Brainpan.AI

FULL BENCHMARK REPORT

Insurance AIVI Full Report

Q3 2026

A complete account of the Q3 Core 5 Expanded benchmark, the additive Q3 Expanded 6 view, Meta AI, Sentiment, Accuracy, methodology, scoring, quality assurance, and Q2 movement.

300 PROMPTS | 1,800 RESPONSES | 226 ENTITIES | 6 SURFACES

Published by Brainpan.AI | Kevin Walsh, Founder

October 2026

CONTENTS

Report structure

Q3 now presents the visibility benchmark alongside Sentiment and Accuracy, scored on the same response corpus.

01	Executive summary	3-4
02	Benchmark results	5-9
03	Quarter movement	10-12
04	Entity universe	13-15
05	Sentiment analysis (NEW)	16-18
06	Accuracy analysis (NEW)	19-22
07	Prompt intent	23
08	Surface analysis	24-30
09	Citation landscape	31
10	Methodology	32-38
11	QA and governance	39-40
12	Appendix	41

THE RESULT

A stable leader across visibility, sentiment, and accuracy.

State Farm remains first on visibility across every view and surface. Q3 adds two new analyses layered onto the same corpus: how brands are actually discussed, and whether specific factual claims about them hold up.

1,500
CORE 5 RESPONSES
Official benchmark

226
ENTITY UNIVERSE
Canonical count

6,917
SCORED SENTIMENT MENTIONS
Product-eligible

1,065
VERIFIED ACCURACY CLAIMS
98.4% match rate

- **Leadership.**
  State Farm ranks first in both visibility reporting views and on all six measured surfaces.
- **Challenger.**
  USAA advances to second in the official Core 5 Expanded benchmark.
- **Expansion.**
  Open-world discovery increases the entity universe from 134 to 226.
- **Depth.**
  Sentiment and Accuracy are scored for the first time this quarter, on the same 1,800-response corpus used for visibility.
- **A real accuracy gap.**
  Verified Accuracy Rate is 59.1% overall, with two systemic failure patterns (invented ratings and stale facts), each confirmed across multiple independent AI engines.

STUDY AT A GLANCE

Three analyses, one governed response corpus.

Visibility, Sentiment, and Accuracy are each scored independently, with their own locked taxonomy and validation discipline, against the same 1,800 Q3 responses.

ANALYSIS	UNIT SCORED	ROWS	KEY RATE(S)
Visibility (AIVI)	Entity mention / response	7,702 mentions	AIVI composite, 0-100
Sentiment	Entity mention / response	6,917 scored	64.8% positive share
Accuracy	Factual claim / response	1,065 resolved	59.1% verified accuracy

Naming guardrail

Core 5 is the five-engine panel measured against the 134-entity Q2 reference set, reserved for Q2-to-Q3 trend math. Core 5 Expanded is the same five engines measured against the full current-quarter universe (226 entities) and is the official visibility benchmark. Use Q3 Expanded 6 only when explicitly including Meta AI. Do not use the label Core 6. Sentiment and Accuracy are reported on the full Expanded 6 response set.

CORE 5 EXPANDED LEADERBOARD

Official benchmark leaders 1-10.

AIVI is normalized to the leading entity within this reporting view. Full 1-20 leaderboards are unchanged from the prior release; see the data appendix for the complete table.

RANK	ENTITY	TYPE	MENTIONS	REC RATE	CITATIONS	AIVI
1	State Farm	Carrier	521	53.4%	157	100.0
2	USAA	Carrier	336	57.1%	101	67.7
3	GEICO	Carrier	258	54.7%	94	56.9
4	Progressive	Carrier	293	46.8%	80	53.8
5	Nationwide	Carrier	248	53.6%	106	50.4
6	Amica	Carrier	182	63.7%	105	50.2
7	AM Best	Rating agency	194	100.0%	46	49.6
8	Travelers	Carrier	164	64.0%	99	44.7
9	Allstate	Carrier	237	37.6%	51	40.6
10	Chubb	Carrier	160	55.6%	52	38.3

EXPANDED 6 LEADERBOARD

Additive market view leaders 1-10.

