MBA AI Engine Perception Index

Methodology · ChatGPT release

Publication revision 1.2.2 | September 2026 | Arcalea

ARCALÉA

arcalea.com/aeo-index/mba-aepi

MBA-AEPI | Scope clarification

September 14, 2026 | Supplement to publication revision 1.2.2

Full-time MBA focus

This release examines 25 tracked U.S. business schools using questions focused on full-time MBA education. Of 600 prompts, 498 explicitly specify full-time study; 102 use broader MBA or business-school wording. Full-time includes one-year and two-year options. Scoring identifies school mentions without independently verifying the MBA format discussed in each answer.

Conditions for interpreting the results

Results reflect this cohort and question mix. Studies targeting other MBA formats, specializations, applicant profiles, geographies, or school cohorts could produce different mention rates, rankings, and configuration differences. Those alternative outcomes have not been measured here. Full-time focus is a study condition; results should not be generalized to part-time, executive, or online MBAs. The 540-prompt sensitivity check addresses exclusion of named-school factual prompts, not robustness across MBA formats.

This clarification changes the description of the study scope, not its raw responses, scoring, numerical results, or statistical classifications. Original page numbering is retained on the following pages.

ChatGPT-only release covering the no-tool, optional-search, and required-search conditions. The filename retains its earlier version for link compatibility; the revision history below identifies the current publication text.

Version History

Version	Date	Changes
1.0.0	September 2026	Initial release
1.1.0	September 2026	API surface clarification; search invocation disclosure; model versioning note; ECM renamed to CSC; CI scope clarification; entity detection validation disclosure; valence interpretation correction; Stable language tightened
1.2.0	September 2026	Entity detection pattern corrections (MIT Sloan, Kenan-Flagler); corrected M1-AGSO scores; per-stratum citation table; exact McNemar + Benjamini-Hochberg replaces divergence test; rank-on-rank correlation added; score-space misspecification documented
1.2.1	September 2026	ChatGPT-only publication review; appendix regenerated from CSV; BH labels distinguished from legacy Option D and materiality; reused baseline and collection dates clarified
1.2.2	September 2026	Recovered original raw captures, registry, and collection scripts; reconstructed both paired comparisons; corrected the entity-free claim (60 named-school factual prompts); documented request-setting and logging limitations. Numerical results unchanged.

1. Overview

The MBA AI Engine Perception Index (MBA-AEPI) measures how often programs appear in generated answers under specified API configurations. It compares a no-tool baseline (NGMO, also called TDK) with optional search (AGSO) and required search (AGSO_FORCED). These outputs do not directly measure training-data contents or program quality.

This release covers 25 target MBA programs and 600 prompt IDs. The September 2 collection comprises 1,200 responses under two configurations; the required-search condition adds 600 responses and reuses the earlier no-tool baseline, for 1,800 distinct captures across the two collections. Other-engine results are outside this publication.

> **DISCLOSURE**
>
> Arcalea offers AEO consulting services. The study uses the v1.9.0-FROZEN design and Amendment v1.9.1. The publication revision reconciles descriptions and labels with the saved scoring outputs; it does not change raw captures, mention counts, or test statistics. BH significance and practical materiality are reported separately (Section 6.4).

2. Index Architecture

2.1 Dual-Mode Capture Protocol

In the original design, each prompt was executed once in each of two configurations.

TDK Mode (Training Data Knowledge / Non-Grounded Model Output, NGMO). The model receives the prompt with no access to live web search. The response reflects baseline behavior without retrieval augmentation. TDK is a research construct, not a user-facing product state; "what the model knows" is colloquial shorthand and does not imply direct access to training-data representations.

LSS Mode (Live Search Surface / Answer-Generation Search Output, AGSO). The model receives the same prompt with search tools enabled. In the optional-search condition, invocation was optional and was not logged directly. The required-search condition set an invocation flag in every response. Citation

presence and invocation are distinct variables; AGSO should not automatically be interpreted as confirmed retrieval.

