



Behing every step of
investment decision making

Orbit Insight vs Enterprise LLMs

An independent evaluation for investment research teams

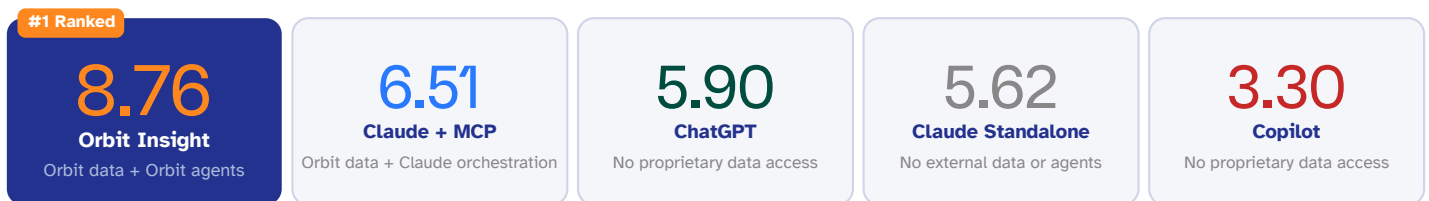
Proprietary 10-dimension scoring framework

Generic enterprise AI is trained on the world. Orbit Insight is built for investment research.

Investment teams are already using ChatGPT, Copilot, and similar tools for research, and in many cases those tools produce **answers that feel complete and authoritative**. The problem is that a **confident answer is not the same as a correct one**.

Generic AI models are trained on publicly available information and have **no access to primary financial documents, earnings call transcripts, or institutional investor meeting records**. When asked a research question, they draw on whatever was indexed on the internet, often news coverage, analyst summaries, or secondary commentary, and present it with the same confidence they would use for a verifiable fact. For investment professionals, that distinction matters enormously.

Orbit Insight sources every answer from primary financial documents, with a citation back to the original paragraph. The evaluation below tests both approaches on the same ten questions. Until you see what a properly sourced answer looks like, it is easy to assume what you already have is sufficient.

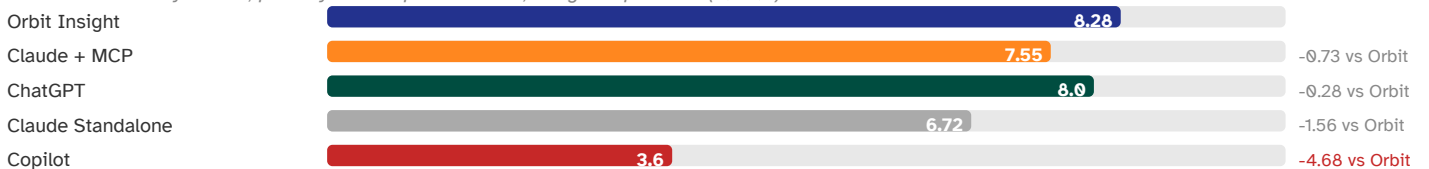


We ran **10 institutional-grade investment research questions across five AI platforms**: Orbit Insight, Claude with Orbit MCP connection, ChatGPT, Claude Standalone, and Microsoft Copilot. Each response was scored across 10 dimensions using a proprietary framework measuring source quality, data depth, exclusive content access, and analytical rigour. **Orbit Insight won every single question.**

Category Breakdown

Source-traceable & auditable

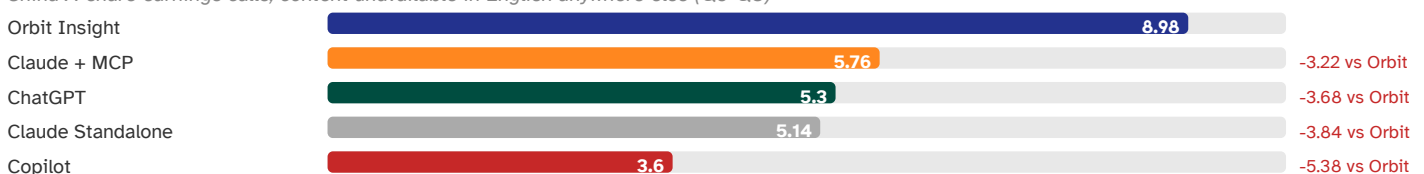
Cross-document synthesis, primary transcript attribution, filing comparisons (Q1-Q4)



ChatGPT performs strongly here — web-accessible filings partially substitute for primary documents. Copilot was unable to access transcript content on three of four questions.