Meta AI changes the mix below the leader and increases the prominence of citation-led entities.

RANK	ENTITY	TYPE	MENTIONS	REC RATE	CITATIONS	AIVI
1	State Farm	Carrier	644	54.3%	426	100.0
2	USAA	Carrier	424	58.0%	239	66.4
3	Progressive	Carrier	389	47.6%	244	57.3
4	GEICO	Carrier	326	52.5%	224	55.2
5	AM Best	Rating agency	242	100.0%	101	49.0
6	Nationwide	Carrier	319	56.1%	194	48.3
7	Amica	Carrier	233	64.4%	189	47.3
8	Travelers	Carrier	234	59.8%	212	44.4
9	Allstate	Carrier	317	39.1%	177	44.4
10	Chubb	Carrier	205	57.1%	124	37.8

META AI LEADERBOARD

Meta AI supplementary leaders 1–10.

Meta is reported independently and then added to Core 5 Expanded for the Expanded 6 view.

RANK	ENTITY	TYPE	MENTIONS	REC RATE	CITATIONS	AIVI
1	State Farm	Carrier	123	58.5%	269	100.0
2	Progressive	Carrier	96	50.0%	164	69.6
3	USAA	Carrier	88	61.4%	138	66.1
4	Allstate	Carrier	80	43.8%	126	56.6
5	Travelers	Carrier	70	50.0%	113	53.8
6	GEICO	Carrier	68	44.1%	130	52.9
7	Nationwide	Carrier	71	64.8%	88	52.6
8	AM Best	Rating agency	48	100.0%	55	50.2
9	Amica	Carrier	51	66.7%	84	48.6
10	Erie Insurance	Carrier	62	61.3%	88	47.1

CORE 5 EXPANDED V EXPANDED 6

Rank changes after adding Meta AI.

Positive change means an entity improves its position in Expanded 6.

ENTITY	C5 EXP RANK	EXP. RANK	CHANGE	C5 EXP MENT.	META	EXP. MENTIONS
State Farm	1	1	+0	521	123	644
USAA	2	2	+0	336	88	424
Progressive	4	3	+1	293	96	389
GEICO	3	4	-1	258	68	326
AM Best	7	5	+2	194	48	242
Nationwide	5	6	-1	248	71	319
Amica	6	7	-1	182	51	233
Travelers	8	8	+0	164	70	234
Allstate	9	9	+0	237	80	317
Chubb	10	10	+0	160	45	205

What changes

Progressive moves to third, AM Best to fifth, and J.D. Power to twelfth. State Farm and USAA remain first and second.

SCORE COMPOSITION

Reach, recommendation, position, citations, and top-three presence.

Each component is normalized to the strongest entity in the reporting view before weighting.

35%
MENTION SHARE

25%
RECOMMENDATION POINTS

20%
AVERAGE POSITION SCORE

15%
CITATION CREDITS

5%
TOP-THREE MENTIONS

TIER	COMPLETE MENTIONS
robust_30_plus	30 or more
established_10_to_29	10-29
emerging_3_to_9	3-9
trace_1_to_2	1-2
not_observed	0

Q2 BASELINE

Mentions and citation credits, counted consistently.

Entity-response pairs are counted once per response, including multi-entity passages.

134	4,328	1,348	PASS
Q2 REFERENCE ENTITIES	Q2 MENTIONS	Q2 CITATION CREDITS	BASELINE
Continuity universe	Complete entity-response pairs	Many-to-many attribution	Used for Core 5 trend

- **Mention reach increased 13.3% on a like-for-like basis.**
  The same 134-entity cohort grew from 4,328 to 4,905 mentions. The full Q3 Core 5 Expanded universe of 226 entities reaches 5,748, with newly tracked entities contributing 843 of that increase.
- **Citation credits increased 27.5% on a like-for-like basis.**
  The same 134-entity cohort grew from 1,348 to 1,719 citation credits. The full Q3 Core 5 Expanded universe reaches 1,928, with newly tracked entities contributing 209 of that increase.
- **Expanded 6 is not the trend denominator.**
  Meta AI was added after the original Q3 core design.