ΔM1 is the difference in mention rates between a search configuration and the no-tool baseline. It is a measured configuration comparison; attributing the difference specifically to retrieval requires further assumptions and controls.

2.2 Temporal Synchronization (CSPS Protocol)

CSPS (Collect Simultaneously, Publish Sequentially) is the intended design for minimizing temporal differences between comparisons. The required-search condition is an exception: its captures are timestamped September 5, 2026 UTC, while the reused no-tool baseline was collected September 2. Prompt pairing does not remove the possible influence of that timing gap. Future dual-mode collections should run both conditions close together and record an exact model snapshot.

3. Prompt Registry

3.1 Design Principles

> **CORRECTION (revision 1.2.2): THE REGISTRY IS NOT ENTIRELY ENTITY-FREE**
>
> The recovered registry contains 600 prompts. All 60 Factual prompts (MBA-PROD-0541 through MBA-PROD-0600) explicitly name schools. They repeat 20 school-specific questions in three variants: 19 schools belong to the scored cohort and Emory Goizueta does not, so 57 prompts directly name tracked schools and six tracked programs receive no direct factual question. The other 540 prompts cover Generic, Constraints, and Persona scenarios; an automated name-token scan found no target-name matches in those prompts. The prior claim that every prompt was entity-free was incorrect. Aggregate M1 is therefore a mixed discovery/lookup index: it combines discovery-style output with named-school lookup responses and the 540-prompt sensitivity analysis isolates the discovery-style portion. No scores were changed to conceal this design limitation.

3.2 Stratification

Stratum	Description	Share	AGSO responses
Generic	Open-ended program discovery (“best full-time MBA programs”)	40%	240
Constraints	Filtering criteria (location, cost, specialization, format)	30%	180
Persona	Role-framed (career switcher, international, STEM applicant)	20%	120
Factual	Named-school data lookups (rankings, salary, acceptance rate, class size)	10%	60

Stratification weights reflect research-team judgment about relative prevalence of query types; they are a design choice, not an empirical claim about observed query-log distributions. A post-hoc sensitivity analysis excluding the Factual stratum is reported in the companion document.

3.3 Registry Integrity

The included registry contains 600 unique prompt IDs and SHA-256 hashes. All 600 hashes reproduce from UTF-8 prompt text, and IDs and strata align with both raw collections. The recovered collection scripts do not validate those hashes at request time; present-day hash consistency is not proof that capture-time verification occurred. See SOURCE_AUDIT.md for provenance and limits.

4. Data Collection

4.1 Engine and Model Specification

Parameter	Value
Engine	ChatGPT (OpenAI Responses API)
Model identifier recorded	gpt-4o (alias; snapshot not recorded)
TDK configuration	Responses API without tools in the recovered dual-mode script
Original output limit	Script default max_output_tokens=1500, CLI-overridable; actual invocation not archived
Forced-search output limit	Not specified in the recovered script
LSS configuration	Responses API with web_search_preview enabled; required search adds tool_choice="required"
No-tool / optional-search timestamps	September 2, 2026, 16:59:24–18:49:55 UTC
Required-search timestamps	September 5, 2026, 01:14:52–02:16:52 UTC
Captures	1,200 original (600 NGMO + 600 AGSO); 600 AGSO_FORCED; baseline reused

API surface note. Both recovered scripts call client.responses.create. These are source-code settings, not a complete archive of historical API requests; original CLI overrides and SDK versions are not recorded, and model_version_reported stores the requested model argument, not an observed server snapshot. Findings describe GPT-4o API outputs under these configurations and are labeled accordingly.

4.2 Execution Protocol

The recovered scripts implement retries and record per-response timestamps, status codes, attempt counts, and response text. They write the completed batch at the end (CSV; the forced script also writes JSONL). They do not implement the previously described every-ten-prompts resumable checkpoint or capture-time hash verification. Their output-token settings differ (original default 1500; forced script no explicit cap). This difference and the collection-date gap limit causal attribution to retrieval.