Exclusive content — proprietary & foreign language

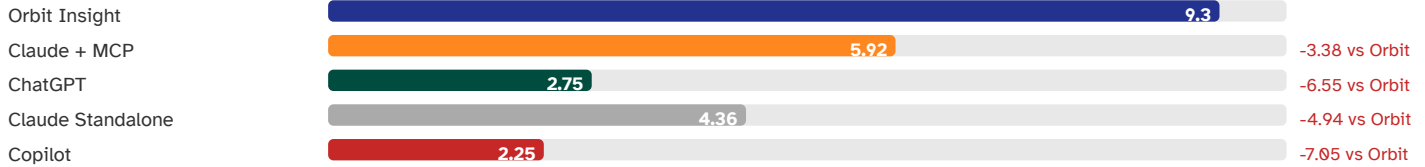
China A-share earnings calls, content unavailable in English anywhere else (Q5-Q8)



Data access dominates. All generic tools admitted inability to source primary institutional investor meeting transcripts. Copilot fully conceded on CATL and Tesla questions.

Time-series analytics & quantitative backtesting

Sentiment trends, news flow tracking, price-sentiment correlation backtesting (Q9-Q10)



Orbit's decisive advantage. Neither ChatGPT nor Copilot can run time-series sentiment analytics or price correlation backtesting without access to proprietary structured data pipelines.

WHY ORBIT WINS

9.30 / 10

Proprietary data you cannot index

Orbit processes 70M+ documents annually including China A-share institutional investor meeting transcripts not available in English anywhere else. ChatGPT and Copilot explicitly admitted they could not source primary transcript content for these questions.

+5.70 pts ahead of Copilot on exclusive content

8.28 / 10

Source citations you can verify

Every Orbit answer traces back to a specific document, date, and paragraph. For investment teams, a confident answer from a generic LLM that cannot be traced to a primary source is a liability, not an insight. Copilot scored 3.6 on source-traceable questions.

+4.68 pts ahead of Copilot on sourced answers

9.30 / 10

Quantitative analytics built for the buy-side

Orbit runs time-series sentiment analysis and price-sentiment correlation backtesting from proprietary structured data. ChatGPT scored 2.75 and Copilot 2.25 on these questions. Both admitted they cannot perform this category of analysis.

+7.05 pts ahead of Copilot on time-series analytics

KEY FINDINGS

Primary driver

Access to data, particularly proprietary foreign language content and unique structured time-series data that cannot be found elsewhere.

Primary driver

Purpose-built agent orchestration delivers significantly more value from the same underlying data.

EVALUATION METHODOLOGY

10 domain-specific questions were evaluated using Orbit Insight, Claude with MCP integration, and Claude standalone. Each answer was scored on ten custom criteria per question, including source clarity, data depth, exclusive content access, analytical rigor, and response consistency. Orbit Insight outperformed the others in every question across all categories. The full methodology and detailed results at the question level are available in the appendix and upon request.

Built by the industry, for the industry

Orbit is an award-winning financial intelligence platform built from the ground up for institutional investment teams. Every capability in the platform reflects a real workflow need from real clients: portfolio monitoring, earnings analysis, thematic research, alpha generation, regulatory tracking, ESG reporting, niche market coverage, and quantitative research. Generic AI platforms are general-purpose products that financial teams have adapted for research use. Orbit was designed the other way around. The data, the agents, the document processing, and the analytical outputs all exist because investment professionals asked for them. The answers are fundamentally different because they are created specifically for use by analysts and investment teams.



Behing every step of
investment decision making

Orbit Insight vs Enterprise LLMs

Appendix - The 10 Questions

Question design, rationale, and per-question scores across all five models

Each question below was designed to expose a specific capability gap between Orbit Insight and generic enterprise AI tools. The scoring uses a proprietary 10-dimension framework per question, with each dimension graded 1-10 based on hard data points including financial figures, direct quotes, document references, specific dates, and named individuals. Scores shown for Orbit Insight, Claude + Orbit MCP, and Claude Standalone are averages across three independent responses. ChatGPT and Copilot scores are assessed from documented POC testing responses using the same rubric.

Q1 [Source-traceable]

Question

What specific guidance did TSMC's management give on AI chip demand for the next two quarters in their most recent earnings call, and which analysts pushed back on it — by name?

