

# Multi Agent Whitepaper

*What I learned writing equity research on my own, building a ten-agent AI version of the same process, and running both on the same stocks*

Between June and October 2026 I wrote equity research reports on eighteen companies with help from an AI model, then built a second version of the process in which ten AI agents gather evidence, argue and rule on each call. I have run both on six companies. They gave the same rating on five of the six, caught mistakes the single process missed, and disagreed once: on Alphabet the agents called a SELL where I had a HOLD, off the same closing price. On October 17 I changed the rating bands and re-priced all six agent runs to the October 16 close without re-running their debates. This paper covers what I built, what it cost, what went wrong, and what I will check when the calls come due in 2027. It replaces the 10-page edition of October 11, which had replaced the 37-page version of September 21.

**Vidwaan Singhania** · BCom student, Gustavson School of Business, University of Victoria, graduating 2027 · Code:

[github.com/Vidwaansinghania/Multi-Agent-Equity-Research](https://github.com/Vidwaansinghania/Multi-Agent-Equity-Research) · Feedback: [linkedin.com/in/vidwaansinghania](https://www.linkedin.com/in/vidwaansinghania)

Current holdings: [blossomsocial.com/users/Vidwaan-Singhania\\_d7oVAMV5EWqGMPon](https://blossomsocial.com/users/Vidwaan-Singhania_d7oVAMV5EWqGMPon)

A student project, written for learning. Nothing in it is investment advice. I own some of the stocks discussed; the disclosure is on the last page.

"Method 2" is the ten-agent process, called multi-agent equity research in the repository.

**18**

companies covered by Method 1, June to October 2026

**6**

Method 2 runs, each now with a Method 1 call

**5 of 6**

head-to-head tests with the same rating

**24 of 56**

bull and bear challenges the judge accepted in whole or part

**\$71–88**

API cost of a Method 2 run, first three runs counted

**0 of 30**

open calls scored; first review June 20, 2027

## The short version

- Method 1 and Method 2 agreed on the rating on Microsoft, Coca-Cola, Visa, Moog and Caterpillar, all HOLD, with targets 2.6% to 6.2% apart. On Alphabet they split: Method 2 said SELL at \$268.81 where Method 1 said HOLD at \$340.
- Agreement proves less than it looks like. I built both methods, they follow the same rules and the same AI does the reasoning in each.
- The agents were more useful as auditors than as second opinions. They found arithmetic errors in their own valuation, a beta taken from the wrong company, a currency cost that made one of my buy instructions lose money, and on Moog an earnings basis mismatch worth 7 points of expected return.
- On Coca-Cola and Visa the two methods reached HOLD from opposite relationships between the DCF and the target: Method 1 trusted the market multiple, Method 2 tied its target to the DCF.
- A Method 2 run cost about \$71 to \$88 at API prices, two to three and a half times a Method 1 report, and more than half of it went on agents re-reading their own instructions.
- On October 17 I moved the bands to BUY above +8% and SELL below –8%, and re-priced the six runs to the October 16 close. Coca-Cola turned SELL, mostly because the ten-year yield rose 41 basis points. The debates were not re-run, so those figures are updates, not new judgments.
- No open call can be checked against the market before June 2027, so I cannot yet say whether either method is any good.

## 1 Overview, aims and background

I'm a business student at the University of Victoria. In 2025 I was one of four UVic students on the CFA Institute Research Challenge team that wrote a 20-page report on Air Canada. I wanted to keep doing that kind of work on companies I chose, at a standard close to what bank research departments publish, and to find out how far AI tools could take a student working alone. Method 1 is that work: me and Claude (Anthropic's AI model) writing a report section by section inside Obsidian, a note-taking app. In August I built Method 2, in which ten Claude agents split the job: four gather evidence, one values the company, two argue for and against that valuation, a judge makes the call, one checks it against my small portfolio and one writes the report. Then I ran both on the same companies and compared them.

I set four aims: learn how a report is built by writing real ones with every number traced to a filing; make the process repeatable so reports can be compared; test whether agents that argue catch things a single line of reasoning misses; and record every call so it can be checked a year later. Status in October 2026: eighteen reports written; both processes fixed in writing; errors caught but one disagreement only; scoring impossible before June 2027. The paper is not investment advice and does not claim ten agents beat one analyst. Six runs cannot show that.

**About me.** I graduate from Gustavson in 2027 and am on exchange at SKEMA Business School in France this term. I sat the CFA Level 1 exam in August 2026. I've done case competitions, a consulting internship and the Research Challenge, but I've never worked in an equity research department, so my standards come from reading published research, not from training on the job. I invest a small amount in Canadian dollars through tax-free savings accounts (TFSAs). That matters here twice: Method 2's risk agent checks every call against what I hold, and I own some of the companies I write about, so every report states my position.

| Term                  | What it means here                                                                                                                                                                                                                                                                                              |
|-----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DCF, reverse DCF      | Discounted cash flow: forecast free cash flow and discount it to today. The reverse version asks what growth the share price assumes.                                                                                                                                                                           |
| Cost of capital, beta | The DCF discount rate, built from interest rates, the market risk premium and beta. Beta is how much a stock tends to move with the market; 0.5 means about half as much. The Blume adjustment pulls a measured beta part of the way toward 1.0.                                                                |
| Target, rating        | The probability-weighted twelve-month price across bull, base and bear scenarios. Expected return is the move to target plus dividends. Since October 17, BUY is above +8%, SELL is below -8% and HOLD is between; every call before that date used $\pm 15\%$ . A call within 5 points of an edge is marginal. |
| Alpha, agent, tokens  | Alpha is a stock's return minus the index's over the same dates. An agent is a separate Claude session with one role, its own instructions and its own files. Tokens are what models are billed in; "effective tokens" weight each kind by price so runs compare.                                               |

**What I built on.** TradingAgents (Xiao, Sun, Luo and Wang, 2024) models a trading firm with language-model agents in analyst, bull, bear, trader and risk roles and reports backtests that beat baselines. I took its roles, its debate and its log of calls. It makes trading decisions judged by backtest; Method 2 makes twelve-month BUY, HOLD or SELL calls with a report and model, judged against the index a year later, beside a person-led process. My report standard comes from an RBC Capital Markets report on Apple and the Research Challenge format. As far as I know, scoring a person-led and a multi-agent process side by side on the same stocks is unusual; I'd like to hear about similar work.