4.3 Search Invocation in the AGSO Condition

Enabling a search tool does not guarantee the model invoked it. The optional-search condition recorded no invocation field; the only proxy is citation presence, which varies sharply by stratum:

Stratum	AGSO responses	With citations	Citation rate	NGMO citations
Factual	60	50	83.3%	0 / 60
Constraints	180	19	10.6%	0 / 180
Generic	240	14	5.8%	0 / 240
Persona	120	0	0.0%	0 / 120
Total	600	83	13.8%	0 / 600

Interpretation. Citation-confirmed retrieval was most common in Factual responses, which are named-school lookups. Zero citations do not imply zero mention-rate differences: Stanford's Persona rates were 0.800 AGSO and 0.617 NGMO, a difference of +0.183. The answers are separate generations paired by prompt ID. Without invocation logs, the role of retrieval cannot be isolated; more paired observations improve precision but cannot recover missing invocation evidence. The required-search condition used tool_choice="required": its invocation flag is set on 600/600 responses, but the flag is set by either a web_search_call item or a citation annotation and is not a retained raw tool-call trace (17 flagged responses have no citations).

4.4 Data Quality Attestation (September 2 collection)

The recovered original captures confirm 1,200 rows, all recorded status 200, all first-attempt, minimum response length 207 characters, and complete prompt/mode pairing. Status codes and model identifiers are harness-recorded fields, not an independent transport audit.

Check	Result
Capture completeness	1,200 / 1,200 (100%)
HTTP 200 rate (recorded)	100%
Model identifier consistency	100% gpt-4o
First-attempt success (recorded)	100%
Minimum response length	207 characters
Empty responses	0
Truncated responses	Not independently established; completion-status metadata was not retained
Identical TDK/LSS pairs	0 / 600
Entity universe coverage	25 / 25 programs detected

5. Entity Universe

The index tracks 25 target MBA programs, each with a canonical name, abbreviations, and parent-university references (Appendix A).

6. Scoring Methodology

6.1 Entity Detection

A regex-based pattern matcher identifies mentions per response, scored as a binary indicator (mentioned at least once, or not).

> **PATTERN CORRECTIONS (v1.2.0)**
>
> **MIT Sloan:** the pattern `MIT.*(?:MBA|business|management)` lacked a word boundary and matched “submit,” “commit,” and “admitted”; 37 of 716 detections (5.2%) referenced neither MIT nor Sloan. Corrected to `\bMIT\b`: M1-AGSO 0.597 → 0.558, rank 4 → 6. **Kenan-Flagler:** the pattern matched “uncommon” and “unconnected”; corrected to `\bUNC\b`, affecting 11 of 47 detections: 0.045 → 0.032, rank unchanged.

The remaining 23 patterns produced no detections lacking a canonical token. Short or ambiguous names (Ross, Johnson, Smith, Foster, Marshall, Anderson) require co-occurring context markers whose effectiveness is untested against a validation sample. No human-coded validation sample has been run; precision and recall are unknown. Future collection should include human-coded validation and quantified error rates.

6.2 M1 (Mention Rate)

M1-AGSO is computed over the 600 search-enabled responses of a given condition: the optional-search ranking and the forced-search ranking are computed and reported separately. Neither should be labeled simply the ChatGPT rank without specifying the condition. M1-NGMO is computed over the 600 no-tool responses. Both are bounded [0,1]. M1 captures whether a program appeared at least once;

recommendation strength, prominence, citation support, attribute association, share of answer, and enrollment influence are separate dimensions reserved for later instruments.

6.3 Confidence Intervals

All M1 estimates carry Wilson score 95% intervals (z = 1.96), preferred over Wald for coverage near 0 or 1. They capture binomial sampling uncertainty only, not prompt-selection uncertainty, repeated-run variance, day-to-day AGSO variation, model drift, or inter-prompt correlation, and should not be read as predicting independent replications.