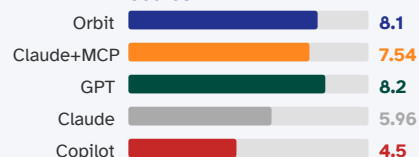
Why this question matters

Tests whether the tool can name specific analysts, their affiliations, and their precise concerns from a primary earnings call transcript. Generic tools describe calls from news articles without naming who said what.

Orbit result

Orbit returned named analysts with timestamps.

Scores



Q2 [Source-traceable]

Question

How has ASML's language around export control risks changed across their last four earnings calls, from early uncertainty through to their most recent guidance?

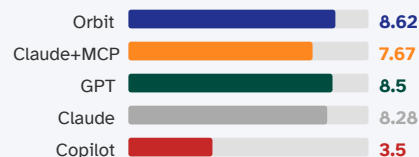
Why this question matters

Requires cross-document synthesis across four quarters from primary transcripts. Orbit tracks language evolution with direct quotes per call. Copilot admitted no export control discussion was found in any indexed transcript.

Orbit result

Orbit and ChatGPT both performed well here. ASML transcripts are publicly accessible in English.

Scores



Q3 [Source-traceable]

Question

Which new risk factors appeared in Tesla's most recent 10-K that were not present in the prior year's filing?

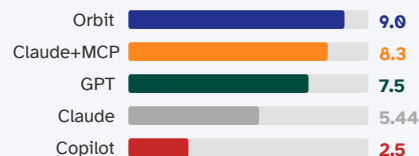
Why this question matters

Requires direct document-level comparison of two annual filings. Orbit reads both and identifies additions. Copilot stated the 2025 10-K was not yet indexed and could not answer the question.

Orbit result

Orbit's strongest result in this category. Direct filing access is the decisive differentiator.

Scores



Q4 [Source-traceable]

Question

Has Pfizer added or removed any language around pipeline risk in their two most recent annual reports, and in what direction has the tone shifted?

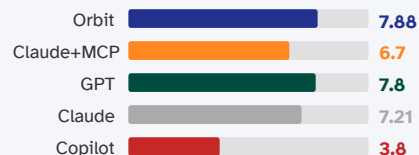
Why this question matters

Clause-level comparison across two annual reports. Orbit traces specific wording changes. Copilot cited a third-party tool reporting "no changes" — missing the co-funding and competitive landscape additions ChatGPT and Orbit both identified.

Orbit result

ChatGPT found specific clause additions from indexed filing text, although the search time was noticeably longer.

Scores





Q5 [Exclusive content]

Question

What did CATL's management say about their European gigafactory pricing strategy and timeline in their most recent institutional investor meeting?

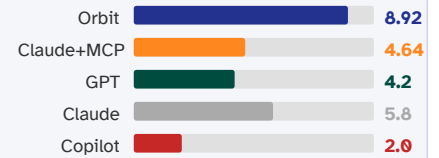
Why this question matters

Content exists only in Orbit's China A-Share dataset. Copilot explicitly stated CATL had not publicly released any institutional investor meeting statements. ChatGPT confirmed the same and offered only publicly reported press release content.

Orbit result

Orbit returned specific production figures, capacity targets, and pricing discipline commentary from the March 2026 meeting.

Scores



Q6 [Exclusive content]

Question

What did BYD's management say about EV price competition risks and their mitigation strategy in their most recent earnings call?

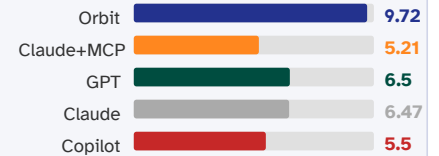
Why this question matters

BYD's institutional investor communications contain management admissions that do not appear in public filings. Orbit accessed the primary transcript. ChatGPT and Copilot sourced from news aggregators.

Orbit result

Orbit's highest score in the evaluation. Direct transcript access produced specific risk framing and mitigation detail absent from news sources.

Scores



Q7 [Exclusive content]

Question

What guidance did Kweichow Moutai give on domestic demand recovery in their most recent earnings communication, and how did management characterize the pace of any rebound?