**Tools.** Claude Code runs Claude against my files and keeps the session logs the cost figures come from. Method 1 and the Method 2 stages from valuation onward use Claude Opus (Opus 5, and Opus 5.5 on Method 2's decision stages from September 26); the four evidence agents use Claude Sonnet 5. From October 12 the standard for Method 1 is Claude Sonnet 5.5 at high reasoning effort, and the October 17 revision agents ran on it too. I name the versions because every claim about cost, debate behaviour and what the judge accepted is a claim about particular models in a particular harness. Python builds the models and the PDF, Word and Excel files. Data comes from SEC EDGAR (which refuses requests without a name and email), Alpha Vantage on keys capped at 25 calls a day, FRED Treasury yields, stockanalysis.com, and since September 24 a temporary Refinitiv (LSEG) licence that runs to about January 2027. Claude is the only tool that produces analysis; a number that comes from the model is marked as an estimate.

## 2 Method 1: writing reports on my own

One Claude session works through a company one section at a time into the same template. I wrote the first report in mid-June 2026.

The bar was a checklist built from the RBC Apple report and my team's Air Canada report: tone, citations, metric definitions, rating language, valuation method, risk layout and four disclosure blocks. A report is not finished until it passes. Financials come from SEC filings; every figure goes into a working-notes file with its source first. Each report has ten sections (overview, thesis, industry, financials, valuation, risks, catalysts, scenarios and recommendation, appendices) and builds into a PDF, a Word file and an Excel model. The DCF is the main method with comparable companies as the cross-check; real costs are charged, including stock-based pay for software; bull, base and bear get probabilities that sum to one.

| Company        | Called | Call | Price    | Target  | What decided it                                                                                                                                                   |
|----------------|--------|------|----------|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ServiceNow     | Jun 20 | BUY  | \$95.12  | \$138   | Down 52% on fears for seat-based pricing when over half of new contract value was already usage-based; a DCF charging stock pay as cash still gave \$134.         |
| IBM            | Aug 8  | HOLD | \$237.28 | \$235   | A record 25% one-day fall removed the overvaluation without making it cheap; DCF counts about \$22B of acquisitions as a cost of growth.                          |
| IREN           | Aug 8  | HOLD | \$41.23  | \$44    | About \$29M of capital per megawatt earning about \$8.3M a year; the first five-year GPU cycle barely pays back, the second rewards shareholders.                 |
| Mogotes Metals | Aug 8  | BUY  | C\$0.64  | C\$1.05 | No revenue or defined resource; sum of the parts and deal prices (Rio Tinto bought about 5% at C\$0.70), with a 25% chance of losing two thirds.                  |
| Apple          | Aug 8  | HOLD | \$313.33 | \$270   | About \$0.11 of EPS came from a one-off \$2.19B tariff refund, so a record quarter was a two-cent beat; 1.4 points from SELL.                                     |
| Alphabet       | Aug 9  | HOLD | \$354.30 | \$330   | \$6.26 of the quarter's \$9.11 EPS was an unrealised investment gain; 24.0x 2027 earnings, not the 17.8x a screener shows. Refreshed Sep 20 at \$349.54 to \$340. |
| Coca-Cola      | Aug 10 | HOLD | \$87.05  | \$82    | DCF \$67; the reverse DCF needs 7.88% free cash flow growth for a decade; a \$20.5B IRS case sits outside guidance. Refreshed Aug 21 at \$91.10.                  |
| Microsoft      | Aug 11 | HOLD | \$506.06 | \$500   | Reported free cash flow of \$67.0B falls to \$26.8B counting unpaid capital spending and lease-funded data centres.                                               |
| Costco         | Aug 16 | HOLD | \$961.10 | \$923   | 52.8x earnings; the reverse DCF needs 15.6% free cash flow growth for ten years against a 9.9% record.                                                            |
| Aluula         | Sep 23 | HOLD | C\$3.95  | C\$3.50 | Initiation on a TSX Venture composites maker; expected return -11.4%. No Method 2 run.                                                                            |
| Visa           | Sep 24 | HOLD | \$361.52 | \$395   | DCF \$341 below the price; target from a held 24x multiple; +10.5%, marginal toward BUY.                                                                          |
| Webull         | Oct 7  | HOLD | \$5.87   | \$5.50  | Initiated the day a House committee report called its China ties a national security risk; low conviction. No Method 2 run.                                       |
| Waste Mgmt     | Oct 8  | HOLD | \$208.94 | \$220   | Initiation, moderate conviction. No Method 2 run.                                                                                                                 |

| Company         | Called | Call | Price    | Target | What decided it                                                                                                                                                                                                            |
|-----------------|--------|------|----------|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Caterpillar     | Oct 8  | HOLD | \$813.83 | \$730  | DCF \$391, under half the price; \$72.1B backlog and data-centre power demand; low conviction, 5.5 points from SELL. Oct 17 refresh at \$730.76: HOLD, \$730, +0.8%.                                                       |
| Rockwell        | Oct 12 | HOLD | \$432.59 | \$420  | DCF \$199; the price implies a 6.8% discount rate against the model's 10.83%; -1.5% expected return, 13.5 points from SELL. No Method 2 run.                                                                               |
| ATS             | Oct 12 | BUY  | C\$25.19 | C\$29  | Initiated as a BUY: +15.6%, 0.6 points past the line, low conviction. Repriced Oct 17 at C\$25.78: C\$29, +13.0%. HOLD on the first bands that day, BUY on the final +8% line, and the node now says BUY. No Method 2 run. |
| Parker-Hannifin | Oct 12 | HOLD | \$962.56 | \$950  | Modelled as if the \$9.25B Filtration Group purchase, closed Aug 13, were already in the numbers; DCF \$452; -0.5%. No Method 2 run.                                                                                       |
| Moog            | Oct 12 | HOLD | \$363.09 | \$360  | A \$30M tariff refund sits in the June quarter's EPS; DCF \$208; -0.2%. Written six days after the Method 2 run, without opening it. Oct 17 refresh at \$368.60: HOLD, \$360, -1.7%.                                       |

Dates 2026; prices are each report's reference price. Mogotes and Aluula trade in Canadian dollars on the TSX Venture Exchange and ATS on the Toronto Stock Exchange. The four October 12 prices are October 9 closes. Calls are as first issued. On October 17 three nodes (ATS, Caterpillar, Moog) were refreshed to the October 16 close; the rest were not.