6.4 ΔM1, statistical classification, and materiality

ΔM1 = M1-AGSO minus M1-NGMO. The original comparison sets optional search against the no-tool baseline; the second compares required search on September 5 UTC with the reused September 2 baseline. A positive value means more mentions in the search configuration; attributing it to retrieval alone requires a concurrent baseline with matched request settings, model-version controls, and retained invocation evidence.

Classification	Criteria (primary rule)
Search-Amplified	BH q ≤ 0.05 and ΔM1 > 0
Stable	No BH-significant difference
Search-Suppressed	BH q ≤ 0.05 and ΔM1 < 0

Primary statistical rule. A two-sided exact McNemar test on discordant prompt pairs, then Benjamini-Hochberg correction across all 25 programs at α = 0.05. Stable means the threshold was not met; it is not proof of equivalence. **Legacy Option D** (non-overlapping Wilson intervals and a difference of at least five percentage points) is retained in the forced-search master CSV as legacy_option_d_class; delta_M1_class matches the BH label. **Materiality is separate:** Amendment v1.9.1 describes a paired-difference interval assessment that the current pipeline does not supply, so no materiality claim is made. The 13 / 3 / 9 result is a statistical classification; the significant increases for Carey (+0.030), Foster (+0.028), and Smith (+0.012) are below five percentage points even as point estimates.

Optional-search results. Five programs reached nominal p < 0.05, all in the suppression direction; none survived BH correction. All 25 labels are Stable. Equivalence is a separate question this test leaves open.

Entity	Off-only (b)	On-only (c)	Discordant	ΔM1	McNemar p	BH q
McCombs	24	7	31	−0.028	0.0033	0.0505
Ross	55	28	83	−0.045	0.0040	0.0505
Tuck	64	37	101	−0.045	0.0093	0.0664
Johnson	18	5	23	−0.022	0.0106	0.0664
Wharton	26	11	37	−0.025	0.0201	0.1004

Required-search results. Thirteen programs were BH-significantly lower and three higher relative to the earlier baseline; nine were Stable. Wharton, Tuck, and Ross were significant in the suppression direction in both the nominal optional-search test and the BH-adjusted required-search test; McCombs and Johnson were Stable under required search. The largest negative difference was Columbia (−0.230). Timing, request-setting, and model-version limitations remain even with the invocation flag set.

Stratum interpretation. BH labels are program-level tests over the full 600-prompt set, not per-stratum tests. Low or absent citation evidence does not force ΔM1 to zero; it limits attribution to retrieval.

6.5 CSC (Cross-Stratum Consistency)

CSC = 1 − (σ / μ) of the four stratum-level M1-AGSO values, clipped to [0,1]; undefined if μ = 0. A diagnostic metric without published precedent. Note that the Factual stratum consists of named-school lookups.

6.6 Valence Scoring

Keyword-pattern matching in a 200-character window around the first mention. The detector found zero negative-framing matches; this is a directional finding, not a validated behavioral characterization. Future semantic scoring should be validated against human-coded examples.

6.7 Rank Correlation with External Rankings

Primary specification. Rank-on-rank OLS of the original optional-search AI rank on the recorded U.S. News rank; descriptive, not causal, and not recomputed for forced search. **Retained score-space fit** models the bounded mention proportion, predicts negative values from U.S. News rank 22 downward (Mendoza, Smith, Olin, Carey), and its residuals depend on that functional form.

Metric	Value
Spearman rho (rank-on-rank)	0.973
Rank-on-rank OLS R^2	0.944
Unexplained rank variance	5.6%
Score-space Pearson r (reference)	−0.941
Score-space OLS R^2 (reference)	0.886
Score-space unexplained	11.4%

The 5.6% and 11.4% figures describe different outcome scales and are not interchangeable; neither identifies an independent AEO effect. A positive rank residual means a numerically worse AI rank than predicted (under-indexed); negative means better (over-indexed). These are positions relative to the fitted relationship, not evidence about merit or cause.