TOP-TEN MOVEMENT

Top-ten positions shift beneath State Farm.

The comparison uses the Q2 baseline and Q3 Core 5 Expanded.

ENTITY	Q2	Q3 CORE 5 EXP.	RANK CHANGE	MENTION CHANGE	CITATION CHANGE
State Farm	1	1	+0	+64	+28
Progressive	2	4	-2	-15	-33
USAA	3	2	+1	+66	-2
GEICO	4	3	+1	-7	+0
Nationwide	5	5	+0	-4	-13
Allstate	6	9	-3	-39	-32
Travelers	7	8	-1	-6	+22
The Hartford	8	11	-3	-10	-6
Chubb	9	10	-1	+20	-8
Amica	10	6	+4	+58	+50

Notable movement

Amica rises four ranks to sixth. USAA and GEICO each gain one. Allstate and The Hartford each fall three positions.

ESTABLISHED MOVERS

The largest changes with repeatable evidence.

This view requires at least 10 mentions in both Q2 and Q3 to reduce thin-evidence rank volatility.

ENTITY (LARGEST GAINS)	Q2	Q3	RANK	MENTIONS	CITATIONS
Farmers Insurance	93	55	+38	+44	+14
Erie Insurance	49	14	+35	+75	+28
Guardian Life	71	38	+33	+29	+22
Banner Life	61	31	+30	+26	+27
Protective Life	75	46	+29	+17	+30

ENTITY (LARGEST DECLINES)	Q2	Q3	RANK	MENTIONS	CITATIONS
Insurance.com	59	153	-94	-10	-3
The General	41	116	-75	-7	-1
Prudential	56	117	-61	+0	+1
NerdWallet	11	69	-58	-62	-44
Gerber Life	104	156	-52	+3	+3

OPEN-WORLD RESOLUTION

The measured market expands from 134 to 226 entities.

The final universe includes Q2 reference entities plus Q3 open-world discovery, each organization resolved to one canonical entity ID.

134	92	226	213
Q2 REFERENCE	Q3 ADDITIONS	FINAL UNIVERSE	OBSERVED
Preserved	Open-world	Canonical count	Expanded 6

- **Open-world discovery.**
  Organizations not on the Q2 reference list are identified directly from response text and added to the canonical universe.
- **Alias normalization.**
  Alternate naming forms are merged as aliases so the same entity is never double-counted under different names.
- **Context-controlled ambiguity.**
  Common-language uses of a brand name do not automatically create a mention; each is confirmed directly against raw response text.

Q3 CORE 5 EXPANDED

Core 5 Expanded entity types

Entity types show how carriers, authorities, marketplaces, and publishers share the visibility landscape.

ENTITY TYPE	UNIVERSE	MENTIONS	CITATION CREDITS
Carrier	132	4,306	1,536
Insurtech	25	442	101
Rating agency	5	391	86
Marketplace	22	285	78
Publisher	9	143	92
Affinity organization	1	39	17
Government marketplace	3	36	0
Broker	8	31	3

Q3 EXPANDED 6

Expanded 6 entity types

Entity types show how carriers, authorities, marketplaces, and publishers share the visibility landscape.

ENTITY TYPE	UNIVERSE	MENTIONS	CITATION CREDITS
Carrier	132	5,740	3,541
Insurtech	25	554	207
Rating agency	5	502	216
Marketplace	22	412	234
Publisher	9	242	228
Affinity organization	1	48	31
Government marketplace	3	46	4
Broker	8	40	13

SENTIMENT TAXONOMY

How brands are actually discussed, not just how often.

A locked seven-aspect taxonomy, built from a 60-row hand-classified pilot rather than assumed from the original design, scores every eligible entity mention for polarity and aspect.