**What it taught me.** Reported cash flow and earnings can be the wrong number: Microsoft's \$19.8B of unpaid capital spending and \$20.4B of lease growth, Alphabet's \$6.26 unrealised gain and Apple's tariff refund all needed adjusting, and when a company puts a number on a one-off I value the shares on what's left. Moog's June quarter was the same pattern: a \$30M tariff refund worth about \$0.70 a share, so guided EPS of \$11.65 is about \$10.95 without it. A DCF tells you what the price assumes: on Costco, Coca-Cola, Microsoft, Alphabet and Caterpillar it came out far below the price, so I show it as "what the price requires" and build the twelve-month target on earnings multiples. On August 9 I cut the scale to BUY, HOLD and SELL and added a rule to say how close a call is to an edge and name the one assumption that would flip it (Coca-Cola's SELL starts above \$99.03). My worst mistakes were not in the models: three reports said I held no position in companies I owned, and one had no ownership disclosure. Every report now ends with four fixed disclosure blocks and the ownership line is copied from my holdings file.

**Where it fell short.** Nothing argued with me, no call would ever be checked, each stock was judged alone (so the cost of converting Canadian to US dollars never entered), and every company folder held its own thousand-line build script, so a fix reached one report. A second method had to answer whether agents that argue catch what a single process can't.

## 3 Method 2: ten agents

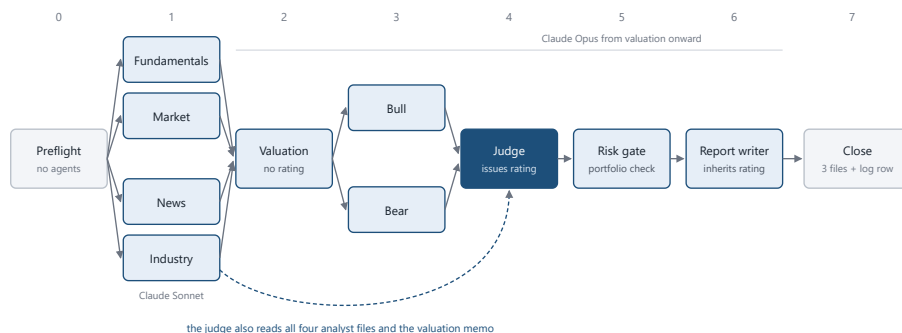
On August 17 I read the code for TradingAgents (Tauric Research, Apache-2.0) and rebuilt the process around Method 1's rules. Five things differ from their design.

| Their design                                                               | What I did instead                                                                                      | Why                                                                    |
|----------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|
| Five-level rating pulled from text by pattern matching, defaulting to Hold | Three ratings in a fixed field; any other value stops the run                                           | A parsing failure should show as an error, not as a call no agent made |
| The judge reads only the debate                                            | The judge also reads the four evidence files and the valuation memo                                     | Otherwise the more persuasive agent wins whatever the evidence says    |
| Bull and bear argue in general terms                                       | Both attack a written base case, name the assumptions they dispute and compute the effect on the target | An argument against a number can be checked                            |
| A social-media sentiment analyst                                           | Dropped                                                                                                 | Noise for a twelve-month call                                          |
| Three risk agents with different temperaments                              | One risk agent measuring concentration, correlation and currency cost; it can't change the rating       | I wanted risk measured, not argued                                     |

I kept their best idea unchanged: a log of every call, never edited afterwards, reviewed later against what the market did.

### Ten agents in eight stages

Stages run left to right. Stacked boxes run at the same time and can't see each other's work.



**Stages.** Preflight (no agents) sets one reference price, downloads the SEC filing index and ten years of financial data, builds the statements, measures beta over several windows with standard errors, takes a Refinitiv snapshot while the licence lasts, and records my position. Four evidence agents then work in parallel; only the market agent may use the market-data API, capped at 75 calls. The valuation agent builds the DCF, comparables, three scenarios and a sensitivity table, and may not give a rating. Bull and bear each attack that base case with numbers.

The judge rules, records which stage's finding decided the call and says how close it is to the edge of its band. The risk agent can disagree in writing but can't overrule. The report writer copies the rating and target without recalculating. The close builds a PDF, a 14-sheet Excel model and a Word file of every agent's full output, checks the decision file against its contract, writes the comparison with Method 1 and logs the call.

**Rules that matter most.** No agent may open my Method 1 report on the same company until the judge has committed, or ten agents would read my conclusion and agree with it. No number is typed twice: the valuation lives in one Python file that the model and the PDF both read, and the rating and target live in one decision file. The DCF is built the same way every time (ten forecast years, mid-year discounting, a terminal value tied to growth and return on capital, plus a reverse DCF), because two runs that discount differently can't be compared. From Method 1 it keeps the three ratings, ownership stated as a fact, price and quality as separate questions, and the checklist.

