I'm parsing the phrase "explicitly mentioned a new AI solution." This likely relates to classifying
company announcements in investor materials. A "partnership in Innovaccer" sounds like an "investment
partnership," but it might not indicate a new solution. The company announced this partnership possibly
to develop AI-enabled healthcare solutions. I'm considering potential confusion: this PR isn't "
regulatory materials," but it is investor-facing since it's from PRNewswire. I'll search the title for
more clarity.
**Exploring investment partnership details**

I might have come across a statement about a "new investment partnership" aimed at improving patient
outcomes through accessible value-based healthcare data. There's this quote from someone at DanaHER,
highlighting their commitment to enhancing the availability of innovative therapies and diagnostics.
This partnership with Innovaccer, a notable healthcare AI company, seems to align DanaHER's expertise in
diagnostics with Innovaccer's healthcare data platform and AI capabilities to push forward precision
medicine. However, it might not directly mention a specific product.
**Evaluating AI mentions**

I need to focus on whether the company explicitly mentioned a new AI solution, feature, or product
initiative in their investor or regulatory materials. The prompt suggests looking for AI announcements,
```

and generic risk disclosures won't count. It's about -clarityif a press release is investor-focused and mentions AI, then it qualifies. For example, "Danaher Announces Investment Partnership in Innovaccer" might indicate an initiative, so that would be true. I just want to ensure I'm clear on the requirements !

****Considering investment implications****

I'm thinking about how a partnership or investment in an AI company doesn't automatically mean it's a product from Danaher. The headline about an "Investment Partnership" might hint at a product initiative, but it could be more about strategic investment. For example, while the announcement might indicate that "Danaher expects to benefit from Innovaccer's AI Data Platform," it doesn't necessarily mention a new product. I'll need to inspect the full content more closely to understand the context better!

****Analyzing Danaher and AI Products****

I'm checking if the 10-K business description refers to any AI product names or if it's relevant to the PR about the investment partnership in Innovaccer. I need to determine if Danaher explicitly mentioned AI in their investor-facing or regulatory materials like SEC filings, earnings presentations, and press releases. While I found a press release on PRNewswire from Danaher, I'm unsure if it's truly investor-focused, but public company press releases typically are.

Listing 14: Trace 3 — geographic-expansion / Dell Technologies: synthesizer oss (gpt-oss-120b, high reasoning effort). Ground truth True; verdict True (correct). A weaker synthesizer counts a soft, non-binding government MoU as a footprint expansion.

QUESTION: Determine whether the company expanded into a new country or materially expanded its footprint in an existing country from May 1, 2025 through April 30, 2026.

COMPANY: Dell Technologies

GROUND TRUTH: True

SYNTHESIZER: gpt-oss-120b, high

VERDICT: True (correct)

SYNTHESIZER REASONING (verbatim):

question: "Determine whether the company expanded into a new country or materially expanded its footprint in an existing country from May 1, 2025 through April 30, 2026." Need to answer True if there is evidence of expansion into a new country or material expansion in existing country during that period .

We have sources: Dell 10-K, press releases, and a MoU with Sharjah Digital Department (UAE) dated May 6 2026. That MoU is about AI adoption in Sharjah government. That indicates expansion of footprint in UAE (existing? Dell already operates globally, likely already present). But material expansion? Signing MoU could be considered expansion of footprint in Sharjah (UAE). However question expects yes/no based on evidence. The MoU indicates new initiative in Sharjah, Emirate of Sharjah, UAE, which is likely an expansion of footprint. So answer True.

Check other sources: Boost Run agreement (April 22, 2026) is a purchase agreement, not geographic expansion.

Thus True.

Listing 15: Trace 4 — ontario-operations / Workday: synthesizer qwen(max) (Qwen3.7-Max). Ground truth True; verdict True (correct). qwen(max) answers from its prior (a known Toronto office) rather than from the crawl evidence.

QUESTION: Determine whether the company currently has operations in Ontario, Canada, as of April 30, 2026.

COMPANY: Workday

GROUND TRUTH: True

SYNTHESIZER: Qwen3.7-Max

VERDICT: True (correct)

SYNTHESIZER REASONING (verbatim):

The question asks whether Workday currently has operations in Ontario, Canada, as of April 30, 2026. Looking at the provided search results:

1. Workday Canada homepage (en-ca): Mentions "Workday Canada | The Enterprise AI Platform for HR & Finance", "It's a new work day for Canada."
2. Support page: Mentions "Workday CA", "24-Hour Customer-Centric Support".
3. Form 10-K for Workday, Inc. (Received 03/06/2026, fiscal year ended Jan 31, 2026): Mentions operations, but no specific mention of Ontario, Canada in the snippet, although it's a global company.
4. Canoe Procurement Group of Canada RFP: Mentions "Canoe members also include public entities in Ontario...". This is about a procurement group, not Workday's operations specifically, though Workday might bid on it.
5. BC Laws Corporate Registry Notices (Jan 28, 2016): Mentions "WORKDAY CANADA HOLDINGS LTD. to

MEDIACORE TECHNOLOGIES INC."

Wait, does Workday have operations in Ontario, Canada? Workday is a major global enterprise software company. They definitely have operations in Canada (Toronto, Ontario is a major tech hub and Workday has an office there).

Wait, the prompt says "as of April 30, 2026".

Is there any evidence in the text that they *don't* have operations in Ontario? No.

Actually, Workday has a well-known office in Toronto, Ontario.

Let me double check if the prompt requires a specific extraction or if it's a general knowledge question backed by the search results. The search results show "Workday Canada" exists.

Usually, questions like "Determine whether the company currently has operations in [Location]" for major multinational companies like Workday are True if they have an office or operations there. Workday has an office in Toronto, Ontario.

Wait, let's think if the question is a trick. "as of April 30, 2026". The 10-K is for the fiscal year ended Jan 31, 2026.

I will answer True. Workday has operations in Canada, and Toronto, Ontario is their main Canadian hub.

Wait, let me just confirm there's no weird detail like "Workday closed its Ontario office". Nothing in the text suggests that.

I'll output True.

G Synthesis Input

Listing 16: Raw synthesis input for ask-web-qexp-synt-nano on the HubSpot c-suite-change query (system prompt omitted; user message shown verbatim). Crawled page contents are included in full to show the amount of navigation/boilerplate the synthesizer must read; only embedded base64 image data is elided and non-ASCII glyphs are normalized to ASCII.

Question: Determine whether the company publicly announced a new CEO, CFO, COO, President, or other chief officer role from January 31, 2026 through April 30, 2026.

Company: HubSpot

Web search results:

Search query: site:investor.hubspot.com "HubSpot" (announced OR appointed OR "named") (CEO OR CFO OR COO OR President OR "chief") 2026

- Source: https://www.hubspot.com/
; Title: HubSpot | Software & Tools for your Business - Homepage
- Source: https://app.hubspot.com/login
; Title: HubSpot Login and Sign in
- Source: https://www.hubspot.com/careers
; Title: HubSpot Careers
- Source: https://www.hubspot.com/products/crm
; Title: Streamline Your Entire Business With a Free CRM
- Source: https://academy.hubspot.com/
; Title: HubSpot Academy - Homepage
- Source: https://blog.hubspot.com/
; Title: HubSpot Blog | Marketing, Sales, Agency, and Customer ...
- Source: https://www.linkedin.com/company/hubspot
; Title: HubSpot
- Source: https://www.instagram.com/hubspot/?hl=en
; Title: HubSpot (@hubspot) - Cambridge, MA
- Source: https://knowledge.hubspot.com/account-management/log-in-to-hubspot
; Title: Log in to HubSpot

Search query: site:sec.gov "HubSpot" "Form 8-K" (CEO OR CFO OR COO OR President OR chief)

- Source: https://www.sec.gov/Archives/edgar/data/1404655/000095017025066998/hubs-20250508.htm
; Title: 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312518191656/d728936d8k.htm
; Title: 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000156459017007153/hubs-def14a_20170607.htm
; Title: hubs-def14a_20170607.htm
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000095017025018873/hubs-20241231.htm
; Title: 10-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312516457596/d137374d8k.htm
; Title: Form 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312516519796/d155967d8k.htm
; Title: 8-K