8,003
TOTAL SCORED MENTIONS
Full Expanded 6 corpus

1,086
CITATION-SOURCE-ONLY
Excluded from product sentiment

6,917
PRODUCT-ELIGIBLE
Base for all rates below

64.8%
POSITIVE SHARE
Of product-eligible mentions

- **Seven locked aspects.**
  Price/value, claims handling, customer service, coverage breadth, digital experience, financial strength, trust/reputation.
- **Citation-source role is separated.**
  A Rating Agency or Publisher cited as the source of a claim about a different entity is tagged and excluded from product sentiment, not scored as if it were an insurance product.
- **Mixed polarity is a real category.**
  14.8% of scored mentions are genuinely mixed (e.g. praised for claims speed but criticized for price), not forced into positive or negative.

POLARITY AND ASPECTS

Coverage and price dominate what AI engines discuss.

Distribution across the 6,917 product-eligible scored mentions.

POLARITY	MENTIONS	SHARE
Positive	4,484	64.8%
Neutral	1,263	18.3%
Mixed	1,024	14.8%
Negative	146	2.1%

ASPECT	CO-OCCURRENCES
Coverage breadth	3,925
Price / value	3,218
Trust / reputation	1,883
Customer service	1,774
Digital experience	1,334
Claims handling	1,154
Financial strength	733

CARRIER SENTIMENT RANGE

Negative/mixed share varies widely across major carriers.

Across nine major carriers with at least 100 scored mentions each, negative/mixed share ranges from 4.5% to 23.8%. Named carrier-level detail is reserved for Brainpan's private brand diagnostic (see Section 3 of the gated-content approach in Methodology).

9

MAJOR CARRIERS COMPARED
n≥100 scored mentions each

4.5%–23.8%

NEG/MIXED SHARE RANGE
Across the comparison set

97.7%

TOP POSITIVE SHARE
Ladder, among all scored entities

64.8%

INDUSTRY POSITIVE SHARE
Overall product-eligible baseline

- **The spread at the high end is real, not noise.**
  Row-level read-through confirms the carriers nearest the top of the range show a consistent, repeated pattern across multiple platforms, not a computation artifact.
- **Life insurers skew most positive.**
  Ladder (97.7%), Haven Life (95.7%), and Protective Life (92.6%) lead on positive share among all scored entities.
- **Named carrier-level findings are a private deliverable.**
  Brand-specific negative/mixed share, root-cause drivers, and competitor positioning are available through Brainpan's private brand diagnostic, not published with the brand named.

Reporting discipline

This range reflects AI-generated brand portrayal on tested prompts, not verified customer experience, policyholder satisfaction, or claims outcomes.

ACCURACY SCOPE

Four hard, deterministic, regulated claim types.

Accuracy is deliberately restricted to checkable, regulated-data claims, not general praise. Each verdict is matched against an authoritative source hierarchy with named tolerance rules.

1,082
CLAIMS EXTRACTED
1,800 responses, 0 failures

1,065
ENTITY-RESOLVED
98.4% match rate

59.1%
VERIFIED ACCURACY RATE
Supported / all scoreable

7.9%
FALSE-CLAIM RATE
Contradicted / all eligible

- **Four claim types only.**
  Market share, rating agency score, J.D. Power ranking, binary state availability: general praise stays in Sentiment.
- **Three published rates, not one blended number.**
  Verified Accuracy Rate, Verifiability/Evidence Coverage Rate, and False-Claim Rate each answer a different question.
- **Claims are checked against an authoritative source hierarchy.**
  Government filings and rating-agency releases are weighted above a single AI-generated or unsourced claim.

RATES BY CLAIM TYPE AND PLATFORM

Rating and ranking claims score highest; market share lowest.

Verified Accuracy Rate (VAR) computed over the 1,059 scoreable claims (of 1,065 entity-resolved).

