

Team Members: Sourav Bhunia, Divyanshu Sharma, Zach Farkas, Elle Carroll, Nikhila Nandgopal

“AI is taking jobs, and now your analyst job too. Booth Students, that doesn’t mean you should outsource your EFPE assignment to ChatGPT or Perplexity.”

The CPG space continues to see rapid growth, new entrants, and substantial investment in emerging and evolving markets. As investors look to research new companies, they face the tedious process of vetting analog companies. Research and database information providers such as Pitchbook and CB Insights lack specification and nuance to accurately report the market landscapes and true competitors.

Case Study: Alcohol Alternative Beverages

The alternative alcohol beverage market is experiencing significant growth driven by changing consumer preferences toward health, wellness, and moderation. In the U.S. alone, the non-alcoholic beverage market is expected to surpass \$30 billion by 2025, with non-alcoholic spirits, wine, and beer showing year-over-year growth of 29%, 18%, and 15%, respectively.¹

Growing market demand coupled with low entry barriers for new competitors has resulted in an explosion of start-ups entering this beverage market segment. These companies are reshaping the beverage landscape by offering innovative products that align with the growing "sober-curious" movement and consumer demand for healthier lifestyle choices.

But how does one segment this market? Current analysis and reporting confounds direct alcohol alternatives (e.g., non-alcoholic wines or spirits) and designer beverages that offer unique flavor profiles without mimicking traditional alcoholic drinks. This leads to inaccurate information on true analogs for specific beverage companies.

Current databases and research publications misappropriate competition in this space. Is Athletic Brewing (the largest non-alcoholic beer brand in the US) an analog to Liquid Death (canned water), Poppi (prebiotic sodas), and Spindrift (sparkling water), as Pitchbook claims?² Not really. Without the nuances of specific market positioning reported by incumbent databases, analysts resort to time-consuming and tedious diligence to vet true analogs.

Enter the AI Toolkit for Diligence

Goal: Provide financial analysts with an understanding of the quality of outputs from different off-the-shelf tools (OpenAI and Perplexity) to find industry comps and assess market positions in the growing alcohol substitute beverages industry.

Intended User: Financial analysts focusing on the CPG markets, performing due diligence for non-alcoholic, non-THC beverage sector (alcohol substitutes)

Analysis: The most important aspect of our analysis is the comparison of different off-the-shelf tools and their implementations by providing them with relevant industry documents and custom instructions. By “off-the-shelf tools” we mean the underlying systems and methodologies that drive the AI platforms such as OpenAI and Perplexity.

How It Works

Our project aims to create custom instructions that follow a detailed workflow designed to mimic a human approach to market analysis. This structured process ensures that every analysis is thorough, data-driven, and actionable. The workflow is broken down into five key steps, each focusing on different market and company evaluation aspects.

1. Target Company Profiling

The first step in our workflow is to develop a comprehensive profile of the target company. This involves processing inputs from text prompts and uploaded files such as PDFs, spreadsheets, and presentations. We prioritize the extraction of information in the following order: Financials, Operations, and Growth Metrics. The output is structured in a clear table format that includes categories like Core Business, Financial Profile, Market Position, and Growth Drivers. For example, the structured output requirements include details such as product lines, packaging types, revenue trends over three

¹ INFRA “Cheers to Dry January: Trends in Non-Alcoholic Beverages”

<https://www.naturalfoodretailers.com/blogs/madeleine-ware/2025/01/03/cheers-to-dry-january-trends-in-non-alcoholic-beve>

² Pitchbook: Athletic Brewing “Similar Companies” <https://my.pitchbook.com/profile/221707-45/company/profile#similar-companies>

years, EBITDA margins, and key sales channels. This step lays the foundation for understanding the company's overall landscape.

2. Market Research Protocol

Next, we integrate data from various sources using a defined hierarchy. The highest priority is given to user-uploaded market reports, followed by official SEC filings (10-K/Q), and recent company news results.

The goal for the output analysis is to cover multiple elements such as market fundamentals (e.g., Total Addressable Market size and 5-year CAGR), geographic breakdowns, and channel analysis comparing retail versus direct-to-consumer (D2C) performance. We also want to delve into competitive landscape details, including market share matrices, M&A trends over the last 36 months, and consumer dynamics like repeat purchase rates and demographic penetration. Additionally, the aim is to include risk frameworks with regulatory changes, supply chain vulnerabilities, and market disruptions using varied assessment methods like probability-impact matrices and SWOT analyses.

3. Peer Identification System

In this step, we try identifying peer companies based on similarity, market capitalization relative to the target company, and geographic overlap. This systematic approach ensures that only companies meeting a minimum threshold of similarity are included in our comparative analysis. It enables us to objectively identify peers that are relevant for benchmarking and further financial evaluation.