7. Publication Scope

This release covers ChatGPT only, under three configurations. The article, this methodology, the ChatGPT result tables, the required-search run log, the source audit, and the sensitivity report form the publication narrative. Other-engine data retained elsewhere in the working repository are outside this release.

8. Interpretive Framework

What M1 measures. M1 measures AI visibility: the rate at which the engine surfaces a program in response to the 600 prompts in the registry, including 60 named-school factual prompts. It is reported alongside quality rankings such as U.S. News, which measure something different. **AEO and SEO.** AEO concerns representation in generated answers; SEO concerns ranking in a results list. The MBA-AEPI measures the AEO outcome exclusively; no SEO metrics are inputs. **Diagnostic language.** Programs are described as over- or under-indexed relative to the recorded U.S. News rank and fitted relationship, not as winners or losers.

9. Limitations

Measurement surface. All data are GPT-4o Responses API outputs under specified tool configurations; findings apply to those configurations and are labeled accordingly.

Search invocation unverified for most optional-search responses. No invocation field was recorded; citations confirm retrieval in 83 of 600 (13.8%). Persona ΔM1 is not zero (e.g., Stanford +0.183) but cannot be attributed to live search.

Model alias, no snapshot. gpt-4o is an alias; no snapshot ID was recorded. Replication may encounter a different model version.

Single engine. Cross-engine patterns and stability should be interpreted with caution until multi-engine data are available.

Entity detection unvalidated. No human-coded validation sample; two pattern defects were corrected post-collection; precision and recall remain unknown.

Confidence interval scope. Wilson CIs capture binomial uncertainty only.

Optional-search nonsignificance. Five nominal candidates, none surviving BH; equivalence remains a separate, open question.

Rank correlation scope. Rank-space and score-space fits describe different outcomes; neither identifies an independent AEO signal; forced-search correlations are not reported.

Valence is a keyword proxy. Zero negative matches mean the detector found none, not that framing is uniformly positive.

Different collection dates, settings, and reused baseline. No-tool and optional-search captures span September 2; forced-search captures span September 5 UTC. No concurrent NGMO arm. The original script defaults to max_output_tokens=1500; the forced script supplies no cap. The comparison does not fully isolate retrieval as a cause.

Prompt stratification and named entities. Weights reflect research-team judgment. All 60 Factual prompts name schools with unequal coverage (19 tracked programs × 3; six with none; three about an untracked school). The full-corpus index is not an entity-free discovery measure; the companion document reports a separately labeled 540-prompt sensitivity analysis.

No adjustment for web footprint. Web presence may be associated with both established rankings and AI mentions; measuring and controlling for it, inspecting training data, and testing interventions are follow-on work.

10. Reproducibility and Available Files

Artifact	Availability and scope
Original raw captures (no-tool + optional search)	Included: 1,200 responses (600 NGMO, 600 AGSO), recovered and fingerprinted
Standalone registry	Included: 600 prompt texts, IDs, strata, hashes; all hashes match
Collection scripts	Included unchanged as historical sources; not executed during verification
Required-search raw captures	Included: 600 responses; identical to the previously checked-in copy
Scoring code	Included; both corrected scripts reproduce every checked-in output
Result tables	Included: rates, intervals, strata, citation summaries, paired counts, p/q values

Artifact	Availability and scope
Verification script	verify_chatgpt_release.py reconstructs both raw pair matrices and checks statistics, registry integrity, and appendix transcription
Source fingerprints	chatgpt_source_manifest.json records SHA-256 of the four imported files

The public downloads accompanying the web page are this document and the companion document; the files listed above are retained in the research repository and are not distributed as downloads.