CLAIM TYPE	N	VAR
Rating agency score	357	65.3%
J.D. Power ranking	185	62.5%
Binary state availability	284	58.9%
Market share	239	47.7%
PLATFORM	**N**	**VAR**
Perplexity	44	70.5%
Meta AI supplementary	369	68.6%
Gemini	268	54.5%
OpenAI	68	52.9%
Copilot	217	52.3%
Claude	99	50.5%

Verdict distribution

Supported 626 · Partially supported 185 · Outdated 164 · Contradicted 84 · Unsupported/unverifiable 4 · Insufficiently specific 2 (of 1,065 resolved claims).

ACCURACY FINDING: INVENTED RATINGS

One recurring pattern: AI engines invent a rating-agency grade that doesn't exist.

The lowest-scoring entity in the Accuracy dataset has eight of nine Contradicted claims that independently invent the same kind of financial-strength letter grade, a hallucination pattern, not a one-off error. Named detail is a private diagnostic deliverable.

10.5%
LOWEST VERIFIED ACCURACY RATE
Single entity, n=19 claims

47.4%
HIGHEST FALSE-CLAIM RATE
Of any scored entity

8 of 9
CONTRADICTED CLAIMS
Invent a specific letter grade

3
INDEPENDENT ENGINES
Pattern recurs, not one model's error

- **The pattern recurs across engines.**
  Multiple independent AI engines, across several major platforms, independently invent the same kind of rating-agency grade for this one entity.
- **The entity's own public filings disclose this.**
  Its own disclosures state it does not hold the rating AI engines keep assigning it; the claims are not approximate, they describe a rating that does not exist.

Why this matters

Recurring across multiple independent AI engines is what separates a hallucination pattern from a one-off model error, exactly what the Accuracy Index exists to catch. The named entity and full evidence trail are available through Brainpan's private brand diagnostic.

ACCURACY FINDING: STALE FACTS

AI engines keep citing facts that stopped being true.

One widely-discussed entity in the Accuracy dataset shows a Verified Accuracy Rate of 25.9%, not from invented facts, but from real facts that went stale: a rating-agency downgrade and a market-position shift that many AI answers still haven't caught up to.

25.9%
VERIFIED ACCURACY RATE
Single entity, highest claim volume in the corpus

2
STALE-FACT TYPES
A ratings downgrade; a market-position shift

4+
INDEPENDENT CONFIRMATIONS
Each fact corroborated by separate trade-press sources

59.1%
OVERALL VERIFIED ACCURACY RATE
All 1,059 scoreable claims

- **Both stale facts are independently confirmed.**
  Each is corroborated by multiple independent trade-press sources.
- **Rating and ranking claims score highest overall.**
  65.3% and 62.5% Verified Accuracy Rate respectively; market share claims score lowest (47.7%) across the full dataset.

Why this matters

This reads as AI engines citing stale training-data facts about a widely-discussed brand, a useful signal for any carrier about how long out-of-date facts persist in AI-generated answers after a real change. The named entity and full evidence trail are available through Brainpan's private brand diagnostic.

NO-ENTITY ANALYSIS

Most entity-free answers are informational.

All 614 no-entity responses came from unbranded prompts.

PROMPT TYPE	NO-ENTITY RESPONSES	SHARE OF 614	RATE WITHIN TYPE
Educational	300	48.9%	90.9%
Problem-Solution	202	32.9%	56.1%
Comparison	98	16.0%	27.2%
Commercial	12	2.0%	2.7%
Recommendation	2	0.3%	0.7%

Interpretation

Educational and Problem-Solution prompts account for 502 responses, or 81.8% of the no-entity set. Commercial and Recommendation prompts account for only 14.

PLATFORM OVERVIEW

Shared leader, different supporting order.

Each surface is scored independently against its own leading entity.

SURFACE	LEADER	MENTIONS	CITE CREDITS	ENTITIES
ChatGPT	State Farm	1,015	0	106
Gemini	State Farm	1,547	0	151
Claude	State Farm	1,125	0	130
Perplexity	State Farm	929	939	123
Copilot	State Farm	1,132	989	135
Meta AI	State Farm	1,954	2,605	193

Surface rule

Aggregate visibility should be accompanied by platform-level diagnostics because the entities, source mix, and recommendation structure differ by surface.