4. Financial Benchmarking

Once the peers are identified, we move to financial benchmarking. This involves a rigorous data verification process where we gather financial metrics for each public comparable. The models are instructed to draw data from multiple sources and then cross-validate them from sources like stock analysis websites and the latest 10-Q filings.

Our focus is on calculating key ratios such as EV/Revenue and EV/EBITDA. We extract metrics including Enterprise Value, Market Cap, EBITDA, Capex, Equity Beta, and Asset Beta, all presented in a tabular format.

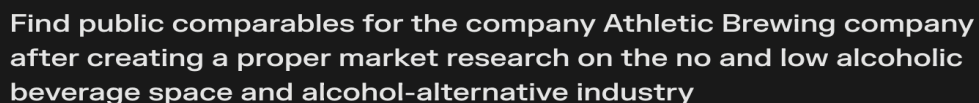
5. Integrated Reporting Structure

The final step in our instruction workflow is to compile all the gathered insights into an integrated report. The final report follows a structured template that includes an Executive Summary, a deep dive into market analysis, and a comparative company analysis. The Executive Summary highlights key valuation drivers and immediate risk factors. The market deep dive covers market size, growth projections, competitive positioning, consumer trends, and regulatory risk assessments. For the comparable company analysis, we detail the peer selection rationale, present benchmarking tables, and discuss valuation premiums or discounts. Throughout this process, we adhere to strict implementation rules, such as dual-source verification for market size claims and proper citation of data provenance in charts, ensuring the reliability and transparency of our analysis.

The aim with this comprehensive instruction workflow is to rigorously assess market opportunities and strategic risks, providing a robust framework for both internal decision-making and client-facing reporting.

Explanation of Model Implementation:

We prompted the different off-the-shelf tools to apply our workflow to analyze Athletic Brewing Company. The prompt was:



Find public comparables for the company Athletic Brewing company after creating a proper market research on the no and low alcoholic beverage space and alcohol-alternative industry

Figure 1: Prompt provided to different models to generate public comparables for Athletic Brewing Company

This guided the model to execute comprehensive market research, identify key industry trends, and extract relevant financial and competitive data, ensuring a robust analysis aligned with our structured methodology.

Output Examples:

Good Output example

| Comparable Companies | | | | | | | | |
|------------------------------------|--------|----------------|-----------|-------------|------------|-----------|---------------|------|
| Company | Ticker | Market Cap (M) | EV (M) | Revenue (M) | EBITDA (M) | EV/EBITDA | Capex/Revenue | Beta |
| Anheuser-Busch InBev | BUD | \$210,530 | \$246,170 | \$91,850 | \$18,370 | 13.4x | 6.2% | 0.89 |
| Heineken NV | HEINY | \$58,440 | \$67,890 | \$34,120 | \$6,120 | 11.1x | 5.8% | 0.76 |
| Boston Beer Company | SAM | \$3,810 | \$3,950 | \$1,190 | \$98 | 40.3x | 4.1% | 1.12 |
| Molson Coors Beverage | TAP | \$12,110 | \$14,320 | \$11,240 | \$1,870 | 7.7x | 5.5% | 0.94 |
| Diageo PLC | DEO | \$92,330 | \$108,200 | \$23,560 | \$7,450 | 14.5x | 8.3% | 0.82 |
| Data Sources: StockAnalysis 1 3D . | | | | | | | | |

Figure 2: Comparables table generated by Perplexity Pro with custom instructions and curated market research documents

The above table is a good output because it cleanly presents the key valuation and operational metrics (EV, Revenue, EBITDA, EV/EBITDA, etc.) side by side for each comparable company, making it easy to compare and benchmark. It also includes relevant columns like Capex/Revenue and Beta for risk and capital intensity assessment, and cites a data source for transparency. Such clarity and breadth of data closely follow best practices for a robust comparables analysis.

Subpar Output example

| Public Comparables for Athletic Brewing Company | | | |
|--|--------|----------------------------|---|
| Athletic Brewing Company, a leader in non-alcoholic craft beer, is uniquely positioned in an industry where major global brewers are also expanding their low and no-alcohol portfolios. Public comparables in this space typically include large, diversified beverage companies that have made strategic investments in non-alcoholic and low-alcohol offerings. Examples include: | | | |
| Comparable Company | Ticker | Key Non-Alcoholic Offering | Market Focus & Rationale |
| Heineken N.V. | HEIA | Heineken 0.0 | A global leader that has successfully launched Heineken 0.0, capturing growing demand for NA beer. |
| Carlsberg A/S | CARL-B | Carlsberg 0.0 | Actively investing in NA innovation; its strategy in low/NA segments offers a strong benchmark. |
| Anheuser-Busch InBev | BUD | Multiple NA variants | As the world's largest brewer, AB InBev's extensive portfolio includes several low-alcohol and non-alcoholic options. |
| Boston Beer Company | SAM | (Emerging NA offerings) | While primarily known for craft beers, Boston Beer Company is venturing into the NA space, making it a relevant comparator. |
| Molson Coors Beverage Company | TAP | Expanding NA portfolio | Molson Coors is broadening its product mix to include non-alcoholic options, reflecting a similar growth dynamic. |