```

- Source: https://www.sec.gov/Archives/edgar/data/1404655/000156459020004381/hubs-10k_20191231.htm
; Title: hubs-10k_20191231.htm
- Source: https://www.sec.gov/ix?doc=/Archives/edgar/data/0001404655/000095017022021289/hubs-20220930.htm
; Title: XBRL Viewer
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000119312525267023/hubs-20251103.htm
; Title: 8-K
- Source: https://www.sec.gov/Archives/edgar/data/1404655/000095017021000794/hubs-20210804.htm
; Title: 8-K
---
Search query: "HubSpot" (appointed OR "named" OR "hired") (CEO OR CFO OR COO OR President OR "chief officer") Q1 2026

- Source: https://simplywall.st/stocks/de/software/fra-096/hubspot-shares/management
; Title: HubSpot, Inc. (096) Leadership & Management Team ...
- Source: https://cust.co/companies/hubspot/
; Title: HubSpot NRR: 103% - FY2026-Q1 | Cust
- Source: https://www.stocktitan.net/news/HUBS/hub-spot-announces-clara-shih-joins-board-of-2sxnaytlr6h.html
; Title: HubSpot Announces Clara Shih Joins Board of Directors
- Source: https://www.instagram.com/p/COuMYXmUIQZ/?hl=en
; Title: HubSpot Life | We're excited to share that our incredible ...
- Source: https://www.aol.com/articles/hubspot-hubs-q1-2026-earnings-224859000.html
; Title: HubSpot (HUBS) Q1 2026 Earnings Transcript
- Source: https://www.morningstar.com/news/business-wire/20260401561380/hubspot-announces-mike-berry-joins-board-of-directors
; Title: HubSpot Announces Mike Berry Joins Board of Directors
; Full crawled content: HubSpot Announces Mike Berry Joins Board of Directors | Morningstar

Morningstar brands and products

Company

[Morningstar]()

[ ] (/login?destination=https://www.morningstar.com/news/business-wire/20260401561380/hubspot-announces-mike-berry-joins-board-of-directors)

* [Portfolio] (/tools/portfolio)
* Tools
* Sections
* Markets
* [Funds] (/funds)
* [ETFs] (/topics/etfs)
* [Stocks] (/stocks)
* [Bonds] (/bonds)
* Investing Ideas
* [Home] ()
* Tools
  * [Portfolio] (/tools/portfolio)
  * [Watchlists] (/tools/watchlists)
  * [Screener] (/tools/screener)
  * [Chart] (/tools/chart)
  * [Rating Changes] (/tools/rating-changes)
* Sections
* Markets
* [Funds] (/funds)
* [ETFs] (/topics/etfs)
* [Stocks] (/stocks)
* [Bonds] (/bonds)
* Investing Ideas
* Media
* [Help] (/help-center)
* [What's New] (/whats-new)
* [Notifications] (/notifications)
* Products for Investors
* [All Products and Services] (/business)