CHATGPT

ChatGPT

Top entities within this surface. Scores are normalized independently and should not be compared as absolute cross-platform performance.

RANK	ENTITY	TYPE	MENTIONS	CITATIONS	AIVI	EVIDENCE
1	State Farm	Carrier	105	0	100.0	robust 30 plus
2	AM Best	Rating agency	57	0	76.9	robust 30 plus
3	GEICO	Carrier	51	0	63.9	robust 30 plus
4	Progressive	Carrier	59	0	58.7	robust 30 plus
5	USAA	Carrier	60	0	55.5	robust 30 plus

Surface readout

State Farm leads ChatGPT. The remaining order reflects this surface's retrieval, answer, and citation behavior.

GEMINI

Gemini

Top entities within this surface. Scores are normalized independently and should not be compared as absolute cross-platform performance.

RANK	ENTITY	TYPE	MENTIONS	CITATIONS	AIVI	EVIDENCE
1	State Farm	Carrier	133	0	100.0	robust 30 plus
2	USAA	Carrier	75	0	65.1	robust 30 plus
3	AM Best	Rating agency	55	0	59.8	robust 30 plus
4	J.D. Power	Rating agency	71	0	58.9	robust 30 plus
5	Progressive	Carrier	82	0	56.8	robust 30 plus

Surface readout

State Farm leads Gemini. The remaining order reflects this surface's retrieval, answer, and citation behavior.

CLAUDE

Claude

Top entities within this surface. Scores are normalized independently and should not be compared as absolute cross-platform performance.

RANK	ENTITY	TYPE	MENTIONS	CITATIONS	AIVI	EVIDENCE
1	State Farm	Carrier	117	0	100.0	robust 30 plus
2	USAA	Carrier	75	0	68.0	robust 30 plus
3	GEICO	Carrier	55	0	60.8	robust 30 plus
4	Progressive	Carrier	75	0	59.8	robust 30 plus
5	AM Best	Rating agency	26	0	53.0	established 10 to 29

Surface readout

State Farm leads Claude. The remaining order reflects this surface's retrieval, answer, and citation behavior.

PERPLEXITY

Perplexity

Top entities within this surface. Scores are normalized independently and should not be compared as absolute cross-platform performance.

RANK	ENTITY	TYPE	MENTIONS	CITATIONS	AIVI	EVIDENCE
1	State Farm	Carrier	74	87	100.0	robust 30 plus
2	Travelers	Carrier	46	54	69.1	robust 30 plus
3	Amica	Carrier	41	51	63.6	robust 30 plus
4	USAA	Carrier	49	40	62.0	robust 30 plus
5	Nationwide	Carrier	44	52	59.6	robust 30 plus

Surface readout

State Farm leads Perplexity. The remaining order reflects this surface's retrieval, answer, and citation behavior.

COPILOT

Copilot

Top entities within this surface. Scores are normalized independently and should not be compared as absolute cross-platform performance.

RANK	ENTITY	TYPE	MENTIONS	CITATIONS	AIVI	EVIDENCE
1	State Farm	Carrier	92	70	100.0	robust 30 plus
2	USAA	Carrier	77	61	82.0	robust 30 plus
3	Nationwide	Carrier	70	54	70.1	robust 30 plus
4	Amica	Carrier	46	54	63.9	robust 30 plus
5	GEICO	Carrier	47	52	60.7	robust 30 plus

Surface readout

State Farm leads Copilot. The remaining order reflects this surface's retrieval, answer, and citation behavior.

META AI

Meta AI

Top entities within this surface. Scores are normalized independently and should not be compared as absolute cross-platform performance.

RANK	ENTITY	TYPE	MENTIONS	CITATIONS	AIVI	EVIDENCE
1	State Farm	Carrier	123	269	100.0	robust 30 plus
2	Progressive	Carrier	96	164	69.6	robust 30 plus
3	USAA	Carrier	88	138	66.1	robust 30 plus
4	Allstate	Carrier	80	126	56.6	robust 30 plus
5	Travelers	Carrier	70	113	53.8	robust 30 plus

Surface readout

State Farm leads Meta AI. The remaining order reflects this surface's retrieval, answer, and citation behavior.