## 4 How I compared the two methods

|                   | Method 1                                                               | Method 2                                                                          |
|-------------------|------------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| Reasoning         | Me, with one Claude session                                            | Ten Claude agents with separate roles; I design the process and review the output |
| Checks            | My review against a checklist                                          | Bull and bear attack the valuation; a judge rules on the evidence                 |
| Valuation         | DCF plus comparables; the target usually comes from earnings multiples | Fixed DCF build plus comparables; the judge decides which to rely on              |
| Portfolio check   | None                                                                   | Risk agent: concentration, correlation, currency cost                             |
| Output            | PDF, Word file, Excel model                                            | PDF, 14-sheet Excel model, Word file of every agent's output                      |
| Cost, active time | \$27 to \$35 on the two test companies, 0.7 to 1.2 hours               | \$71 to \$88, about 2.3 to 2.6 hours                                              |

**EQUITY RESEARCH** | Update | Consumer staples — non-alcoholic beverages | August 21, 2026

**The Coca-Cola Company**  
NYSE: KO - USD 91.10 (Aug 21, 2026 close)

**HOLD**  
12-month price target **USD 82**

**A very good quarter at a price that needs everything to go right**

**My view**  
The June quarter was the best Coca-Cola has printed in years: organic revenue up 6%, unit case volume up 5% led by India, China, the United States and Brazil, comparable operating margin 35.6% against 34.7%, value share gained in four of five segments, and full-year guidance raised on July 28. None of that is in dispute. **The price is.** At \$91.10 the shares trade at 25.8x 2027 consensus, having run 39% off the \$65.35 low on a 1.4% upward revision to 2027 estimates, so almost the whole move is multiple rather than earnings. Three of the guided nine to ten points of earnings growth are currency; the company's own currency-neutral figure excluding divestitures is 7% to 8%. A discounted cash flow at a 7.18% cost of capital values the shares at \$67. This

|        | Bear | Base | Bull  | Target |
|--------|------|------|-------|--------|
| Value  | \$59 | \$85 | \$100 | \$82   |
| Return | -30% | -0%  | +17%  | -3%    |
| Weight | 25%  | 50%  | 25%   | —      |

| Key statistics |          |
|----------------|----------|
| Market cap     | \$392.0B |

Method 1, Coca-Cola, page 1 of 23: call, main points, scenario table, statistics.

**The Coca-Cola Company (NYSE: KO)**  
**HOLD, twelve-month target USD 77.45**

| Item                  | Figure                                                      |
|-----------------------|-------------------------------------------------------------|
| Rating                | HOLD                                                        |
| Reference price       | USD 91.10, close of 2026-08-21                              |
| Twelve-month target   | USD 77.45                                                   |
| Expected total return | -12.6%, including USD 2.16 of dividends                     |
| Conviction            | Low. The call is marginal                                   |
| Market capitalisation | USD 391.96bn                                                |
| Enterprise value      | USD 419.85bn                                                |
| Fiscal year end       | December 31. FY2025A closed 2025-12-31, reported 2026-02-20 |

Method 2, Coca-Cola, page 2 of 24 after a title page: the call and the figure it turns on.

**Questions.** Do the methods reach the same rating? Does Method 2 find mistakes Method 1 misses? Which stage decides? What does the extra work cost? Whose calls beat the index over twelve months (not answerable before 2027)? For each company I record the ratings, the target gap, the inputs that differ, every error found and whether the judge accepted it, the deciding stage, and tokens, dollars and time.

| Condition                                     | MSFT                                                                                                               | KO                    | GOOGL                                                  | V                                     | MOG.A                                                                                                | CAT                                    |
|-----------------------------------------------|--------------------------------------------------------------------------------------------------------------------|-----------------------|--------------------------------------------------------|---------------------------------------|------------------------------------------------------------------------------------------------------|----------------------------------------|
| Same reference price                          | No: \$506.06 against \$480.35, six days apart                                                                      | Yes, \$91.10          | Yes, \$349.54                                          | No: \$361.52 and \$367.38, 1.6% apart | No: \$381.04 and \$363.09, 4.9% apart                                                                | No: \$813.83 and \$796.18, a day apart |
| Method 2 agents kept from the Method 1 report | Yes                                                                                                                | Yes                   | Yes                                                    | Yes                                   | Yes                                                                                                  | Yes                                    |
| Me kept from both answers                     | No                                                                                                                 | No: I knew Method 1's | Partly: Method 2 ran first, but an August call existed | No: Method 1 ran two days earlier     | No: I knew Method 2's answer, but the Method 1 session did not open the run until its call was final | No: Method 1 ran a day earlier         |
| Same DCF build                                | No in every test: Method 2 uses its own fixed build. Same rating bands and same AI doing the reasoning in all six. |                       |                                                        |                                       |                                                                                                      |                                        |

**How calls will be scored.** Each call is reviewed twelve months after it was made. Return is price change plus dividends; alpha is that minus the benchmark's total return over the same dates, the S&P 500 (SPY) for US stocks and the S&P/TSX Capped Composite (XIC.TO) for the two TSX Venture names. A BUY is right if the stock beat its benchmark, a SELL if it lagged, and a HOLD if alpha lands within 15 points either way, which I plan to apply strictly. A script now does the arithmetic on Yahoo adjusted closes; it still runs only when I run it. Each scored call gets a short written reflection that the judge reads on later runs. Unscored calls are never fed forward.

## 5 Six runs

| Company          | Method 1        | Method 2            | Targets apart | Decided by     | What the run showed                                                                                                                                            |
|------------------|-----------------|---------------------|---------------|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Microsoft Aug 17 | HOLD \$500      | HOLD \$487.18 +2.2% | 2.6%          | Market analyst | The bear found a sensitivity-table error that helped the bull and an EPS share-count slip; the risk agent found the currency cost exceeds the expected return. |
| Coca-Cola Aug 21 | HOLD \$82 -7.6% | HOLD \$77.45 -12.6% | 5.6%          | Bull           | Same rating from opposite sides of beta: 0.558 against 0.45. Method 1's beta in Method 2's model gives SELL.                                                   |

| Company              | Method 1                       | Method 2                             | Targets apart | Decided by     | What the run showed                                                                                                                                                                                                                                                                                                     |
|----------------------|--------------------------------|--------------------------------------|---------------|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Alphabet<br>Sep 19   | HOLD \$340<br>–2.5%            | SELL \$268.81<br>–22.8%              | 21%           | Market analyst | First disagreement: Method 2 built its own 2027 EPS 11.6% below consensus, then applied a multiple two points lower.                                                                                                                                                                                                    |
| Visa<br>Sep 26       | HOLD \$395<br>+10.5%, marginal | HOLD \$376.46<br>+3.3%, not marginal | 4.7%          | Valuation      | Same rating, different distance from BUY. The bear showed the bull's 26x was a trailing multiple applied to forward EPS.                                                                                                                                                                                                |
| Moog<br>Oct 6        | HOLD \$360<br>–0.2%            | HOLD \$343.86<br>–9.4%               | 4.7%          | Bull           | The valuation memo said SELL (–16.3%). The bull found it applied adjusted-EPS multiples to stricter EPS; the judge corrected it and the call moved to HOLD. Method 1, written six days later, also said HOLD; about half the 9-point gap in expected return is price, which fell 4.7% between the two reference closes. |
| Caterpillar<br>Oct 9 | HOLD \$730<br>–9.5%            | HOLD \$684.79<br>–13.2%, marginal    | 6.2%          | Valuation      | The gap is the earnings base: Method 1 capitalises FY2028E, Method 2 the twelve months after the target date. 1.8 points from SELL.                                                                                                                                                                                     |

Expected total returns are shown where each method published one. Moog Class A's street mean target is \$446.40, above both methods.