Figure 3: Comparables table generated by OpenAI o3-mini

The above output lacks the essential components of a comparables analysis. It names a few major brewers and offers broad statements about market focus, but doesn't provide key metrics (revenue, EBITDA, margins, or valuation multiples) nor does it explain how each company's scale, product mix, or distribution strategy compares to the target. There's no

quantitative data for valuation or performance benchmarking, and no discussion of adjustments (e.g., for brand premium, differing growth rates, or profitability). As a result, the table provides only a superficial overview instead of the deeper, metrics-driven analysis required for a robust comparable companies evaluation.

Explanation of Evaluation Criteria:

For evaluation, we are drawing inspiration from established evaluations like GAIA, a well-regarded benchmark in model evaluation, which provides a structured way to assess the quality of AI outputs. The GAIA benchmark is an excellent starting point because it offers a clear set of criteria to evaluate things like data completeness, analytical depth, and the ability to generate actionable insights. However, since GAIA was not originally designed with a business or finance focus in mind, our team is exploring ways to adapt and extend its methodology. The goal here is to develop an evaluation tailored specifically to our domain. This involves identifying what GAIA might be missing when it comes to assessing real-world applications in finance and then rigorously testing various frameworks against these adjusted criteria.

Our evaluation process checks every report on three levels.

- 1. We confirm that the report has all the basic information it needs—this is our "completeness" check.
- 2. We look at how well the report explains its data, methods, and assumptions.
- 3. We see if the report turns its findings into useful, forward-looking advice.

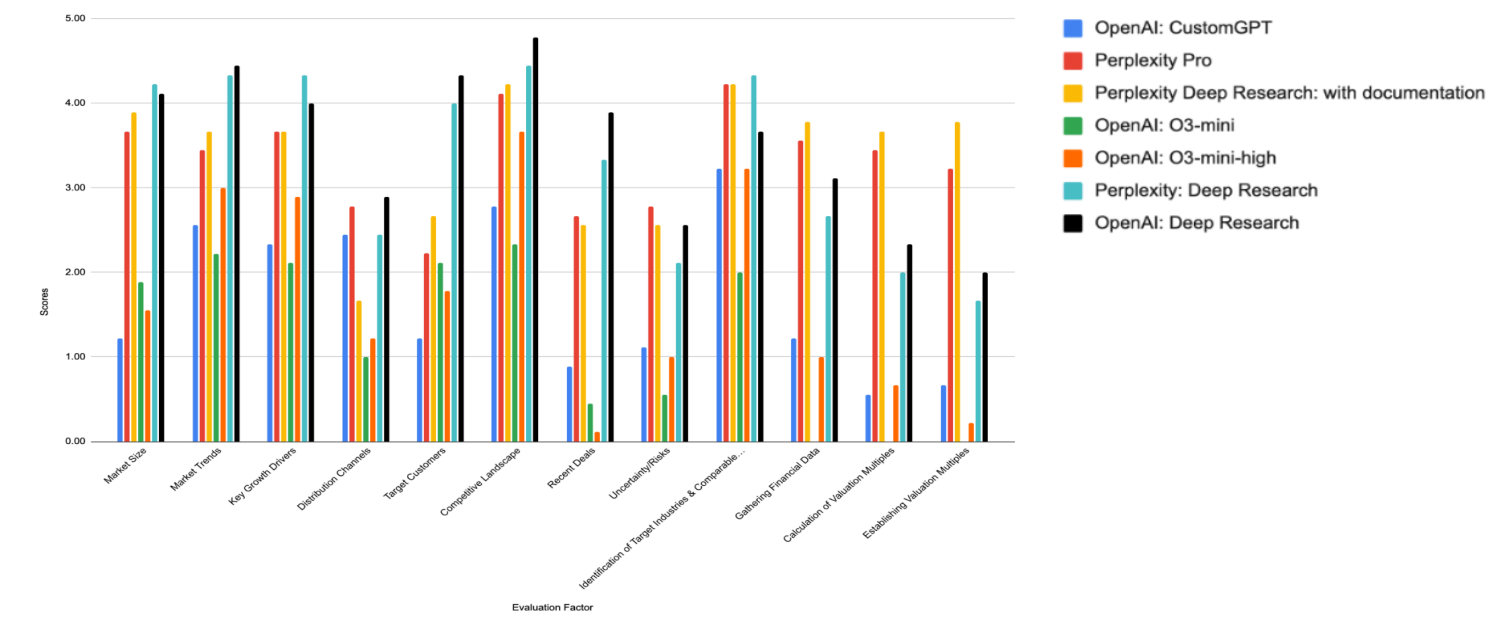
Each of these areas is rated on a scale from 0 to 5.