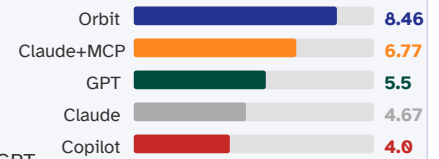
Why this question matters

Requires the actual earnings communication in translation, not analyst notes or media coverage. ChatGPT drew from Morningstar secondary analysis. Copilot produced generic commentary on China consumer sentiment without any primary source citation.

Orbit result

Orbit returned management's specific "rational consumption" framing and 9% revenue target with capacity expansion details.

Scores



Q8 [Exclusive content]

Question

What did Longi Green Energy's management say about module pricing pressure and their overseas capacity strategy in their most recent institutional investor meeting?

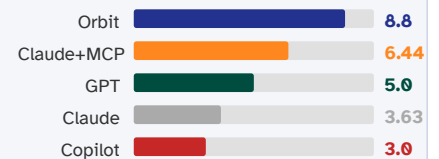
Why this question matters

Solar supply chain dynamics are actively debated across global portfolios. Orbit returns management's own words from the investor meeting including the HPBC 2.0 cost parity target and the cautious overseas expansion stance — detail not available in any public source.

Orbit result

Orbit returned specific production targets (50GW BC capacity by end-2025), international order mix (50%+), and pricing discipline rationale.

Scores





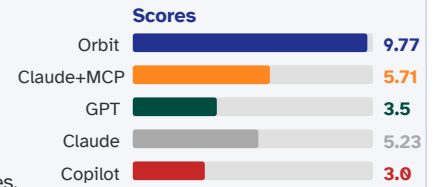
Q9 [Time-series analytics]

Question

Using Orbit's news flow tracker, compare Nvidia and AMD's sentiment trends and key themes over the last 6 months. How has the tone shifted for each company and where do they diverge most significantly?

Why this question matters Orbit produces a month-by-month sentiment score table with exact figures, identifies the precise peak divergence point (September 2025, 9.7-point gap), and explains what drove it from measured data across hundreds of news items. ChatGPT and Copilot produce narrative without any quantification.

Orbit result Orbit's second-highest score. Month-by-month sentiment table, competitor mention frequency mapping, and divergence cause analysis.



Q10 [Time-series analytics]

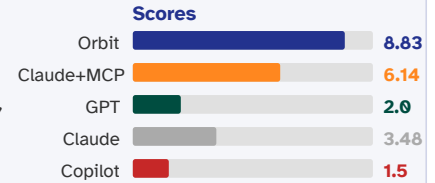
Question

Using Microsoft's news sentiment alongside end-of-day price data, run a backtesting analysis to measure the correlation between sentiment shifts and price movements over the last 12 months, and explain the key findings.

Why this question matters

Orbit computed Pearson and Spearman correlations, tested lagged effects, segmented performance by sentiment regime, and generated six statistical charts. ChatGPT described what such analysis would show conceptually. Copilot admitted it cannot access time-series data to run the analysis.

Orbit result Orbit ran full statistical backtesting including lag effect testing and regime segmentation — output quality is equivalent to a quant research report.



SCORING RUBRIC

Each dimension is scored 1-10 based on hard data point count in the response:

10 — Perfect	25+ hard data points	Exceeds professional research report standards
9 — Outstanding	18-25	Near professional research report quality
8 — Excellent	12-18	Deep analysis, data-dense, clearly structured
7 — Strong	8-12	Rich content, verifiable document references
6 — Good	5-8	Fairly complete, clear data and timeline
5 — Average	3-5	Covers main aspects, partial data support
4 — Below average	2-3	Basic framework, lacks depth and concrete data
3 — Weak	1-2	Minimal relevant information, no sources
1-2 — Poor	0-1	Vague, blank, or completely irrelevant response

Hard data points include: financial figures, percentages, specific dates, quarter references, named individuals, direct quotes of 15+ characters, document references (10-K, 10-Q, Annual Report, Transcript), and source links. A noise penalty is applied where non-content text exceeds 15% of the response. Full methodology and question-level scoring breakdown available on request.



Orbit is an award-winning investment research platform designed to empower informed, data-driven decision-making ahead of the curve. Orbit's extensive coverage, coupled with rigorous quality control measures, ensures access to reliable and timely information. Orbit facilitates automated workflows and comprehensive data analysis, streamlining research processes and uncovering critical insights. By committing to data integrity and clarity, Orbit enables users to confidently navigate complex financial landscapes and make strategic decisions with precision.



orbitfin.ai



info@orbitfin.ai



orbit-fintech



@orbitfin