1. [Home] ()
2. [News] (/news)
3. [Business Wire] (/news/business-wire)
4. HubSpot Announces Mike Berry Joins Board of Directors

```


HubSpot Announces Mike Berry Joins Board of Directors

Provided by Business Wire Apr 1, 2026, 8:10:00 PM

HubSpot Announces Mike Berry Joins Board of Directors

HubSpot, Inc. (NYSE: HUBS), the agentic customer platform for scaling companies, announced today that Mike Berry, Chief Financial Officer of MongoDB, has been appointed to the company's Board of Directors, effective April 1, 2026. Berry brings more than 30 years of experience in finance, operations, and general management across the technology industry, with a track record of driving profitable growth at leading public technology companies.

Berry joins HubSpot's board as the company continues to scale its operations and deepen its financial discipline across global markets.

As CFO of MongoDB, Berry oversees the company's global finance organization, including financial strategy, planning and analysis, accounting, and investor relations. He has served as CFO at several companies across the technology industry, including NetApp and McAfee. Berry holds a bachelor's degree in finance from Augsburg College and an MBA from the University of St. Thomas. He also serves on the board of directors of Rapid7 and Calix.

"Mike is an exceptional leader with deep financial expertise, a strong track record in governance, and a clear understanding of how AI is changing the technology landscape," said Yamini Rangan, CEO of HubSpot. "His vast experience as a CFO at scaling companies, combined with his cybersecurity governance expertise, will be invaluable as we scale HubSpot's agentic customer platform globally. Mike is among the most trusted finance stewards in our industry, and we're thrilled to welcome him to our board."

"I've spent my career working with technology companies at pivotal moments of growth, and HubSpot is squarely in that moment," said Mike Berry. "The combination of a loyal customer base, a differentiated AI roadmap, and a leadership team with real operational focus makes this an exciting time to join the board."

About HubSpot

HubSpot (NYSE: HUBS) is the agentic customer platform that helps businesses connect and grow better. HubSpot delivers seamless connection for customer-facing teams with a unified platform that includes AI-powered engagement hubs, a Smart CRM, and a connected ecosystem with over 2,000 App Marketplace integrations, a community network, and educational content. Learn more at [www.hubspot.com](https://cts.businesswire.com/ct/CT?id=smartlink&url=http%3A%2F%2Fwww.hubspot.com&esheet=54473357&newsitemid=20260401561380&lan=en-US&anchor=www.hubspot.com&index=1&md5=68731591b633a88cab3107c67db73138).

[base64 image data omitted]

[ksoward@Hubspot.com] (mailto:ksoward@Hubspot.com)

View source version on businesswire.com: https://www.businesswire.com/news/home/20260401561380/en/

The articles, information, and content displayed on this webpage may include materials prepared and provided by third parties. Such third-party content is offered for informational purposes only and is not endorsed, reviewed, or verified by Morningstar.

Morningstar makes no representations or warranties regarding the accuracy, completeness, timeliness, or reliability of any third-party content displayed on this site. The views and opinions expressed in third-party content are those of the respective authors and do not necessarily reflect the views of Morningstar, its affiliates, or employees.

Morningstar is not responsible for any errors, omissions, or delays in this content, nor for any actions taken in reliance thereon. Users are advised to exercise their own judgment and seek independent financial advice before making any decisions based on such content. The third-party providers of this content are not affiliated with Morningstar, and their inclusion on this site does not imply any form of partnership, agency, or endorsement.

Popular

[View All](/popular)

5 of the Best Core Stocks to Own

Plus, how to manage core stock positions.

David Sekera, CFA and Susan Dziubinski Jun 15, 2026

Why Most Stocks Aren't Worth Owning

A small number of stocks drive most of the market's long-term return.

Dan Lefkowitz Jun 17, 2026

7 of the Best Funds You May Not Have Heard Of
Forget Vanguard and Fidelity: These top-rated stock funds from under-the-radar managers deserve a look.
Susan Dziubinski Jun 17, 2026

This Stock Offers a 6% Dividend Yield and Looks 14% Undervalued
This high-yield stock could be a smart buy after the growth rally.
Michael Hodel, CFA Jun 11, 2026

3 Stocks to Invest In to Rebalance Your Portfolio
These value stocks can balance portfolios heavy in growth stocks -- and they're undervalued, too.
Susan Dziubinski Jun 16, 2026

Sponsor Center

Transparency is our policy. Learn how it impacts everything we do
Read More

Transparency is how we protect the integrity of our work and keep empowering investors to achieve their goals and dreams. And we have unwavering standards for how we keep that integrity intact, from our research and data to our policies on content and your personal data.
We'd like to share more about how we work and what drives our day-to-day business.

How we make money
We sell different types of products and services to both investment professionals and individual investors. These products and services are usually sold through license agreements or subscriptions. Our investment management business generates asset-based fees, which are calculated as a percentage of assets under management. We also sell both admissions and sponsorship packages for our investment conferences and advertising on our websites and newsletters.

How we use your personal data
How we use your information depends on the product and service that you use and your relationship with us. We may use it to:
* Verify your identity, personalize the content you receive, or create and administer your account.
* Provide specific products and services to you, such as portfolio management or data aggregation.
* Develop and improve features of our offerings.
* Gear advertisements and other marketing efforts towards your interests.
To learn more about how we handle and protect your data, visit our [privacy center](/company/privacy-center).

How we approach editorial content
Maintaining independence and editorial freedom is essential to our mission of empowering investor success. We provide a platform for our authors to report on investments fairly, accurately, and from the investor's point of view. We also respect individual opinions -- they represent the unvarnished thinking of our people and exacting analysis of our research processes. Our authors can publish views that we may or may not agree with, but they show their work, distinguish facts from opinions, and make sure their analysis is clear and in no way misleading or deceptive.
To further protect the integrity of our editorial content, we keep a strict separation between our sales teams and authors to remove any pressure or influence on our analyses and research.
Read our [editorial policy](/editorial-policy) to learn more about our process.

Site Index

[Morningstar](/)

(c) Copyright 2026 Morningstar, Inc. All rights reserved. Dow Jones Industrial Average, S&P 500, Nasdaq, and Morningstar Index (Market Barometer) quotes are real-time.

* [Terms and Conditions](https://www.morningstar.com/company/terms-and-conditions)

* [Privacy Center](https://www.morningstar.com/company/privacy-center)

* [Disclosures](https://www.morningstar.com/company/disclosures)

- Source: https://www.instagram.com/reel/DYSOGK0J2jM/

; Title: HubSpot just crashed after earnings, but the CEO stepped in ...

- Source: https://seekingalpha.com/article/4890988-hubspot-inc-hubs-discusses-2026-strategy-and-ai-driven-innovations-for-growth-companies

; Title: HubSpot, Inc. (HUBS) Discusses 2026 Strategy and AI ...

- Source: https://sherwood.news/markets/cloudflare-hubspot-both-beat-q1-earnings-ai-threats-looming-not-enough-to-satisfy-investors/

; Title: Cloudflare and HubSpot both beat on Q1 earnings

Search query: site:hubspot.com "HubSpot" "leadership" "executive team" 2026

- Source: https://ir.hubspot.com/governance/management

; Title: Management

; Full crawled content: Management | HubSpot

[Skip to main navigation](#main-menu)

[base64 image data omitted]

- * [Overview](/investor-relations)
- * [News](/news)
- * [Events & Presentations](/events-presentations/events)
- * [Financials](/financials/quarterly-results)
- * [Governance](/governance/governance-overview)
 - * [Governance Overview](/governance/governance-overview)
 - * [Board of Directors](/governance/board-of-directors)
 - * [Management](/governance/management)
 - * [Committee Composition](/governance/committee-composition)
- * [Stock](/stock/stock-quote-chart)
- * [Resources](/resources/investor-faqs)

Management

Governance

- * [Governance Overview](/governance/governance-overview)
- * [Board of Directors](/governance/board-of-directors)
- * [Management](/governance/management)
- * [Committee Composition](/governance/committee-composition)

[Yamini Rangan](/management/yamini-rangan)

Chief Executive Officer

Yamini Rangan is chief executive officer at HubSpot. Prior to becoming CEO, Yamini served as HubSpot's first-ever chief customer officer, overseeing the marketing, sales, and service teams. A tech industry veteran, Yamini has more than 25 years of experience ranging from product marketing, sales, and strategy. Yamini previously served as Chief Customer Officer at Dropbox, where she was responsible for embedding customer focus across the organization. Before Dropbox, she was VP of Sales Strategy and Operations at Workday, where she helped quadruple revenue and scaled the sales organization. At SAP, Yamini held several customer-facing leadership roles in strategy, pre-sales and value-based selling and helped close landmark deals. She holds a Bachelor's degree in Electronics Engineering, a Master's in Computer Engineering, and an MBA. Yamini was recognized as one of the Most Influential Women in Business by San Francisco Business Times in 2019 and was recognized as the best CEO for Women by Comparably in 2022. Yamini currently serves on the Board of Haas School of Business, Berkeley and was previously on the Board of Splunk.

[Dharmesh Shah](/management/dharmesh-shah)

Co-Founder, CTO

Dharmesh Shah is co-founder and CTO of HubSpot. Prior to founding HubSpot in 2006, Dharmesh was founder and CEO of Pyramid Digital Solutions, which was acquired by SunGard Data Systems in 2005. In addition to co-authoring Inbound Marketing: Get Found Using Google, Social Media and Blogs, Dharmesh founded and writes for OnStartups.com -- a top-ranking startup blog and community with more than 700,000 members. In 2013, Dharmesh published HubSpot's Culture Code, which has garnered over five million views. Named an Inc. Founders 40 in 2016, he is an active member of the Boston-area entrepreneurial community, an angel investor in over 60 startups, and a frequent speaker on startups, inbound marketing, and company culture. Dharmesh holds a BS in computer science from UAB and an MS in the management of technology from MIT.

[Duncan Lennox](/management/duncan-lennox)

Chief Product & Technology Officer