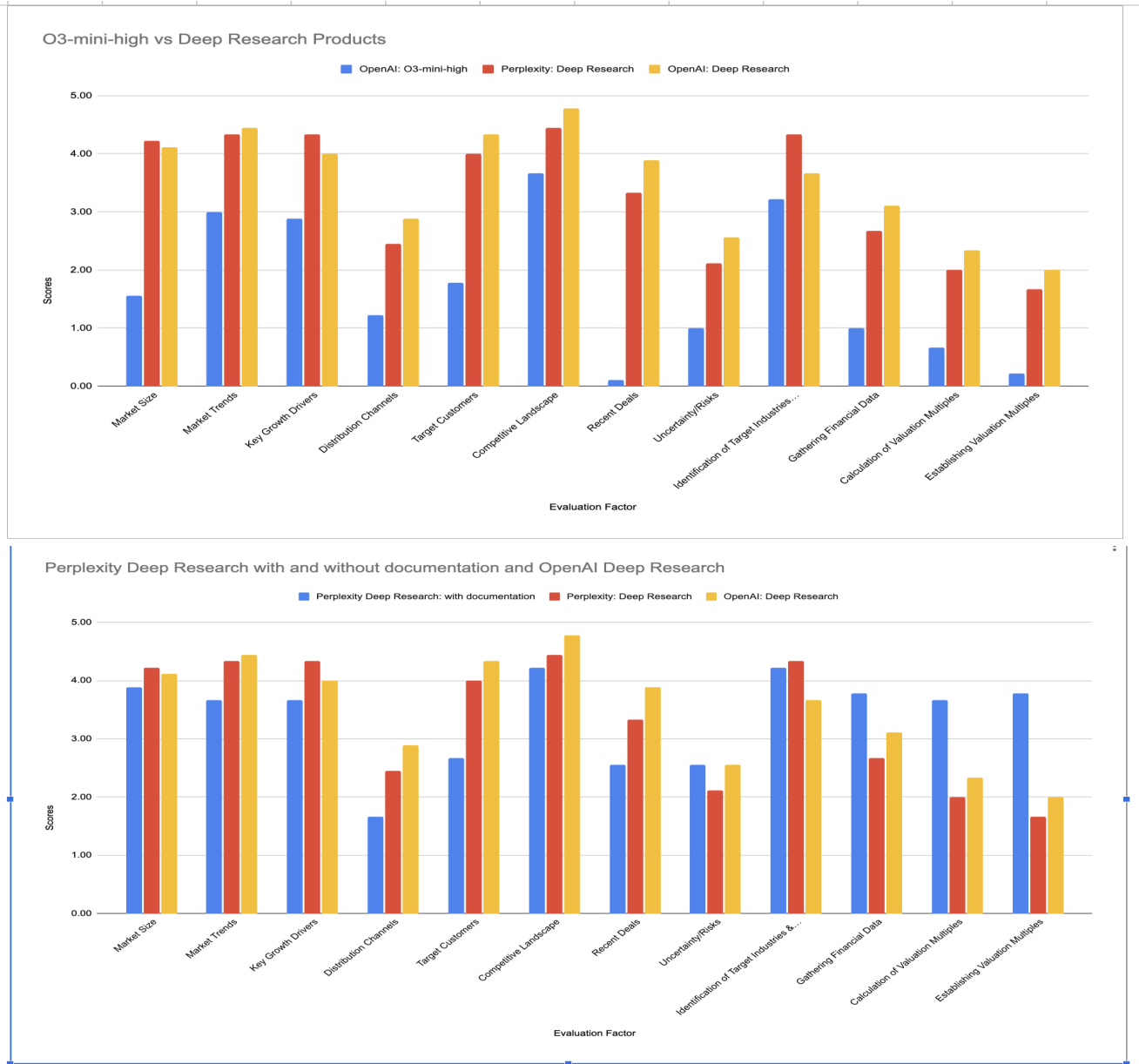
We use this three-level approach for both market reports and evaluations of similar companies. For example, when looking at a market report, we check if it covers important topics like market size, trends, key growth drivers, and risks. For each topic, we verify that the necessary details are there, that the analysis is thorough, and that it offers clear recommendations on what actions to take next.

We are not only evaluating the technical capabilities of these AI systems but also looking at how they can drive strategic insights for practical business applications. For instance, after comparing results from the different platforms, we analyze how well each system translates its findings into actionable recommendations. This means looking beyond raw data and asking questions like: How effectively does the report highlight market trends? Are the insights detailed enough to suggest specific business strategies? How well does the analysis tie back to the assumptions and data sources used? Answering these questions helps us understand whether a particular platform is capable of not just processing data, but also providing strategic guidance that businesses can rely on.