OBSERVED DOMAINS

Citation volume and entity credit are different measures.

The source-domain inventory records observed citation records before entity-level eligibility and local-context attribution.

DOMAIN	RECORDS
moneygeek.com	660
instagram.com	636
reddit.com	588
nerdwallet.com	531
forbes.com	364
insurify.com	351
facebook.com	345
cnbc.com	316
thezebra.com	308
marketwatch.com	286

Interpretation

High citation-record volume does not guarantee entity credit. Sponsored evidence, source-card labels, unresolved context, and unsupported markers are excluded from scoring.

STUDY DESIGN

A standardized 300-prompt six-surface capture.

The same governed prompt inventory is measured across each surface, with Core 5 frozen before the supplementary Meta AI addition.

300	6	1,800	1,826
PROMPTS	MEASURED SURFACES	ACCEPTED CELLS	RAW ATTEMPTS
Frozen inventory	Five core + Meta	Complete coverage	Retries retained

- **One accepted cell per prompt and surface.**
  Scoring uses the latest completed evidence for each cell.
- **Extra attempts remain immutable.**
  Copilot 6, Gemini 4, and Meta AI 14 are retained but not double-counted.
- **Surface-specific capture contracts.**
  API and browser-derived evidence are standardized into one response schema.
- **Frozen prompt hashes.**
  The exact prompt and SHA-256 remain attached to each accepted response.

EVIDENCE STANDARDIZATION

Raw artifacts remain immutable. Derived records are reproducible.

Every accepted cell retains its raw evidence reference and a standardized answer record.

EVIDENCE ASSET	COUNT	ROLE
Raw JSON attempts	1,826	Immutable source attempts
Accepted response cells	1,800	One per surface-query cell
Hashed evidence files	3,328	SHA-256 manifest
Screenshots	1,346	Browser evidence where required
Standardized responses	1,800	JSONL analytical record
Observed citations	17,423	Pre-eligibility citation records

ENTITY RESOLUTION

Canonical entities require response-level evidence.

Candidates come from the Q2 reference list and Q3 open-world discovery.

20,033	226	5	13
CANDIDATE RECORDS	RESOLVED ENTITIES	ALIAS MERGES	NOT OBSERVED
Open-world discovery	Canonical universe	Not double-counted	Retained unranked

- **Canonical names and types.**
  Each valid organization maps to one entity ID and one reporting type.
- **Context-controlled ambiguity.**
  Common-language uses do not automatically create a brand mention; each candidate is confirmed directly against raw response text.
- **Source-card separation.**
  Copilot publisher labels are removed from answer prose before entity matching.
- **Zero-mention retention.**
  Entities remain in the universe even when not observed in the reporting view.

SENTIMENT METHODOLOGY

A piloted taxonomy, not an assumed one.

60 entity mentions were hand-classified across all six platforms before any scaled scoring.

- **Seven locked aspects.**
  Price/value, claims handling, customer service, coverage breadth, digital experience, financial strength, and trust/reputation are scored independently per mention.
- **Citation-source role excluded from scoring.**
  28% of the pilot sample was a Rating Agency or Publisher entity cited as a source, not itself being evaluated.
- **Alias matching is curated, not mechanical.**
  Entity aliases are reviewed individually rather than auto-generated, so common English words are never mistaken for a brand name.
- **Response-level, not snippet-level, judgments.**
  Full responses are read in context (roughly 4 entity-sentiment judgments per response) rather than isolated 300-character windows around each mention.

Scope note

Sentiment is scored on the same 1,800-response corpus as Visibility and Accuracy, so all three analyses share one evidentiary base.

ACCURACY METHODOLOGY

Two-call verification, not one, for reliability.