The audit confirms available counts and arithmetic. Open items beyond the arithmetic are human-validated detector accuracy, exact historical command-line settings or runtime, capture-time hash validation, and reproducibility of future model outputs. The forced-search tool_invoked field is set by either a web_search_call item or a citation annotation and is not a retained raw trace; citation-count definitions also differ between extractors (the original deduplicates URLs, the forced counts annotations). Imported file hashes identify recovered bytes, not proof that those exact bytes ran historically.

Appendix A: Entity Universe and Disambiguation Notes

The fourth column lists the regex alternatives exactly as run in wave1_corrections.py (case-insensitive; a response counts as a mention if any alternative matches). Short names such as Ross, Johnson, Smith, Foster, Marshall, and Anderson require a co-occurring context token via lookahead.

Entity	Canonical Name	Abbreviations	Detection patterns (regex, as run)

Harvard Business School	Harvard Business School	HBS	`Harvard Business School`
			`\bHBS\b`
			`Harvard\s+Business`

Stanford GSB	Stanford Graduate School of Business	GSB, Stanford GSB	`Stanford\s+G(?:raduate\s+)?S(?:chool\s+of\s+)?B(?:usiness)?`
			`Stanford\s+GSB`
			`Stanford\b(?!.*University(?! (?:Graduate|Business)))`

Wharton	The Wharton School	Wharton, Penn, UPenn	`Wharton`
			`University of Pennsylvania.*(?:MBA|business|school)`
			`UPenn.*(?:MBA|business)`

Kellogg	Kellogg School of Management	Kellogg, Northwestern	`Kellogg` `Northwestern.*(?:MBA|business|school|Kellogg)`

Booth	Chicago Booth	Booth, University of Chicago	`Chicago\s*Booth`
			`\bBooth\b(?!.*(?:phone|call))`
			`University of Chicago.*(?:MBA|business|school)`

MIT Sloan	MIT Sloan School of Management	MIT Sloan, Sloan	`MIT\s*Sloan`
			`Sloan\s+School`
			`\bMIT\b.*(?:MBA|business|management)`

Columbia Business School	Columbia Business School	Columbia, CBS	`Columbia\s+Business` `Columbia\b.*(?:MBA|school|business)`
Tuck	Tuck School of Business	Tuck, Dartmouth Tuck	`\bTuck\b` `Dartmouth.*(?:MBA|business|school|Tuck)`

Yale SOM	Yale School of Management	Yale SOM, Yale	`Yale\s+S(?:chool\s+of\s+)?(?:OM|Management)`
			`Yale\s+SOM`
			`Yale\b.*(?:MBA|business|management|school)`

Fuqua	Duke Fuqua	Fuqua, Duke Fuqua	`\bFuqua\b` `Duke.*(?:MBA|business|school|Fuqua)`

Ross	Michigan Ross	Ross, Michigan Ross	`Michigan\s*Ross`
			`\bRoss\b(?=.*(?:Michigan|MBA|business|school))`
			`University of Michigan.*(?:MBA|business|school|Ross)`

Stern	NYU Stern	Stern, NYU Stern	`NYU\s*Stern`
			`\bStern\b(?=.*(?:NYU|business|school))`
			`NYU.*(?:MBA|business|school)`

Darden	Darden School of Business	Darden, UVA Darden	`\bDarden\b`
			`UVA.*(?:MBA|business|school|Darden)`
			`University of Virginia.*(?:MBA|business|school|Darden)`

Haas	Berkeley Haas	Haas, UC Berkeley	`Berkeley\s*Haas`
			`\bHaas\b`
			`UC\s*Berkeley.*(?:MBA|business|school|Haas)`

McCombs	McCombs School of Business	McCombs, UT Austin	`\bMcCombs\b`
			`UT\s*Austin.*(?:MBA|business|school|McCombs)`
			`University of Texas.*(?:MBA|business|school|McCombs)`

Anderson	UCLA Anderson	Anderson, UCLA	`UCLA\s*Anderson`
			`Anderson\b(?=.*(?:UCLA|MBA|business|school))`
			`UCLA.*(?:MBA|business|school)`