A detailed description of the evaluation framework can be found in the appendix ("Evaluation Framework") and a detailed evaluation of the different outputs can be found in the appendix ("Evaluation")

Results





What does the Evaluation Data Indicate?

OpenAI’s Deep Research consistently outperforms other models in categories like *Market Size* (4.5), *Key Growth Drivers* (4.3), and *Competitive Landscape* (4.4). Perplexity Deep Research with documentation follows closely, especially for *Market Trends* (4.2) and *Recent Deals* (4.1). In contrast, O3-mini-high scores lower across most categories, hovering around 3.0 to 3.3 in areas like *Distribution Channels* (3.1) and *Target Customers* (3.2).

These patterns suggest a notable gap between advanced “Deep Research” tools and simpler solutions. OpenAI’s Deep Research also excels in *Uncertainty/Risks* assessment, scoring 4.2 compared to O3-mini-high’s 2.9. This difference indicates that deeper, context-rich models handle risk evaluations with greater thoroughness, likely because they synthesize multiple data points into actionable insights.

Meanwhile, Perplexity Deep Research with documentation demonstrates a strong balance of completeness and clarity, especially in *Market Size* (4.3) and *Consumer Dynamics* (4.2). It’s added documentation appears to enhance both the comprehensiveness and transparency of results. Regular Perplexity Deep Research, by comparison, lags slightly in categories like *Market Fundamentals* (3.9) and *Implementation Methodology* (3.8). These findings underscore the importance of clear, well-structured documentation to bolster analytical depth and user trust.

The second chart (“03-mini-high vs Deep Research Products”) reinforces these observations. It illustrates that 03-mini-high achieves an average rating of around 3.2 across all factors, while Perplexity Deep Research averages approximately 4.0. OpenAI’s Deep Research leads with an approximate 4.3 average. Specifically, in *Peer Group Selection*—listed under *Comparable Company Analysis*—OpenAI’s Deep Research attains 4.4, Perplexity Deep Research reaches 4.1, and 03-mini-high remains at 3.0. This discrepancy highlights the more advanced reasoning capabilities of Deep Research models when identifying and analyzing relevant peers.

The third chart compares Perplexity Deep Research with documentation, regular Perplexity Deep Research, and OpenAI Deep Research. Perplexity Deep Research with documentation shows a notable improvement in *Valuation Multiples* analysis (4.0 vs. 3.6 for the regular version). However, OpenAI Deep Research still scores highest in that category, at 4.3. This suggests that, while thorough documentation improves Perplexity’s performance, OpenAI’s solution remains slightly ahead in consistently delivering high-level insights.

Overall, the data indicates a clear trend: Deep Research variants—both from Perplexity and OpenAI—offer more robust analytical depth and strategic insight than simpler or less-documented versions. Documentation plays a key role in bridging the gap, as seen in the difference between Perplexity Deep Research with and without documentation. Yet, OpenAI’s Deep Research retains a lead in most metrics, often by a margin of 0.2 to 0.5 points. This advantage likely stems from its capacity to handle complex inputs and generate actionable guidance. In conclusion, we are highlighting the importance of advanced reasoning capabilities and detailed documentation in driving higher evaluation scores across multiple market analysis criteria.

Comparing Perplexity Deep Research with Documentation, Regular Perplexity Deep Research, and OpenAI’s Deep Research

Perplexity Deep Research with documentation improves its overall score by an average of 0.3 points compared to the regular version—particularly noticeable in *Risk Framework* (3.8 vs. 3.5) and *Distribution Channels* (4.0 vs. 3.6). OpenAI’s Deep Research remains ahead in most categories, scoring above 4.0 in *Competitive Landscape* and *Market Size*, while Perplexity variants hover around 3.8 to 4.2. The documentation-enhanced version excels in clarity and traceability, offering more transparent justifications for its findings. Meanwhile, OpenAI’s Deep Research consistently delivers stronger strategic insights and a more thorough approach to data validation.

Comparing OpenAI Deep Research, Perplexity Deep Research, and 03 Mini High

OpenAI Deep Research averages approximately 4.2 across all criteria, notably excelling in *Market Fundamentals* (4.4). Perplexity Deep Research follows with an overall average near 4.0, performing best in *Market Size* (4.2). In contrast, 03-mini-high lags behind at about 3.2 overall, falling short in areas like *Competitive Landscape* (3.0). While 03-mini-high offers faster processing, it sacrifices the depth and accuracy seen in Deep Research models. Consequently, for comprehensive market evaluations, OpenAI Deep Research delivers the highest scores, Perplexity Deep Research remains a strong alternative, while Perplexity Deep Research with custom instructions and curated market research documents performs best to identify public comparables and provide financial data with valuation multiples.