## October 17: new bands and re-priced runs

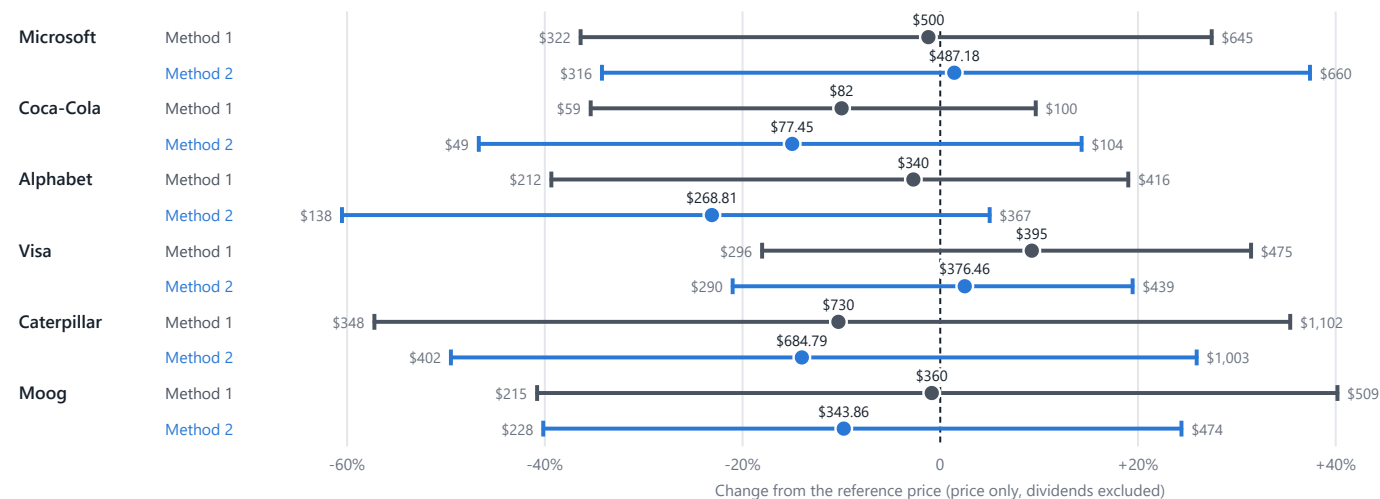
On October 17 I changed the rating bands twice that day, ending at BUY above +8%, HOLD from –8% to +8% and SELL below –8%. Calls made before then keep  $\pm 15\%$ . The runs were written to the first version (BUY at +15% or more, HOLD from –8% to +15%), and none of their ratings changes under the final one. I also had one Sonnet 5.5 agent update each closed run to the October 16 close, the ten-year yield (now 5.10%) and current consensus, and for this one pass I set aside the rule that a closed run is never edited. The debate and judge did not run again, so these are updates and not new judgments. The originals stay in each run's *superseded* folder and in the calls log, which got five new rows.

| Run         | First issued          | Revised Oct 17        | Price    | What moved it                                                                                                                                             |
|-------------|-----------------------|-----------------------|----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|
| Microsoft   | HOLD \$487.18, +2.2%  | HOLD \$486.63, –7.3%  | \$529.05 | The price rose 10% against a flat target; 0.7 points from SELL                                                                                            |
| Coca-Cola   | HOLD \$77.45, –12.6%  | SELL \$69.32, –23.9%  | \$93.90  | The 41 bp rise in the yield costs 8.6 points through the DCF-based target                                                                                 |
| Alphabet    | SELL \$268.81, –22.8% | SELL \$273.22, –19.5% | \$340.59 | Rating unchanged                                                                                                                                          |
| Visa        | HOLD \$376.46, +3.3%  | HOLD \$379.71, –4.9%  | \$402.61 | The price rose 9.6%                                                                                                                                       |
| Moog        | HOLD \$343.86, –9.4%  | HOLD \$339.97, –7.4%  | \$368.60 | 0.6 points from SELL                                                                                                                                      |
| Caterpillar | HOLD \$684.79, –13.2% | HOLD \$684.79, –5.4%  | \$730.76 | Target unchanged. A first pass that restated the bull multiple from the new price came out SELL; I reversed it because it derived an input from the price |

On the same October 16 close, Method 1 now stands at HOLD \$360 (–1.7%) on Moog against Method 2's –7.4%, and HOLD \$730 (+0.8%) on Caterpillar against –5.4%. Coca-Cola's Method 2 SELL faces a Method 1 HOLD that was not refreshed, so that gap is price and rate, not a difference of method.

## Scenario ranges and targets, both methods

Each line runs from the bear case to the bull case; the dot is the twelve-month target. Dollar labels are scenario values.



Sources: Method 1 reports and comparison files; Method 2 decision records, August 17 to October 9, 2026. Method 1's Microsoft scenarios use fiscal 2027 earnings and Method 2's fiscal 2028, so that pair is not built on identical inputs. Base-case markers from the earlier edition are left out.

On Microsoft the two sit within a few points everywhere but the bull case. On Coca-Cola Method 2 is lower everywhere except the bull case. On Alphabet the whole Method 2 line sits below Method 1's, which is what a rating disagreement looks like on this scale. On Visa the gap is in the bull and base legs; on Caterpillar Method 1's bull case reaches +35% where Method 2's reaches +26%, while its bear case is the harsher of the two. On Moog the two overlap closely, Method 1's range a little wider at both ends (\$215 to \$509 against \$228 to \$474).