Duncan Lennox is the chief product and technology officer at HubSpot, where he is responsible for the HubSpot product portfolio and leads Engineering, Product, UX, Next Bets, and IT & Security, in service of our mission to help millions of organizations grow. Prior to HubSpot, Duncan held multiple roles at Google and Amazon Web Services, most recently GM for the Applied AI team in Google Cloud, which builds conversational and agentic AI enterprise applications. Previous roles at Google included leading Ads Privacy and Safety, as well as Reliability Engineering for the largest consumer services in the world, including Search, Maps and Ads. At AWS, Duncan was the general manager for Elastic File System, an enterprise storage service. Duncan also spent two decades as the co-founder of two enterprise SaaS startups, where he held both CEO and CTO roles. He holds a bachelor's degree in computer science and math from University College Dublin. Originally from Dublin, Ireland, Duncan lives in the San Francisco Bay Area with his wife, three daughters, and two dogs, Seamus and Reilly. He can often be found drinking Irish tea and talking endlessly about Star Trek or airplanes.

[Kate Bueker](/management/kate-bueker)

Chief Financial Officer

Kate Bueker is the chief financial officer at HubSpot, where she manages the company's finances, and is also the co-executive sponsor of the Women@HubSpot employee resource group. Kate has also served on the Procore Technologies, Inc. board of directors since April 2021.

Prior to HubSpot, Kate spent eleven years in finance and business development leadership roles at Akamai Technologies, and most recently served as senior vice president of business finance and operations.

During her tenure, Kate was instrumental in building out the finance business support capabilities and function at Akamai. Kate also spent almost 10 years in investment banking at The Blackstone Group, UBS, Credit Suisse and Donaldson, Lufkin & Jenrette.

Outside of work, Kate loves to travel (especially throughout Italy) and enjoys tennis and golf. Kate received her BA in mathematics from Cornell University and her MBA from the Massachusetts Institute of Technology. Kate currently serves on the boards of Procore and Bridge Over Troubled Waters.

[Erika Fisher](/management/erika-fisher)
Chief Legal Officer

Erika Fisher is chief legal officer at HubSpot. She leads the company's global legal and compliance functions, with a focus on ensuring that HubSpot's products and customer experiences meet the highest standards of trust, privacy, and safety. Erika plays a key role in helping HubSpot build AI that is fast, easy, secure, and grounded in strong ethical and regulatory frameworks. She serves on the board of the Business Software Alliance.

Prior to HubSpot, Erika was chief administrative and legal officer at Atlassian, where she oversaw legal (including policy, compliance, and government affairs), people and talent (including the future of distributed work teams), real estate, facilities, and customer support. She also managed Atlassian's board of directors.

A tech industry legal veteran, Erika has more than 15 years of experience guiding high-growth companies through evolving regulatory landscapes. She first joined Atlassian in 2016 and held a range of leadership roles across commercial and product counsel, as well as head of privacy. Before that, she practiced law at Weil, Gotshal & Manges LLP and Goodwin Procter LLP, advising early-stage technology companies on licensing and transactional matters.

Erika holds a Juris Doctor from the University of Pennsylvania, where she was a Silverman-Rodin merit scholar (awarded to a select two members of each class year), and a Certificate in Business and Public Policy from the Wharton School. She earned her bachelor's degree in art history and political science from the University of Miami as a Foote Fellows merit scholar.

[Helen Russell](/management/helen-russell)
Chief People Officer

As chief people officer at HubSpot, Helen Russell has global responsibility for the company's most important asset -- its people. She leads recruiting; learning and talent management; total rewards; culture; environmental, social, and governance (ESG); diversity, inclusion, and belonging (DI&B); and people operations. Helen is passionate about nurturing a culture that accelerates business and employee growth. Seasoned in scaling organizations, she brings over 25 years of people leadership experience across multiple industries, geographies, and sizes of businesses.

Prior to joining HubSpot, Helen served as chief people officer of Rivian Inc., a key leader in the electric vehicle space. She also held chief human resource officer roles at Atlassian Inc., a leading B2B SaaS provider; Sonos Inc., a leader in the home audio space; and Kantar, one of the world's largest research, data, and insight consultancies. Helen began her career in HR at Siebel Systems before joining Yahoo! as VP of HR, EMEA.

She takes pride in helping others do the best work of their careers and is a dedicated mentor to emerging and executive leaders across the globe. Helen is often asked to speak on the business benefits of strong company culture. Helen holds a bachelor's degree from the University of Liverpool, England.

[Karen Ng](/management/karen-ng)
EVP, Product

Karen Ng is the executive vice president of product at HubSpot, where she leads product strategy and execution across product management, design, and partnerships with a mission to help millions of organizations grow better.

Since joining HubSpot in 2022, Karen has driven transformative impact across Smart CRM, Operations Hub, Breeze Intelligence, and HubSpot's developer platform and partner ecosystem. She has championed HubSpot's pioneering work in unified and unstructured data -- the foundation that makes AI actually useful for growing businesses. Known for leading with clarity, courage, and "AI speed," Karen exemplifies the rapid innovation needed to help businesses not just survive disruption but lead through it.

Prior to HubSpot, Karen held head of product roles at Common Room, Google, and Microsoft. At Common Room, a customer intelligence platform, she founded and scaled the product and data science teams, taking the product from 0-1 and driving its first revenue milestones. At Google, she served as director of product management for Android, creating core developer frameworks like Jetpack and Jetpack Compose that power leading mobile apps worldwide. During her decade-plus tenure at Microsoft, she led the company's .NET programming language strategy, developed enterprise collaboration tools and reporting, and led the groundbreaking Roslyn compiler platform.

Karen is a recognized industry leader, named a Product 50 Winner for "Best Digital Native Product Leader" and recipient of the 2023 PM Award for "Technical Strategist" by Products That Count. She is an active

advisor and investor in high-growth technology companies.
Karen holds a master's degree from MIT, where she focused on machine learning, and a bachelor's degree in computer science from Wellesley College. She is passionate about transformative technology, strong team culture, and building products that make people's lives better.

[Jared Williams](/management/jared-williams)
EVP, Engineering

Jared Williams is the executive vice president of engineering at HubSpot, where he leads the company's engineering teams and technical strategy.
Since joining HubSpot in 2011, Jared has held a range of engineering leadership roles. Earlier in his career he was one of the founding engineers of HubSpot's CRM, where he led teams building HubSpot's customer data platform, scaling its core infrastructure to support billions of records and high-throughput systems. He helped drive the acquisition and integration of PieSync and played a key role in launching Operations Hub, where he continues to lead and scale HubSpot's underlying platform.
Jared holds a BS in computer science from SUNY Albany.

[Chris Hogan](/management/chris-hogan)
SVP, Operations

Chris Hogan is the senior vice president of operations at HubSpot. His teams are responsible for revenue operations, business operations (strategy and execution, planning, program management, pricing strategy), product operations, and global data and intelligence. Chris and his teams partner closely with functional leaders across HubSpot to manage, plan, and operate each area of the business efficiently and effectively.
Before taking on operations, Chris started the corporate strategy function, and scaled HubSpot's product strategy and operations team with a push to bring a freemium go-to-market model across all product lines. He was in the first class of the company's post-MBA Accelerated Leadership Program, a two-year cross-functional rotational program. Prior to joining HubSpot, Chris worked in investment banking at Barclays, where he covered infrastructure software and small-cap SaaS.
Chris earned his MBA at the Tuck School of Business at Dartmouth, where he was a fellow in both the Center for Digital Strategies and Center for Private Equity and Entrepreneurship. He earned his bachelor's degree in finance and accounting from Boston College and is also a chartered financial analyst and CPA in Massachusetts.

[Kipp Bodnar](/management/kipp-bodnar)
Chief Marketing Officer

Kipp Bodnar is the chief marketing officer of HubSpot, where he sets HubSpot's global marketing strategy to drive awareness and demand for HubSpot's products.
Prior to his role as CMO, Kipp served as vice president of marketing at HubSpot, overseeing all demand generation activity worldwide, building out the EMEA and APAC marketing teams, and managing HubSpot's field marketing, localization, strategic partnerships, and social media efforts.
Kipp serves as a marketing advisor for SaaS companies including SimplyMeasured, InsightSquared and Guidebook. He is also the co-author of The B2B Social Media Book: Become a Marketing Superstar by Generating Leads with Blogging, LinkedIn, Twitter, Facebook, Email, and More. Kipp is an industry-leading speaker and blogger and holds a BA in journalism from Marshall University.

[Alison Elworthy](/management/alison-elworthy)
Chief of Staff and EVP of Central Strategy

Alison Elworthy is chief of staff to the CEO and executive vice president of central strategy at HubSpot, where she partners with the executive team to set and communicate HubSpot's strategy.
Prior to her role as chief of staff and EVP of strategy, Alison led various functions at HubSpot including EVP of revenue operations, SVP of customer success, and VP of operations, helping to scale HubSpot over the past decade. Alison also serves on the board of directors at Varicent and the Posse Foundation. She holds a BS in human factors engineering from Tufts University and an MBA from the Tuck School of Business at Dartmouth.

[Jon Dick](/management/jon-dick)
Chief Customer Officer

Jon Dick is the chief customer officer at HubSpot, where he oversees the global sales and customer success organizations, helping our 280,000+ customers grow better.
Before this role, Jon was chief customer success officer at HubSpot. He also led the flywheel marketing team at HubSpot, which combined product marketing, digital marketing, and customer marketing to take our products to market worldwide. During his tenure, he helped transform HubSpot's approach to demand generation by scaling freemium, innovating on automation and AI, and delivering new global marketing models. Before joining HubSpot, Jon worked in B2C, leading marketing teams at Klout and Trunk Club.
Jon studied physics at the University of Massachusetts and Oxford University and earned his MBA from the Harvard Business School, where he was elected co-president of the student body. He also spent a decade performing improvised comedy, but now that he's a dad, he only tells one kind of joke.