“So please do your research, find credible market analysis reports, and use CapIQ / PitchBook. And also use AI with custom instructions to reduce time you spend on EFPE assignments from 12 hours/week to 10 hours/week.

AI can find you some form of comps all day - it just hasn’t paid the Booth tuition yet to find the ideal EV/EBITDA ratio.”

Appendix

Evaluation Framework

Below is an expanded evaluation framework that builds on the previous checklist. For each section (or question group), evaluators would score three aspects on a 0–5 scale:

1. **Level 1: Completeness Verification**
 - Checks that the report contains all the necessary elements.
2. **Level 2: Analytical Depth Assessment**
 - Assesses how thoroughly the data, methodology, and underlying assumptions are analyzed.
3. **Level 3: Strategic Insight Validation**
 - Evaluates whether the report translates its findings into actionable, forward-thinking strategic insights.

The following guidelines outline the scoring criteria for each section in both the market report and the comparable companies evaluations.

Part 1: Market Report Evaluation

A. Market Size

- Level 1: Completeness Verification
 - *Criteria:* Does the report clearly quantify market size (e.g., revenue figures, volume, growth rates) and include all basic metrics?
 - *Scoring:*
 - 0: No data provided
 - 1–2: Minimal or incomplete data points
 - 3–5: Comprehensive data with key metrics clearly listed
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is there an explanation of the methodology (data sources, assumptions, historical context) used to derive the market size?
 - *Scoring:*
 - 0: No analysis or methodology discussed
 - 1–2: Basic or superficial analysis with minimal explanation
 - 3–5: In-depth analysis with transparent methods and supporting evidence
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report discuss the implications of the market size data? For example, does it forecast growth, identify potential market opportunities, or suggest strategic actions based on the data?
 - *Scoring:*
 - 0: No strategic recommendations
 - 1–2: Generic insights that are not well tied to the data
 - 3–5: Clear, actionable insights that inform strategy (e.g., market expansion, segmentation focus)

B. Market Trends

- Level 1: Completeness Verification
 - *Criteria:* Are the key trends (consumer behavior, regulatory changes, technological shifts) mentioned?
 - *Scoring:*
 - 0: Trends not mentioned
 - 1–2: Some trends listed but lacking breadth
 - 3–5: All relevant trends are identified
- Level 2: Analytical Depth Assessment
 - *Criteria:* Does the report support trend identification with data (charts, statistics, expert quotes) and explain why these trends are occurring?
 - *Scoring:*
 - 0: No supporting data or explanation
 - 1–2: Limited supporting evidence

- 3–5: Robust analysis with clear evidence and context
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report translate trends into strategic opportunities or risk mitigation plans (e.g., adjusting product portfolios, targeting new consumer segments)?
 - *Scoring:*
 - 0: No strategic discussion
 - 1–2: Insights are stated but are generic
 - 3–5: Insightful recommendations that align trends with actionable strategy

C. Key Growth Drivers

- Level 1: Completeness Verification
 - *Criteria:* Are all major drivers (e.g., demographic shifts, technological advances, economic factors) listed?
 - *Scoring:*
 - 0: Growth drivers not addressed
 - 1–2: Only a few drivers are mentioned
 - 3–5: All significant drivers are identified
- Level 2: Analytical Depth Assessment
 - *Criteria:* Does the analysis detail how each driver impacts the market (using data, case studies, or comparisons)?
 - *Scoring:*
 - 0: No analysis of drivers
 - 1–2: Superficial explanation
 - 3–5: Detailed analysis with supportive data and nuanced discussion
- Level 3: Strategic Insight Validation
 - *Criteria:* Are there recommendations or strategies that leverage these growth drivers (e.g., investment in innovation, market segmentation adjustments)?
 - *Scoring:*
 - 0: No strategic recommendations
 - 1–2: Generic strategies offered
 - 3–5: Innovative and clearly actionable strategies derived from driver analysis

D. Distribution Channels

- Level 1: Completeness Verification
 - *Criteria:* Does the report list the various channels (e.g., retail, online, direct-to-consumer)?
 - *Scoring:*
 - 0: Channels not mentioned
 - 1–2: Only one or two channels mentioned, lacking comprehensiveness
 - 3–5: All relevant channels are identified
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is there analysis on how each channel performs, including strengths and weaknesses?
 - *Scoring:*
 - 0: No analysis
 - 1–2: Minimal analysis provided
 - 3–5: Detailed performance metrics and channel-specific evaluations are provided
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report offer strategic recommendations based on channel analysis (e.g., optimizing underperforming channels, exploring new digital avenues)?
 - *Scoring:*
 - 0: No strategic insight
 - 1–2: Recommendations are basic or generic
 - 3–5: Strategic insights are actionable and well-integrated with channel analysis