---

## 6 What the runs showed

### Microsoft: errors, a currency cost and a discount-rate gap

The first full test ran six days after my report, and Microsoft's shares fell 5.1% in between, so part of the difference is price. Same rating, same moderate conviction; both found the DCF (\$342 and \$279.54) well below the price and both built the target from multiples, with capital spending as the swing factor. Method 2 gave three things Method 1 had no stage to produce. *The bear rebuilt the model in Python*, matched the base case to the cent, then found a sensitivity cell that applied half of a two-part change (correcting it raised the DCF range's top from \$311.12 to \$323.31, which hurt the bear's own side) and an EPS figure implying a share count used nowhere else. *The risk agent priced the currency*. My Canadian-to-US round trip costs about 3.77% on my own conversions against a +2.23% expected return. Bought at \$480.35, Microsoft needs \$494.58 within a year to break even, and a purchase only pays for the conversion below about \$473, so my Method 1 advice to buy below \$500 lost money at every price from \$473 to \$500 even if the call was exactly right. *The discount rate split us*: Method 2 used 9.22% on a beta of 1.10, the bear argued 9.84% on 1.25 and was rejected, and Method 1 had used 9.89%. We're 67 basis points apart and I haven't settled it. The judge moved the weights from 25/50/25 to 28/50/22 on 39 upward EPS revisions against 13 downward that the memo had recorded and not used.

### Coca-Cola: one number

I refreshed Method 1 that evening on the same \$91.10 close, so price, time and filings are identical. File timestamps show Method 1's valuation saved at 9:17 pm Pacific, Method 2 started a minute later, and no Method 2 agent had written anything when my refresh finished at 9:25 pm; API limits then stretched the run to August 23. Method 1 could not have copied Method 2, but I knew its answer first. Measured beta was 0.34. Method 1's Blume adjustment gave 0.558; the valuation agent reached 0.56. The memo defended it by saying Monster Beverage's 0.55 was the peer high, but the market analyst's own file put Monster at 0.52; the 0.55 belonged to a bottler the file had excluded. The bull caught it. The judge went further: Blume pulls toward 1.00, yet every non-bottler beverage peer measured 0.34 to 0.52, so for a company with a clear peer group the peer average is the better anchor. It used 0.45. That was worth \$8.39 of the \$77.45 target; all other corrections together were \$0.46. Method 1 refused its own \$67 DCF and took a 24x multiple; Method 2 made the DCF the target. Both agreed on the business and split on a number neither had measured more than once, so beta is now measured in preflight.

The bear also found the memo subtracted the \$6,514M IRS deposit from a starting figure that never included it (fixing it raised the target \$0.54, against the bear's own side), and the risk agent caught the judge's claim that currency cost exceeded any capturable return. Method 1 did better on segment detail, on noticing that nothing about the business changed between August 10 and 21 while the shares rose 4.7% on the multiple alone, and on naming what would change the call.

### Alphabet: the first disagreement

Method 2 ran first (September 19); my refresh followed the next day off the same \$349.54 close. It is closer to a blind test than the first two but not one, since my August HOLD existed. *They agree* that two thirds of the June quarter's \$9.11 EPS was a disclosed unrealised gain (so about \$2.85 of real earnings), reached by Method 2's market analyst before any agent could see my report, and that the September 2 ad-tech ruling, which rejected a forced sale of AdX, closed the break-up risk. My August note had that date wrong, reading the September 16 unsealing as the decision; the agents' news file had it right and I corrected my note. *They differ* in two compounding gaps. Method 1 takes consensus 2027 EPS of \$14.87; Method 2 built \$13.14 from a depreciation schedule (servers on six-year lives, structures on twenty) that pushes depreciation from 5.2% of revenue into double digits. That is 11.6% lower before any multiple. Then 23.0x against 21.0x, because the worst cash conversion and highest capital intensity in the peer group earn a discount. Method 2's SELL is not marginal: it reaches HOLD only at a 25.6x base multiple, close to its own 26.0x bull case. The bull's catch (a revenue cut applied to the wrong year, worth \$6.80) was adopted; the bear's catches on share count and terminal depreciation were adopted in part. *What settles it*: Alphabet's 2027 capital spending guidance, due late January or early February. Above \$280 billion supports Method 2's reading; at or below \$240 billion breaks it.

### Visa, Moog and Caterpillar

**Visa.** Same HOLD, different distance from BUY (+10.5% and marginal against +3.3% and not marginal). Earnings differ by 1% or less in every scenario, and the weights and 19.0x bear multiple match. The gap is multiples: Method 1 holds 24.0x base and 27.0x bull; Method 2 tied the base to its DCF rolled forward, 23.1x, and cut the bull from 26.0x to 25.0x after the bear showed the 25.6x to 26.9x history behind it was trailing P/E on GAAP EPS applied to forward economic EPS. This is the second case, after Coca-Cola, where the DCF came in below the price and the methods split: Method 1 trusts the multiple, Method 2 the DCF. If Visa trades near 24x forward earnings on September 26, 2027, Method 1's target was the better one; at 24.5x Method 2's own judge computed \$388.21, still HOLD.

**Moog.** I wrote the Method 1 report on October 12, six days after the run, and its session did not open the run until its own call was final. Both reached HOLD. The valuation memo called it a SELL at -16.3%, 1.3 points past the line. The bull's first attack decided the rating: the peer and history multiples were measured on adjusted EPS, but the memo applied them to a normalised EPS \$0.99 below that (base FY2028E \$12.44 against \$13.43), and the bear's own arithmetic conceded it. Restating onto one basis was worth \$25.00 on the target, enough to cross the band, and the judge moved five points of probability from bull to base. Stock at \$381.04 trades at 31.7x FY2027E adjusted EPS, the 95th percentile of its five-year range; Method 2's call flips to SELL at a 23.1x base multiple. Method 1 found the same one-off, a \$30M tariff refund worth about \$0.70 a share in the June quarter. It put 26.5x on FY2028E EPS of \$13.59 where Method 2's judge put 26.0x on adjusted FY2028E EPS, and its call sits further from the edge: SELL only at 18.6x. Most of the nine-point gap in expected return (-0.2% against -9.4%) is price and target: Moog closed at \$363.09 on October 9 against \$381.04 on October 5, and the targets are 4.7% apart. On October 17 both moved

to the same \$368.60 close: Method 1 at –1.7%, Method 2 at –7.4%, targets \$360 and \$339.97. I bought a small Class A position on October 6, the day of the run, and sold it four hours later; the run's record still shows it as open.