Tepper	Tepper School of Business	Tepper, CMU	`\bTepper\b`
			`Carnegie\s*Mellon.*(?:MBA|business|school|Tepper)`
			`\bCMU\b.*(?:MBA|business|school|Tepper)`

Entity	Canonical Name	Abbreviations	Detection patterns (regex, as run)

Kenan-Flagler	UNC Kenan-Flagler	Kenan-Flagler, UNC	`Kenan[\s-]*Flagler`
			`\bUNC\b.*(?:MBA|business|school|Kenan)`
			`University of North Carolina.*(?:MBA|business|school)`

Marshall	USC Marshall	Marshall, USC Marshall	`USC\s*Marshall`
			`Marshall\b(?=.*(?:USC|MBA|business|school))`
			`USC.*(?:MBA|business|school)`

Mendoza	Mendoza College of Business	Mendoza, Notre Dame	`\bMendoza\b` `Notre\s*Dame.*(?:MBA|business|school|Mendoza)`

Johnson	Johnson Graduate School of Management	Johnson, Cornell	`Cornell\s*(?:Johnson|SC\s*Johnson)`
			`Johnson\b(?=.*(?:Cornell|MBA|business|school))`
			`Cornell.*(?:MBA|business|school)`

Olin	Olin Business School	Olin, WashU	`Olin\s+Business`
			`Wash(?:ington)?\s*U(?:niversity)?.*(?:Olin|MBA|business|school)`
			`\bWashU\b.*(?:MBA|business|school|Olin)`

Smith	Smith School of Business	Smith, Maryland Smith	`Maryland\s*Smith`
			`Smith\b(?=.*(?:Maryland|MBA|business|school))`
			`University of Maryland.*(?:MBA|business|school)`

Carey	Carey Business School	Carey, Johns Hopkins	`Johns\s*Hopkins.*(?:Carey|MBA|business|school)` `\bCarey\b(?=.*(?:Johns\s*Hopkins|MBA|business|school))`

Foster	Foster School of Business	Foster, UW Foster	`UW\s*Foster`
			`Foster\b(?=.*(?:UW|Washington|MBA|business|school))`
			`University of Washington.*(?:MBA|business|school|Foster)`

Appendix B: Scoring Formula Reference

Symbol	Formula	Notes
M1	k / n	k = responses with mention; n = responses in mode
Wilson CI	$(2k+z^2 \pm z\sqrt{(z^2+4k(1-k/n))}) / (2(n+z^2))$	z = 1.96
ΔM1	M1-AGSO − M1-NGMO	Positive = more mentions in the search configuration
CSC	1 − (σ(strata M1) / mean(strata M1))	Clipped [0,1]; undefined if mean = 0
Valence ratio	positive mentions / total mentions	200-char window around first mention
McNemar	exact binomial on min(b,c) of n=b+c	two-sided
BH q	sort p ascending; q(i) = min(1, min over j≥i of m·p(j)/j)	m = 25, α = 0.05

Appendix C: Original Optional-Search Scores (publication revision 1.2.2)

Regenerated from mba_aepi_index_scores_wave1_chatgpt_v2.csv; corrected detection; Wilson 95% CIs; rank residual = AI rank minus OLS-predicted rank (positive = under-indexed, negative = over-indexed).

Rk	Entity	M1-AGSO	CI lo	CI hi	M1-NGMO	ΔM1	Class	McN p	BH q	Rank resid
1	Stanford GSB	0.627	0.587	0.664	0.600	+0.027	Stable	0.0559	0.2329	−1.68
2	Wharton	0.625	0.586	0.663	0.650	−0.025	Stable	0.0201	0.1004	+0.28