Claim extraction (Stage A) is a structured-JSON pass; verification (Stage B) separates grounded research from schema-constrained structuring.

STAGE	WHAT IT DOES	RESULT
A: Extraction	Pulls claims matching 4 locked categories	1,082 claims, 0 failures
B: Verification	Grounded research call, then structured verdict call	1,082 verified, 0 failures
C: Resolution	Entity-resolves claims to the canonical manifest	1,065 resolved, 98.4% match

Source-hierarchy weighting

Government filings and rating-agency releases are treated as authoritative. A single AI-generated or unsourced claim is never treated as confirming evidence on its own.

MEASUREMENT UNITS

Reach, repetition, citation, and credit are kept separate.

Separating units prevents repeated wording or dense source panels from inflating primary visibility.

- **Primary mention.**
  One resolved entity per response.
- **Textual occurrence.**
  Every matched appearance of the entity string.
- **Eligible citation occurrence.**
  A citation record that passes the evidence contract.
- **Entity-citation credit.**
  A supported link between a citation's local claim context and a resolved entity.

Many-to-many rule

One response can create several mentions. One citation can support several entities. Unsupported or context-free markers create no guessed credit.

SCORING

Five normalized components produce the AIVI score.

The Q3 implementation preserves the Q2 weighting model. Sentiment and Accuracy use their own published rates instead of a composite score.

35%	25%	20%	15%	5%
MENTION SHARE	RECOMMENDATION POINTS	AVERAGE POSITION SCORE	CITATION CREDITS	TOP-THREE MENTIONS

- **Recommendation points.**
  Strong recommendations receive more weight than moderate recommendations.
- **Top-three weighting.**
  A smaller, separate 5% credit rewards landing among the top three named options.
- **Independent normalization.**
  Each reporting view and platform has its own maxima and leader.
- **Evidence qualification.**
  Mention-based tiers distinguish robust results from trace observations, applied the same way across Visibility, Sentiment, and Accuracy.

PUBLICATION GATE

Every required gate passed.

The visibility package recorded zero blocking exceptions; Sentiment and Accuracy pass their own pilot-then-scale discipline.

GATE	RESULT
Evidence completeness (visibility)	PASS
Mention adjudication sample	150/150
Citation adjudication sample	150/150
Ambiguous-name controls	PASS
Sentiment gold-check sample	60/60 hand-classified
Accuracy spot-check sample	20/20 verdicts, plus 1 independent web fact-check

LIMITATIONS

Interpret movement within the reporting contract.

The benchmark measures observed AI-answer visibility, sentiment, and factual-claim accuracy: not market share, policy quality, claims performance, or consumer conversion.

- **Gemini continuity is directional.**
  The governed Q3 replacement is not a strict same-model continuation from Q2.
- **Meta AI is supplementary.**
  Excluded from the Q2-to-Q3 headline denominator.
- **Thin evidence can create unstable ranks.**
  Use evidence tiers before interpreting large movement among low-mention entities.
- **Accuracy scope is intentionally narrow.**
  Four regulated-data claim types only; general praise stays in Sentiment.
- **One entity carries trace-tier evidence.**
  Pattern has one occurrence across 1,800 responses, real, but too thin to weight in any score comparison.

Reporting boundary

Use findings as measured visibility, sentiment, and accuracy evidence. Do not treat any of the three as a proxy for insurance product quality or financial performance.

DATA INVENTORY

The release is fully traceable from evidence to report.

The final analysis package contains evidence indexes, normalized records, extraction facts, scoring outputs, QA results, and publication tables across all three analyses.

DATASET	RECORDS	PURPOSE
Accepted response index	1,800	Evidence freeze
Complete mentions (visibility)	7,702	Entity-response facts
Eligible citations	8,401	Citation occurrences
Resolved universe (canonical)	226	Canonical entity manifest
Sentiment-scored mentions	8,003	6,917 product-eligible
Accuracy claims extracted	1,082	1,065 entity-resolved

Data integrity

Every published figure in this report traces back to an immutable raw-response record retained in Brainpan's evidence archive.