**Caterpillar.** Both methods find the DCF (\$391 and \$355) under half the price of \$796 to \$814. The bear showed the memo's 21.5x base multiple is a blend of two regimes: twelve quarter-ends averaged 21.5x, but the eight before September 2025 averaged 16.5x and the latest four 31.5x. The judge kept 21.5x because backlog of \$72.1B opens FY2027 at about 0.96 times prior sales against 0.43 to 0.54 in FY2023 to FY2025, and kept the 20.0% base margin because the first-half 21.7% incremental margin carried about \$871M of tariff and capacity-ramp cost. It did cut the bear multiple from 20.0x to 18.6x, the six-peer median. The bear's 17.2x base would give –21.5%, a SELL, so that is the rejected argument to check. First test: the Q3 margin excluding tariff recoveries, where below 19.2% flips the call. Method 1 and Method 2 date that report differently (about October 28 against November 4).

## The debate, in numbers

| Run         | Challenges | Accepted in whole or part | The accepted ones that mattered                                                                                                |
|-------------|------------|---------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| Microsoft   | 10         |                           | 6 Sensitivity cell, EPS share count, bull EPS above consensus, the 39 to 13 revisions (weights moved)                          |
| Coca-Cola   | 9          |                           | 4 Beta from the wrong company (+\$8.39), IRS double count (+\$0.54), tax rate by scenario, a \$400M fairlife charge            |
| Alphabet    | 12         |                           | 6 Revenue cut in the wrong year (+\$6.80), share count, terminal depreciation (about \$10), margin fade in part                |
| Visa        | 11         |                           | 4 Consensus growth haircut had no basis (+\$0.64), restructuring charge (–\$2.89), forward against trailing multiple (–\$4.39) |
| Moog        | 7          |                           | 3 Basis mismatch (+\$25.00), base multiple on consistent return on capital (+\$11.41, in part), weights moved off the bull     |
| Caterpillar | 7          |                           | 1 Bear multiple 20.0x to 18.6x (–\$7.56 on the target)                                                                         |
| Total       | 56         | 24                        |                                                                                                                                |

The accepted claims are mostly arithmetic, sourcing and basis errors, which can be checked against a file. The rejected ones are mostly forecasts, which can't: terminal growth, exit multiples for a year out, scenario weights. That split seems right to me, and it means the debate is better at auditing a model than at changing a view. Counted this way the stage that decided the call is split evenly across six runs: the market analyst twice, the bull twice, the valuation agent twice. The bull's finding moved Coca-Cola and Moog out of SELL-range numbers; on Alphabet the SELL came from the evidence stage. Six runs prove nothing, but a dozen should show whether the debate does real work. (The long edition lists the claims for the first three runs one by one; I recounted them here as 31 challenges, 16 accepted, where the earlier edition said 30.)

## 7 What it cost

| Company   | Work                     | Effective tokens | At API prices | Active time | Elapsed |
|-----------|--------------------------|------------------|---------------|-------------|---------|
| Microsoft | Method 1 report, Aug 11  | 6.9M             | \$35          | 1.2 h       | 18 h    |
| Microsoft | Method 2 run, Aug 17–18  | 18.8M            | \$74          | 2.6 h       | 12 h    |
| Coca-Cola | Method 1 report, Aug 10  | 5.3M             | \$27          | 0.7 h       | 19 h    |
| Coca-Cola | Method 1 refresh, Aug 21 | 2.3M             | \$11          | 0.3 h       | 0.3 h   |
| Coca-Cola | Method 2 run, Aug 21–23  | 21.2M            | \$88          | 2.6 h       | 34 h    |
| Alphabet  | Method 1 refresh, Sep 20 | 6.8M             | \$34          | 0.5 h       | 7 h     |
| Alphabet  | Method 2 run, Sep 19–20  | 21.5M            | \$71          | 2.3 h       | 15 h    |

Counted from session logs on September 17 and 20, each API response once. Effective tokens weight cache writes 2x, cache reads 0.1x and output 5x against fresh input; dollars use list prices of \$5 per million effective tokens for Opus and \$2 for Sonnet. I have not counted the Visa, Moog and Caterpillar runs or the October 17 revision agents. Method 1 rows cover only the session that wrote each report.

I ran everything on a Claude subscription, so the dollars are what the work would cost on the API, not what I paid. A Method 2 run costs two to three and a half times a Method 1 report and takes two to four times the active time. Most of Coca-Cola's 34 hours was waiting on API limits.

**Where it went.** On Microsoft the fundamentals and news agents made 66 and 90 tool calls, mostly hunting for filing addresses, and evidence gathering was over a third of the cost; filing lists and data are now fetched once, before any agent starts. On Coca-Cola, 85% of about 21M effective tokens sat inside agents, and by my audit's shares about 55% went on re-reading instructions and role briefs, 23% on reading files and running scripts, and under 0.5% on web searches and market data. My first audit said 39.6M tokens and was wrong, because two-part responses were counted twice; the shares survived the recount. The mechanism is that every tool call re-sends the whole conversation, so what an agent carries on every step costs far more than what it looks up once; the fundamentals agent spent 2.5M tokens over 77 calls rebuilding statements the data already fixed.

| Change                      | What it does, and the evidence                                                                                                                                                                                |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Preflight builds statements | A script extracts ten years of statements before any agent starts. On Coca-Cola data it matched the hand-built figures on every line checked except 2016 diluted EPS, two cents off from a later restatement. |
| Agents save as they go      | API limits killed three Coca-Cola stages; the two holding work in memory restarted from zero (about 1.8M tokens), the one saving each section resumed.                                                        |

| Change            | What it does, and the evidence                                                                                                                                             |
|-------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Beta in preflight | Weekly regressions over one, two, three and five years with standard errors, for the company and its peers. It settled the Alphabet argument on the first run that had it. |

None removes an agent or lets the judge skip the evidence. Since then I have added a contract check that refuses to build files when a decision record breaks its schema, moved the market agent's consensus pull to a cited Alpha Vantage source, and added the Refinitiv snapshot to preflight. Alpha Vantage calls per run so far: 1 on Coca-Cola, 6 on Alphabet, 5 on Visa, 1 on Moog and 0 on Caterpillar.