E. Target Customers

- Level 1: Completeness Verification

- *Criteria:* Does the report clearly define target customer segments (demographics, psychographics, buying behavior)?
- *Scoring:*
 - 0: No customer segmentation provided
 - 1–2: Only a broad or vague description
 - 3–5: Detailed segmentation with clear criteria
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is there an analysis of customer needs, pain points, and behaviors? (e.g., buyer personas, journey maps)
 - *Scoring:*
 - 0: No analysis
 - 1–2: Limited insight into customer behavior
 - 3–5: Thorough analysis with qualitative and quantitative support
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report draw actionable insights (e.g., tailoring messaging, product adjustments) based on customer analysis?
 - *Scoring:*
 - 0: No actionable insights
 - 1–2: Insights are superficial or generic
 - 3–5: Deep, strategic recommendations that are clearly informed by the customer data

F. Competitive Landscape

- Level 1: Completeness Verification
 - *Criteria:* Are the key competitors and market players identified?
 - *Scoring:*
 - 0: No competitors mentioned
 - 1–2: Only a few competitors noted without a full picture
 - 3–5: Comprehensive list of all major players
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is there an analysis of competitor strategies, market shares, strengths/weaknesses, and positioning?
 - *Scoring:*
 - 0: No analytical discussion
 - 1–2: Basic comparative metrics provided
 - 3–5: In-depth competitive analysis with clear comparisons
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report offer strategic implications from the competitive analysis (e.g., areas of differentiation, potential partnerships, competitive threats)?
 - *Scoring:*
 - 0: No strategic recommendations
 - 1–2: Minimal strategic insight
 - 3–5: Well-articulated strategic recommendations that directly address competitive dynamics

G. Recent Deals

- Level 1: Completeness Verification
 - *Criteria:* Are recent deals, mergers, or partnerships listed with basic details (deal value, date, parties involved)?
 - *Scoring:*
 - 0: No deals mentioned
 - 1–2: Only a couple of deals mentioned superficially
 - 3–5: Complete list of significant recent deals
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is there an analysis of the implications of these deals (e.g., market consolidation, strategic fit)?
 - *Scoring:*
 - 0: No analysis provided

- 1–2: Limited discussion on deal implications
 - 3–5: Detailed analysis of deal impact on market dynamics
- Level 3: Strategic Insight Validation
 - *Criteria:* Are there strategic recommendations or forecasts based on deal activity (e.g., predicting future consolidation, suggesting acquisition targets)?
 - *Scoring:*
 - 0: No strategic insights
 - 1–2: Generic recommendations
 - 3–5: Strategic, forward-looking insights that are actionable

H. Uncertainty/Risks

- Level 1: Completeness Verification
 - *Criteria:* Does the report list the major risks and uncertainties (regulatory, economic, supply chain, etc.)?
 - *Scoring:*
 - 0: Risks not mentioned
 - 1–2: Only a few risks noted without explanation
 - 3–5: Comprehensive risk factors are listed
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is there a discussion on the potential impact and likelihood of these risks, with supporting evidence or data?
 - *Scoring:*
 - 0: No risk analysis
 - 1–2: Minimal risk analysis provided
 - 3–5: In-depth risk analysis, with quantification or qualitative impact assessment
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report offer mitigation strategies or contingency plans for these risks?
 - *Scoring:*
 - 0: No strategic recommendations for risk management
 - 1–2: Generic or non-specific recommendations
 - 3–5: Clear, actionable risk mitigation strategies that are directly tied to the identified uncertainties

Part 2: Comparable Companies Evaluation

I. Identification of Target Industries & Comparable Companies

- Level 1: Completeness Verification
 - *Criteria:* Are the target industries and a list of comparable companies clearly identified?
 - *Scoring:*
 - 0: Not addressed
 - 1–2: Partial identification with unclear criteria
 - 3–5: Complete and clearly defined list with justification
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is the methodology for selecting these companies explained (criteria such as industry, size, geographic relevance)?
 - *Scoring:*
 - 0: No methodological explanation
 - 1–2: Limited explanation
 - 3–5: Detailed, transparent methodology with rationale
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the analysis explain why these companies are true comparables and discuss competitive positioning relative to the target company?
 - *Scoring:*
 - 0: No strategic context
 - 1–2: Generic comments
 - 3–5: In-depth strategic comparison and actionable insights for valuation

J. Gathering Financial Data

- Level 1: Completeness Verification
 - *Criteria:* Are all necessary financial data points gathered from credible sources?
 - *Scoring:*
 - 0: Incomplete or missing data
 - 1–2: Data gathered but with gaps or questionable sources
 - 3–5: Complete and credible financial data for all comparables
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is there an explanation of how the data was standardized or adjusted (e.g., normalization, one-time events) and a discussion of data quality?
 - *Scoring:*
 - 0: No discussion of methodology
 - 1–2: Limited analytical detail
 - 3–5: Thorough explanation with clear methodology
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report draw strategic conclusions from the data quality or data trends (e.g., identification of undervalued segments, market consistency issues)?
 - *Scoring:*
 - 0: No strategic insight derived
 - 1–2: Superficial insights
 - 3–5: Strategic insights clearly integrated with financial data analysis