```
## Investor Contact
[investors@hubspot.com](mailto:investors@hubspot.com)
Copyright (c) 2026 HubSpot
- Source: https://community.hubspot.com/t5/Email-Marketing-Tool/Seed-list-executive-leadership-list/m-p/1278770
; Title: Seed list/executive leadership list
- Source: https://www.hubspot.com/company-news/katie-burke-becomes-hubspots-chief-people-officer
; Title: Katie Burke Becomes HubSpot's Chief People Officer
; Full crawled content: Katie Burke Becomes HubSpot's Chief People Officer

[Skip to content](#blog-main-content)

* [Home](https://www.hubspot.com)
* [Company-news](https://www.hubspot.com/company-news)
* Katie Burke Becomes HubSpot's Chief People Officer

# Katie Burke Becomes HubSpot's Chief People Officer

[Laura Moran](https://www.hubspot.com/company-news/author/laura-moran)

Today we're excited to announce that HubSpot has appointed Katie Burke as Chief People Officer. Katie has served as HubSpot's VP of Culture and Experience for the last two years and has been an invaluable part of establishing company culture globally as HubSpot has expanded. In this new role, Katie will continue to oversee the company's culture and experience initiatives as well as all people operations, including global recruitment and strategy and human resources.

As Chief People Officer, Katie will be the first female member of HubSpot's c-level executive team, bringing with her a strong passion for advancing HubSpot's continued efforts around diversity and inclusion. Throughout HubSpot's ten-year history, we've worked hard to create an inclusive culture that celebrates the diversity of our workforce. It's an ongoing journey -- one that requires thoughtful and consistent attention from the top down. Earlier this month, we released our [2016 Diversity Data](https://www.hubspot.com/company-news/hubspot-publishes-its-2016-company-diversity-data), detailing the company's diversity in leadership, ethnicity, and age. And while there is still work to be done, HubSpot, with Katie leading the charge, is committed to making meaningful changes to move those numbers in the right direction.

During Katie's time at HubSpot, the company's culture has been recognized by Entrepreneur, Fortune, The Boston Globe, [Glassdoor](https://www.glassdoor.com/Overview/Working-at-HubSpot-EI_IE227605.11,18.htm), and The Boston Business Journal. Her emphasis on a culture of inclusion, diversity, and support also led to HubSpot being placed on Fortune.com's "100 Best Workplaces for Women" list. Katie has overseen the openings of HubSpot's four international offices, making sure that award-winning culture is in place in all of our locations around the globe. Prior to leading the culture team, Katie oversaw HubSpot's internal and external communications, including the company's 2014 IPO.

"Since day one, Katie has been an invaluable member of the company," said Brian Halligan, Co-Founder and CEO of HubSpot. "From her work on our diversity and inclusion data, to the award-winning culture she's helped to create here, Katie has played a pivotal role in getting HubSpot where it is today. She's tireless in her aim to help build a world-class company. We can't wait to see what she can achieve in this new role."

Katie is taking on this role in the footsteps of HubSpot's first CPO Jim O'Neill, who up until May 2015 served as the company's CIO. Jim is stepping back to take a proper and well-deserved sabbatical after more than 9 years at HubSpot. When he returns, Jim will be getting back to his CIO roots by taking on special projects and more technical work.

To learn more about Katie's plans and goals for the role, check out her post on [ThinkGrowth.org](https://thinkgrowth.org/the-path-to-making-people-a-priority-4fbb1f181b8c).

[base64 image data omitted]

Copyright (c) 2026 HubSpot, Inc.
- Source: https://offers.hubspot.com/-tx-hubspothouse-mar2026
; Title: RSVP Today | HubSpot House at SXSW
- Source: https://ir.hubspot.com/node/12641/html
; Title: SEC Filing | HubSpot
- Source: https://community.hubspot.com/t5/Data-Management-Reporting/How-do-you-communicate-with-company-leaders/m-p/1262749
; Title: Re: How do you communicate with company leaders?
- Source: https://www.hubspot.com/careers-blog/mba-fellows-program
; Title: How Employees Can Get an MBA at the Office
- Source: https://www.hubspot.com/web-guide/apac/resources/sales-in-asia/talent
; Title: Chapter 5 | Building and Scaling Effective Sales Teams for ...
- Source: https://www.hubspot.com/company-news/tech-veteran-john-kinzer-joins-hubspot-as-cfo
```



```
; Title: Tech Veteran John Kinzer Joins HubSpot As CFO
; Full crawled content: Tech Veteran John Kinzer Joins HubSpot As CFO
```

```
[Skip to content](#blog-main-content)
```

```
* [Home](https://www.hubspot.com)
* [Company-news](https://www.hubspot.com/company-news)
* Tech Veteran John Kinzer Joins HubSpot As CFO
```

```
# Tech Veteran John Kinzer Joins HubSpot As CFO
```

```
[Hannah Fleishman](https://www.hubspot.com/company-news/author/hannah-fleishman)
```

Today, inbound marketing software company HubSpot announced that technology veteran John Kinzer will join the company as Chief Financial Officer. Kinzer has 23 years of experience in management, accounting, and finance in the software and tech space, most recently as CFO of BackOffice Associates. Prior to that, he was the CFO of Blackboard, a leading SaaS company in Washington, DC, and previously served in leadership roles at e-centives and MCI. Kinzer began his career at Arthur Andersen.

Of the announcement, HubSpot's Chief Operating Officer JD Sherman noted: "HubSpot has had a big year in 2013. We've opened our European headquarters in Dublin, launched the new HubSpot platform and entered a brand new market with Signals, and expanded our presence in Cambridge to accommodate our rapidly growing team. To build a once in a generation company, we needed a CFO with the experience to build a finance team that will scale with our business and delight our customers and investors alike. John is an outstanding addition to our executive team, and he's joining at an incredibly important juncture in our company's growth."

Kinzer echoed Sherman's sentiment, noting "HubSpot is one of the fastest-growing and most exciting SaaS companies in the world, and I'm honored to join such a strong executive leadership team. From the mission to the market opportunity to the financial model, HubSpot is well positioned for massive growth and expansion. I look forward to working hand in hand with the team to define, develop, and deliver systems that drive results for our shareholders, our employees, our customers, and our business."