Rk	Entity	M1-AGSO	CI lo	CI hi	M1-NGMO	ΔM1	Class	McN p	BH q	Rank resid
3	Harvard Business School	0.603	0.564	0.642	0.605	−0.002	Stable	1.0000	1.0000	+0.32
4	Columbia Business School	0.592	0.552	0.630	0.590	+0.002	Stable	1.0000	1.0000	−3.48
5	Kellogg	0.573	0.533	0.612	0.598	−0.025	Stable	0.1006	0.3594	−0.56
6	MIT Sloan	0.558	0.518	0.598	0.573	−0.015	Stable	0.3492	0.7856	+1.40
7	Haas	0.473	0.434	0.513	0.495	−0.022	Stable	0.1539	0.4810	−0.48
8	Booth	0.442	0.402	0.482	0.442	+0.000	Stable	1.0000	1.0000	+2.44
9	Tuck	0.280	0.246	0.317	0.325	−0.045	Stable	0.0093	0.0664	−1.35
10	Yale SOM	0.275	0.241	0.312	0.268	+0.007	Stable	0.6358	1.0000	+0.61
11	Stern	0.245	0.212	0.281	0.250	−0.005	Stable	0.8304	1.0000	−0.31
12	Ross	0.188	0.159	0.222	0.233	−0.045	Stable	0.0040	0.0505	−0.27
13	Fuqua	0.177	0.148	0.209	0.190	−0.013	Stable	0.3409	0.7856	+0.73
14	Darden	0.122	0.098	0.150	0.122	+0.000	Stable	1.0000	1.0000	−0.19
15	Anderson	0.117	0.093	0.145	0.130	−0.013	Stable	0.3123	0.7856	+0.81
16	Tepper	0.087	0.067	0.112	0.097	−0.010	Stable	0.3771	0.7856	−2.98
17	Marshall	0.073	0.055	0.097	0.078	−0.005	Stable	0.5811	1.0000	−0.06
18	Foster	0.062	0.045	0.084	0.060	+0.002	Stable	1.0000	1.0000	−1.94
19	McCombs	0.060	0.044	0.082	0.088	−0.028	Stable	0.0033	0.0505	+1.94
20	Johnson	0.048	0.034	0.069	0.070	−0.022	Stable	0.0106	0.0664	+3.90
21	Kenan-Flagler	0.032	0.020	0.049	0.028	+0.003	Stable	0.6875	1.0000	+0.10
22	Mendoza	0.020	0.011	0.035	0.018	+0.002	Stable	1.0000	1.0000	+0.14
22	Olin	0.020	0.011	0.035	0.020	+0.000	Stable	1.0000	1.0000	−1.77
24	Carey	0.015	0.008	0.028	0.013	+0.002	Stable	1.0000	1.0000	−0.73
25	Smith	0.013	0.007	0.026	0.012	+0.002	Stable	1.0000	1.0000	+3.14

Addendum: Required-Search Condition

The included raw captures contain 600 unique prompt IDs timestamped September 5, 2026, 01:14:52–02:16:52 UTC. All 600 record the invocation flag; 583 contain citations (97.2%), mean 6.79 citation annotations (not deduplicated). Persona citation presence is 115/120 (95.8%). The earlier NGMO baseline is reused; no concurrent control was collected. Under the primary BH rule, 13 programs are Search-Suppressed, three Search-Amplified, and nine Stable; this is not a materiality-gated result. The forced-search top five are Wharton (0.447), Kellogg (0.408), Harvard (0.387), MIT Sloan (0.385), and Stanford GSB (0.383), distinct from the optional-search ranks in Appendix C. A post-hoc 540-prompt sensitivity analysis excluding the named-school Factual prompts preserved the optional-search ranking, all BH classifications, and the forced-search top six; see the companion document. Future dual-mode collections should pair NGMO with required search, log invocation explicitly, collect both arms close together, and validate detection. This release makes no claims about other engines.

MBA-AEPI ChatGPT publication revision 1.2.2 | September 2026. Based on v1.9.0-FROZEN and Amendment v1.9.1, with statistical classification distinguished from unassessed materiality.