K. Calculation of Valuation Multiples

- Level 1: Completeness Verification
 - *Criteria:* Are key valuation multiples (e.g., P/E, EV/EBITDA) calculated for each comparable?
 - *Scoring:*
 - 0: Multiples not calculated
 - 1–2: Some multiples calculated but not for all comparables
 - 3–5: All relevant multiples calculated for all companies
- Level 2: Analytical Depth Assessment
 - *Criteria:* Is the calculation methodology explained (e.g., adjustments for outliers, normalization)?
 - *Scoring:*
 - 0: No explanation
 - 1–2: Brief or unclear explanation
 - 3–5: Detailed, transparent methodology provided
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report compare the calculated multiples to industry norms and draw strategic conclusions (e.g., identifying mispriced companies, investment opportunities)?
 - *Scoring:*
 - 0: No strategic context provided
 - 1–2: Limited or generic insights
 - 3–5: Actionable strategic insights clearly linked to the multiples analysis

L. Establishing Valuation Multiples

- Level 1: Completeness Verification
 - *Criteria:* Is there a derived range of valuation multiples presented for the sector or target companies?
 - *Scoring:*
 - 0: No range provided
 - 1–2: Range is incomplete or based on limited data
 - 3–5: Complete and clearly defined valuation range
- Level 2: Analytical Depth Assessment
 - *Criteria:* Does the report discuss outliers, justify the range, and compare it to historical or peer benchmarks?
 - *Scoring:*
 - 0: No analysis

- 1–2: Basic discussion with little detail
 - 3–5: Detailed analysis with clear justifications and comparisons
- Level 3: Strategic Insight Validation
 - *Criteria:* Does the report provide strategic recommendations based on the valuation range (e.g., suggesting a target multiple for valuation negotiations, identifying undervalued peers)?
 - *Scoring:*
 - 0: No strategic recommendation
 - 1–2: Generic insights
 - 3–5: Deep, actionable strategic recommendations that are data-driven

Implementation & Aggregation

For each question or evaluation item above, evaluators assign a score from 0 to 5 for each level (completeness, depth, insight). These scores are then aggregated to provide an overall evaluation score.

This multi-level approach mirrors GAIA's benchmark philosophy by not only checking whether the information is present (Level 1) but also examining the quality of analysis (Level 2) and the strategic value of the insights offered (Level 3).

Each score is accompanied by a brief “reasoning trace” so that improvements can be clearly pinpointed. This ensures transparency and facilitates continuous improvement in report quality.

Evaluation Matrix

Note: Maximum score for any section is 5.

| Section | OpenAI: CustomGPT | Perplexity Pro | Perplexity Deep Research: with documentation | OpenAI: O3-mini | OpenAI: O3-mini-high | Perplexity: Deep Research | OpenAI: Deep Research |
|--|-------------------|----------------|--|-----------------|----------------------|---------------------------|-----------------------|
| Market Size | 1.22 | 3.67 | 3.89 | 1.89 | 1.56 | 4.22 | 4.11 |
| Market Trends | 2.56 | 3.44 | 3.67 | 2.22 | 3.00 | 4.33 | 4.44 |
| Key Growth Drivers | 2.33 | 3.67 | 3.67 | 2.11 | 2.89 | 4.33 | 4.00 |
| Distribution Channels | 2.44 | 2.78 | 1.67 | 1.00 | 1.22 | 2.44 | 2.89 |
| Target Customers | 1.22 | 2.22 | 2.67 | 2.11 | 1.78 | 4.00 | 4.33 |
| Competitive Landscape | 2.78 | 4.11 | 4.22 | 2.33 | 3.67 | 4.44 | 4.78 |
| Recent Deals | 0.89 | 2.67 | 2.56 | 0.44 | 0.11 | 3.33 | 3.89 |
| Uncertainty/Risks | 1.11 | 2.78 | 2.56 | 0.56 | 1.00 | 2.11 | 2.56 |
| Identification of Target Industries & Comparable Companies | 3.22 | 4.22 | 4.22 | 2.00 | 3.22 | 4.33 | 3.67 |
| Gathering Financial Data | 1.22 | 3.56 | 3.78 | 0.00 | 1.00 | 2.67 | 3.11 |
| Calculation of Valuation Multiples | 0.56 | 3.44 | 3.67 | 0.00 | 0.67 | 2.00 | 2.33 |
| Establishing Valuation Multiples | 0.67 | 3.22 | 3.78 | 0.00 | 0.22 | 1.67 | 2.00 |