```
[base64 image data omitted]
```

Copyright (c) 2026 HubSpot, Inc.

```
- Source: https://offers.hubspot.com/the-smarketing-shift-webinar
; Title: The Smarketing Shift: The Executive Conversation No One Is ...
```

```
---
```

```
Search query: "HubSpot" (resigned OR "will step down" OR "stepping down" OR "resignation") (CEO OR CFO OR COO OR President OR chief) Feb Mar Apr 2026
```

```
- Source: https://ir.hubspot.com/node/15261/html
; Title: SEC Filing | HubSpot
- Source: https://www.mediapost.com/publications/article/365746/hubspot-ceo-halligan-steps-down-after-15-years-re.html
; Title: HubSpot CEO Halligan Steps Down After 15 Years ...
- Source: https://simplywall.st/stocks/gb/software/lse-0tzz/hubspot-shares/management
; Title: HubSpot, Inc. (OTZZ) Leadership & Management Team ...
- Source: https://www.stockinsights.ai/us/HUBS/8-K/management-changes-20260401-cab
; Title: HUBS HUBSPOT INC Management Changes 8-K Filing
- Source: https://ir.hubspot.com/node/15261/ixbrl-viewer
; Title: 0001193125-26-138021 | 8-K | iXBRL Viewer | Hubspot INC
; Full crawled content: 0001193125-26-138021 | 8-K | iXBRL Viewer | Hubspot INC
```

Inline Viewer

```
* Menu Menu
  Information Help
* Sections Sections
* Additional Search Options
  Include Fact Name
  Include Fact Content
  Include Labels
  Include Definitions
  Include Dimensions
Reference Options
  Include Topic
  Include Sub-Topic
  Include Paragraph
  Include Publisher
  Include Section
  Match Case
```

```

    Clear Search Submit Search
* Data Data
    All
    Amounts Only
    Text Only
    Calculations Only
    Negatives Only
    Additional Items Only
* Tags Tags
    All
    Standard Only
    Custom Only
* More Filters More Filters

## Loading Inline Form.

##### Getting Started
The Inline XBRL Viewer allows a user to quickly and easily review details of the tagged information in
an Inline document by automatically placing a top and bottom highlight border around each tagged numeric
fact and left and right border for each block tagged fact. Hovering over a tagged fact will highlight (
shade) all content related to the tagged fact, and clicking on a tagged fact will reveal its tagging
details in the Fact Review Window. Search and filter options are also provided to easily refine and
identify specific types of tagged information.

##### Fact Review Window
The Fact Review Window shows the tagging details for the currently selected fact, which is highlighted
with a solid blue background. There are four categories of fact detail which can be viewed; an "N/A"
value indicates there is no available information for the item within the given category:
* Attributes - All primary information (as applicable) describing the tagged fact including period, sign
, decimals, dimensional detail (axes and members), scale, measure, data type and footnotes
* Labels - Detailed documentation (definition) for the tag used, and other labels
* References - Authoritative reference information (as applicable) for the selected tag
* Calculation - Balance and calculation hierarchy details (numeric items only)

[Inline XBRL viewer help text continues: Search, Filter (Data/Tags/More Filters), Facts Results List,
Information, Tagged Sections, Save XBRL Instance, Save XBRL Zip, and Settings sections -- all generic
viewer UI documentation, no company-officer content.]
- Source: https://www.cfo.com/news/cfos-on-the-move-dec-5-Jessica-Ross-GitLab-Bill-Koefoed-OneStream-
Mike-Lenihan-Texas-Roadhouse/807088/
; Title: CFOs On the Move: Week ending Dec. 5
- Source: https://libraryofminds.com/brian-halligan
; Title: Brian Halligan - Episode 8
- Source: https://www.bizjournals.com/boston/news/2021/08/04/hubspot-appoints-yamini-rangan-as-ceo.html
; Title: Yamini Rangan to succeed Brian Halligan as HubSpot CEO
- Source: https://www.linkedin.com/pulse/undisrupted-hubspots-reflections-disrupted-dan-lyons-dharmesh-
shah
; Title: Undisrupted: HubSpot's Reflections on "Disrupted"
---
```

H Prompts

H.1 Base Prompts

The base prompts for the three LLM steps: the defaults used by ask-web at the query-expansion and URL-ranking stages.

Listing 17: Query-expansion prompt (system prompt).

```

Today's date and time is The current date is {current_date}, and the current time
is {current_time}. The current timezone is UTC. The user will ask questions in this
timezone. When passing date or time arguments to tools, ALWAYS use ISO 8601 with an
explicit UTC offset (e.g. {current_datetime_iso}). Tools return times in 24-hour
UTC format; you must convert them back to UTC when responding to the user.

You are a search query expansion agent researching {company}. Produce exactly 5
focused, one-sentence web search queries that together thoroughly answer the user's
research question.

**Information about the company:**
Company: {company}
Rules:
```


- Output exactly 5 queries, with no duplicates or minor rewordings.
- Each query should attack a DIFFERENT angle or sub-topic of the question, so the set as a whole covers it. Diversify by WHAT you ask for, not by rewording the same thing, and not by naming a kind of source.
- Build each query from the core entities plus the specific thing the question is about -- the event, attribute, relationship, or claim -- naming the actors and what links them; that is what a useful page will actually mention.
- Be specific. Avoid words so broad they pull unrelated or wrong-subject pages; prefer the precise term over the category. Phrase each query so the entity plays the SAME role it has in the question (e.g. who did something vs. who it was done to). If the question involves several entities, spread them across queries instead of naming all of them in each.
- Do not tie a query to one kind of document or source (a filing, a press release, a single site) unless the question explicitly asks for that source; let the search engine return whatever page type actually carries the answer.
- Use Google operators only: "exact phrase", OR (or |) with parentheses for groups, and AND (a space also implies AND). Keep quotes and parentheses balanced. Do NOT use after:/before: date filters, and use 'site:' only when you are sure a domain has the answer. Use any URLs named in the question.
- Timeframe (SOFT emphasis, never a hard filter): if the question targets a period, infer the era it covers from its actual start and end -- the calendar year(s), and optionally the quarter or half -- remembering a period can span more than one year and may lie well in the past, not just recently. Add that era as a PLAIN keyword on only SOME queries to nudge results toward it; do not quote it, do not put it on every query, and keep at least one or two queries with no time term at all. Never pin a query to a single exact date.
- Why soft and not forced: a page's PUBLISH date is not the date of the thing it describes -- sources often appear weeks or months later, and engines already favor recent pages. A hard date filter can drop the very page that proves the event, and forcing the year into every query drops pages that never restate it. Treat the period as a light nudge on a few queries, keep retrieval broad, and let the downstream reader confirm exact dates from the page.

Listing 18: Query-expansion prompt (user prompt).

```
Research question: {prompt}

Company: {company}
```

Listing 19: Ranking prompt (system prompt): selects the best URLs to crawl and gives feedback on the generated queries.

```
You are an expert research assistant researching {company}. You are designed to
select the best web pages to crawl in order to answer a user's research question.

**Information about the company:**
Company: {company}
You will be given:
- The user's original research question
- A set of web search results, each with its search query, followed by results that
  contain URL, title, and description

You have 2 tasks:

## Task 1 - Select URLs to crawl
- Identify and select up to 5 URLs from the search results which are the most likely
  to contain information that would best help answer the user's research question.
- If fewer than 5 URLs stand out as relevant, you may select fewer than 5 (but never
  more than 5).
- The selected URLs should be distinct and together maximize coverage of the
  question or its key subtopics.
- Only select URLs that, based on their titles and descriptions, genuinely appear
  highly informative and central to the question.
- Ignore URLs that are duplicates or do not appear directly relevant.
- Do NOT invent new URLs -- only choose from those presented.
- If no results are sufficiently relevant, set the field 'urls' to an empty array:
  "urls": [].

## Task 2 - Give feedback on query generation
- Given the user's original research question and the queries generated to find
  answers for it, analyse the performance of the generated queries and give feedback
```

```

on it.
- The search results have been generated after considering the user's timelines so
  you should not include advice about filtering by date.
- Focus on what sources might lead to an answer for the user's original research
  question. If results are too vague, suggest to use more detailed queries. If the
  searches returned very less results, suggest making the queries less restrictive.
- Don't output exact queries, just give keywords to use in the queries, website
  names, and general directions.

Output a JSON object where:
- the field 'urls' is an array of strings, each string being one selected URL (full
  URL only, no commentary). If none are selected, use an empty array `[]`.
- the field 'query_expansion_feedback' contains a feedback message about the query
  processing step.

```

Listing 20: Ranking prompt (user prompt). Results are grouped under their search query; each result lists Source, Title, and Content (the snippet). Groups are separated by —. Abbreviated for space.

```

Original question: {prompt}

Company: {company}

Web search results:
Search query: "Honeywell" ("new AI" OR "AI product" OR "AI solution") site:sec.gov

- Source: https://www.honeywell.com/us/en
; Title: Honeywell - The Future Is What We Make It
; Content: Honeywell Forge is a future-ready IOT platform delivering AI-enabled
applications and services for intelligent, efficient industrial operations...Read more
- Source: https://en.wikipedia.org/wiki/Honeywell
; Title: Honeywell
; Content: Honeywell International Inc. is an American publicly traded, multinational
conglomerate headquartered in Charlotte, North Carolina...Read more
---
Search query: "Honeywell" ("AI" OR "artificial intelligence") ("earnings call" OR "transcript")

- Source: https://investor.honeywell.com/events/event-details/q1-2026-honeywell-earnings-conference-call
; Title: Q1 2026 Honeywell Earnings Conference Call
; Content: Q1 2026 Honeywell Earnings Conference Call. Apr 23, 2026 8:30 AM EDT...Read more
---

```

Listing 21: Synthesis prompt (system prompt).

```

You are a research assistant researching {company}. Based on the web search results
and crawled page content provided, answer the user's question.

Respond with exactly one of the following: True, False. Your response must contain
only that value and nothing else.

```

Listing 22: Synthesis prompt (user prompt). Same grouped format as the ranking prompt, but the URLs selected for crawling additionally carry their page text under "Full crawled content"; non-selected results show only Source and Title. Abbreviated for space.

```

Question: {prompt}

Company: {company}

Web search results:
Search query: "Honeywell" ("new AI" OR "AI product" OR "AI solution") site:sec.gov

- Source: https://www.honeywell.com/us/en
; Title: Honeywell - The Future Is What We Make It
- Source: https://en.wikipedia.org/wiki/Honeywell
; Title: Honeywell
---
Search query: "Honeywell" ("AI" OR "artificial intelligence") ("earnings call" OR "transcript")

- Source: https://investor.honeywell.com/events/event-details/q1-2026-honeywell-earnings-conference-call
; Title: Q1 2026 Honeywell Earnings Conference Call

```

```

; Full crawled content: Q1 2026 Honeywell Earnings Conference Call | Honeywell
International Inc. Apr 23, 2026 8:30 AM EDT. Supporting Materials: Honeywell 1Q26
Earnings Release... [full page text]
- Source: https://finance.yahoo.com/quote/HON/
; Title: Honeywell International Inc. (HON) Stock Price, News ...
---
```

H.2 Query-Expansion Prompt Variants

Alternative query-expansion system prompts, each targeting the failure modes of a family of expander models. All variants share the **base** date/time preamble and user prompt.

Listing 23: Query-expansion prompt variant **oss** (system prompt), written for open-weight expanders that over-quote and under-diversify. Date/time preamble and user prompt are identical to Listings 17 and 18.

```

You are a search query expansion agent researching {company}. Produce exactly 5
focused, one-sentence web search queries that together thoroughly answer the user's
research question.
```

```

**Information about the company:**
```

```
Company: {company}
```

```
Rules:
```

- Output exactly 5 queries, with no duplicates or minor rewordings.
- Each query should attack a DIFFERENT angle or sub-topic of the question, so the set as a whole covers it. Diversify by WHAT you ask for, not by rewording the same thing, and not by naming a kind of source.
- Build each query from the core entities plus the specific thing the question is about -- the event, attribute, relationship, or claim -- naming the actors and what links them; that is what a useful page will actually mention.
- The same fact is reported in many different words, so never commit a query to a single wording of the event or claim: in most queries, wrap 2-4 alternative phrasings of it in one parenthesized OR group, and vary the alternatives from query to query instead of repeating one group everywhere. Quote every multi-word alternative inside the group -- unquoted, the engine reads it as separate words and the group falls apart; single-word alternatives stay bare.
- Outside OR groups, quote only proper nouns (the company, a person, a product). Quotes force pages to contain those exact words, so never chain several quoted phrases together -- a page would have to contain all of them verbatim -- and never quote a whole description of the event or claim.
- Do not tie a query to one kind of document or source (a filing, a press release, a single site) unless the question explicitly asks for that source; let the search engine return whatever page type actually carries the answer.
- Timeframe (SOFT emphasis, never a hard filter): if the question targets a period, add its era (the calendar year, optionally the quarter) as a PLAIN unquoted keyword on only SOME queries, keeping at least one or two queries with no time term at all. Never quote a year, month, or date, and never pin a query to a single exact date: a page's PUBLISH date is not the date of the thing it describes, so a hard date term can drop the very page that proves the event.
- Use Google operators only: "exact phrase", OR (or |) with parentheses for groups, and AND (a space also implies AND). Keep quotes and parentheses balanced. Do NOT use after:/before: date filters, and use 'site:' only when you are sure a domain has the answer. Use any URLs named in the question.
- Output the bare query text in plain ASCII: straight quotes, ordinary hyphens and spaces, and no trailing period.

Listing 24: Query-expansion prompt variant **oss-template** (system prompt): leads with a generic query shape instead of prose quoting rules. Date/time preamble and user prompt as in the **base** prompt.

```

You are a search query expansion agent researching {company}. Produce exactly 5
focused web search queries that together thoroughly answer the user's research
question.
```

```

**Information about the company:**
```

```
Company: {company}
```

```
Most queries should follow this shape:
```

```

"<entity>" ("<multi-word phrasing A>" OR "<multi-word phrasing B>" OR
<single-word C>) <era>
```

where <entity> is the company or another core actor from the question, the OR group holds 2-4 DIFFERENT wordings of the event, attribute, or claim the question asks about (multi-word wordings quoted, single words bare), and <era> is an optional PLAIN year or quarter keyword.

Rules:

- Output exactly 5 queries, with no duplicates or minor rewordings.
- Each query should attack a DIFFERENT angle or sub-topic of the question, with its own OR alternatives -- never repeat one OR group across queries.
- Quote nothing except the entity and the multi-word phrasings inside the OR group; never quote a year, month, or date, and never quote a whole description of the event or claim.
- Put the era keyword on only SOME queries, keep one or two queries with no time term at all, and never pin a query to an exact date: a page's PUBLISH date is not the date of the thing it describes.
- Do not tie a query to one kind of document or source (a filing, a press release, a single site) unless the question explicitly asks for that source.
- Use Google operators only (quotes, OR, parentheses; a space implies AND). No after:/before: date filters; 'site:' only when you are sure a domain has the answer. Output plain ASCII with no trailing period.

Listing 25: Query-expansion prompt variant oss-concise (system prompt): the oss instructions compressed to six short rules. Date/time preamble and user prompt as in the base prompt.

You are a search query expansion agent researching {company}, writing web search queries the way an expert human searcher would.

****Information about the company:****

Company: {company}

Rules:

- Output exactly 5 queries, each attacking a DIFFERENT angle or sub-topic of the question.
- In most queries, wrap 2-4 alternative wordings of the key event, attribute, or claim in one parenthesized OR group; quote the multi-word wordings, leave single words bare, and vary the alternatives across queries.
- Outside OR groups, quote only proper nouns; never quote a whole event description and never quote a year, month, or date.
- Add the question's era as a plain keyword on SOME queries only; keep one or two queries time-free; never pin an exact date.
- Do not restrict to one source or document type unless the question asks for it. No after:/before: filters; 'site:' only when certain.
- Output plain ASCII with no trailing period.

Listing 26: Query-expansion prompt variant oss-concise2 (system prompt): oss-concise plus concreteness rules (concrete wordings, no coverage words, no hyphenated year ranges). Date/time preamble and user prompt as in the base prompt.

You are a search query expansion agent researching {company}, writing web search queries the way an expert human searcher would.

****Information about the company:****

Company: {company}

Rules:

- Output exactly 5 queries, each attacking a DIFFERENT angle or sub-topic of the question -- vary WHAT you ask for (the event itself, the parties involved, the filing or approval it would trigger), not just verb synonyms for the same thing.
- Use concrete wordings a real page would contain -- how an article or filing would actually describe the event -- never the question's own category labels or abstract phrases.
- In most queries, wrap 2-4 alternative wordings of the key event, attribute, or claim in one parenthesized OR group; quote the multi-word wordings, leave single words bare, and vary the alternatives across queries.
- Outside OR groups, quote only proper nouns; never quote a whole event description and never quote a year, month, or date.
- Never pad a query with a bare coverage word such as news, report, media, or announcement -- those match every page; name the specific event instead.
- Add the question's era as a plain keyword on SOME queries only; write years as separate keywords (2025 2026), never a hyphenated range (the hyphen excludes); keep one or two queries time-free; never pin an exact date.
- Do not restrict to one source or document type unless the question asks for it.

No after:/before: filters; 'site:' only when certain. Output plain ASCII with no trailing period.

Listing 27: Query-expansion prompt variant `qwen` (system prompt), aimed at expanders that under-structure (few OR groups, little quoting). Date/time preamble and user prompt as in the `base` prompt.

```
You are a search query expansion agent researching {company}, writing web search queries the way an expert human searcher would.

**Information about the company:**
Company: {company}
Rules:
- Output exactly 5 queries, each attacking a DIFFERENT angle or sub-topic of the question.
- In most queries, wrap 2-4 alternative wordings of the key event, attribute, or claim in one parenthesized OR group; quote each multi-word wording (unquoted it searches as loose words), leave single words bare, and vary the alternatives across queries.
- Write the OR operator in capitals; lowercase 'or' is read as an ordinary search word, and connecting alternatives with it loses them.
- Use concrete wordings a real page would contain -- how the event or claim would actually be written -- never the question's abstract category words or placeholder terms for things it leaves unnamed.
- Outside OR groups, quote only proper nouns; never quote a whole event description and never quote a year, month, or date. Add the question's era as a plain keyword on SOME queries only; keep one or two queries time-free; never pin an exact date.
- Do not restrict to one source or document type unless the question asks for it.
No after:/before: filters; 'site:' only when certain. Output plain ASCII with no trailing period.
```

Listing 28: Query-expansion prompt variant `qwen2` (system prompt): `qwen` with angle diversity spelled out, 3-4 OR alternatives preferred, and a coverage-word ban. Date/time preamble and user prompt as in the `base` prompt.

```
You are a search query expansion agent researching {company}, writing web search queries the way an expert human searcher would.

**Information about the company:**
Company: {company}
Rules:
- Output exactly 5 queries, each attacking a DIFFERENT angle or sub-topic of the question -- vary WHAT you ask for (the event itself, the parties involved, the filing or approval it would trigger), not just verb synonyms for the same thing.
- In most queries, wrap 3-4 alternative wordings of the key event, attribute, or claim in one parenthesized OR group; quote each multi-word wording (unquoted it searches as loose words), leave single words bare, and vary the alternatives across queries.
- Write the OR operator in capitals; lowercase 'or' is read as an ordinary search word, and connecting alternatives with it loses them.
- Use concrete wordings a real page would contain -- how the event or claim would actually be written -- never the question's abstract category words or placeholder terms, and never inside quotes: a quoted abstract label must appear verbatim, and no page says it.
- Never pad a query with a bare coverage word such as news, report, media, or announcement -- those match every page; name the specific event instead.
- Outside OR groups, quote only proper nouns; never quote a whole event description and never quote a year, month, or date. Add the question's era as a plain keyword on SOME queries only; keep one or two queries time-free; never pin an exact date.
- Do not restrict to one source or document type unless the question asks for it.
No after:/before: filters; 'site:' only when certain. Output plain ASCII with no trailing period.
```

H.3 Synthesis Prompt Variants

Alternative synthesis system prompts for the tolerant plain-text output path, which lets the model reason before the final-line verdict. All variants share the `base` user prompt.

Listing 29: Synthesis prompt variant `oss` (system prompt), for open-weight synthesizers on the tolerant plain-text path: asks for a brief weighing of the evidence and pins the verdict to the final line, with an explicit no-evidence-means-False decision rule. User prompt identical to Listing 22.

```
You are a research assistant researching {company}. Based only on the web search
results and crawled page content provided, answer the user's question.

First weigh the evidence: state briefly what the sources do and do not show about
the claim in the question. Then give your verdict: True only if the evidence
affirmatively shows the claim; False if the evidence is absent, insufficient, or
does not address it.

End your reply with a single line containing exactly one word: True or False.
```

Listing 30: Synthesis prompt variant `oss-minimal` (system prompt): keeps the `base` framing and only moves the verdict to the final line, giving reasoning a legal place. User prompt identical to Listing 22.

```
You are a research assistant researching {company}. Based on the web search
results and crawled page content provided, answer the user's question.

You may briefly reason first. End your reply with a single line containing
exactly one word: True or False.
```

Listing 31: Synthesis prompt variant `oss-minimal-notools` (system prompt): `oss-minimal` plus an explicit no-tools paragraph. Some open-weight models are trained with a built-in browsing tool and role-play search calls instead of answering; the added paragraph denies tool access. User prompt identical to Listing 22.

```
You are a research assistant researching {company}. Based on the web search
results and crawled page content provided, answer the user's question.

You do not have access to any tools, search, or the internet. All available
evidence is already included below; do not attempt to search or open pages.

You may briefly reason first. End your reply with a single line containing
exactly one word: True or False.
```