## 8 What the tests can't tell me, and what went wrong

**The two methods aren't independent.** I built both. Method 2 inherits Method 1's bands, rules, sections and checklist, I wrote the agents' instructions after writing nine reports, and Claude reasons in both. The anchoring rule stops an agent reading my conclusion but not my conclusions shaping its instructions. So agreement mostly shows that Method 2 reproduces my method; the better evidence is where it disagreed or caught something. Other limits: six runs is a small sample, and five of Method 2's six calls and fifteen of Method 1's eighteen, as first issued, are HOLDs. Method 1's BUYs (ServiceNow, Mogotes and ATS) have no Method 2 run, so nothing has tested whether the methods agree on a BUY. The October 17 re-pricing changed Method 2 targets without a new debate, so a revised target is what the model gives at today's price, not what the judge would rule. The one disagreement shows only that a bottom-up earnings estimate and a lower multiple can move a call by twenty points. No call has been scored. An experienced analyst might disagree with the judge's choices, such as a 0.45 beta, and I can't yet tell. Method 1's cost counts only the session that wrote each report. A fairer test would take a company I have no view on, seal Method 2's call, and only then write Method 1's. My Parker-Hannifin and Moog reports of October 12 ran at low reasoning effort, below the high effort I set that day as Method 1's standard, and I did not stop to ask first.

| What went wrong                                                                                                                                                               | How it was found                      | What changed                                                               |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|----------------------------------------------------------------------------|
| Three reports said I held no position in companies I owned; one had no ownership disclosure                                                                                   | My re-read of every report, Aug 9     | Four fixed disclosure blocks; ownership copied from the holdings file      |
| Two months of reports used Helvetica, not the intended serif font                                                                                                             | Checking the finished PDFs            | The build prints the font it used                                          |
| The first IREN export deleted a paragraph I wanted                                                                                                                            | Comparing with the previous PDF       | Every rebuild is diffed against the last                                   |
| My Microsoft advice to buy below \$500 lost money after currency costs                                                                                                        | Method 2's risk agent                 | Risk check in every Method 2 run                                           |
| The first Microsoft market agent made no data calls                                                                                                                           | The run log                           | Each run passes a table of which key is which                              |
| Agents spent a third of the Microsoft run searching for filing addresses                                                                                                      | Token audit                           | Addresses and data fetched once, in preflight                              |
| The valuation agent defended its beta with a figure from the wrong company, copied from the market analyst's file                                                             | The bull agent                        | Beta measured before any agent starts                                      |
| The risk agent was told my US allocation was below target using a figure revised the same day                                                                                 | The risk agent                        | Holdings note corrected                                                    |
| The Coca-Cola judge said currency cost exceeded any capturable return                                                                                                         | The risk agent, on my measured costs  | Left on the record as a disagreement                                       |
| API limits killed three Coca-Cola stages; two restarted from zero                                                                                                             | The run stopping                      | Agents save each section as they finish                                    |
| Method 1's Alphabet node dated the ad-tech ruling September 16, not September 2; the refresh had no calls-log row; the Aluula, first Visa and ATS calls were logged days late | Method 2's news file; reading the log | Node corrected and rebuilt; a refresh that moves a target gets its own row |
| The first October 17 revision of Caterpillar derived a bull multiple from the new price and flipped the call to SELL                                                          | My review of the revision             | Reversed; the multiple stays as the judge set it                           |
| The Parker-Hannifin and Moog reports ran at low reasoning effort                                                                                                              | The session notes, afterwards         | Model and effort checked, and my yes asked for, before any Method 1 job    |
| The first draft of this paper misstated the Microsoft currency example, and my token audit counted responses twice                                                            | Paper review and a recount, Sep 17    | Rewritten; costs recounted                                                 |

Most of my mistakes were in routine steps I didn't think needed checking. The agents caught errors in their own work as well as mine, but several problems above were in Method 2 itself.

## 9 Where I am now

As of October 17, 2026.

| Part of the project | Status                                                                                                                                                                                                                                                                                                      |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Method 1            | Eighteen reports. Coca-Cola and Alphabet were refreshed after their first report; the nine newest were initiated between September 23 and October 12. On October 17 ATS (HOLD on the interim bands, BUY on the final ones), Caterpillar and Moog were refreshed to the October 16 close; the rest were not. |
| Method 2            | Six runs (Microsoft, Coca-Cola, Alphabet, Visa, Moog, Caterpillar), each with a Method 1 call. The latest closed October 9. All six were re-priced on October 17 without a new debate.                                                                                                                      |
| Calls log           | Thirty rows, all open, none scored: the original 25 (the ATS initiation was added on October 17) plus five Method 2 rows from that day. First review June 20, 2027.                                                                                                                                         |

| Part of the project | Status                                                                                                                                                |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Code                | Public on GitHub (MIT licence) with a worked example on an invented company; brought up to the October 10 renaming, but without the October 17 bands. |

| Ticker | M | Called | Call | Price    | Target   | Review due | Ticker | M | Called | Call | Price    | Target   | Review due |
|--------|---|--------|------|----------|----------|------------|--------|---|--------|------|----------|----------|------------|
| NOW    | 1 | 06-20  | BUY  | \$95.12  | \$138    | 2027-06-20 | V      | 2 | 09-26  | HOLD | \$367.38 | \$376.46 | 2027-09-26 |
| IBM    | 1 | 08-08  | HOLD | \$237.28 | \$235    | 2027-08-08 | MOG.A  | 2 | 10-06  | HOLD | \$381.04 | \$343.86 | 2027-10-06 |
| IREN   | 1 | 08-08  | HOLD | \$41.23  | \$44     | 2027-08-08 | BULL   | 1 | 10-07  | HOLD | \$5.87   | \$5.50   | 2027-10-07 |
| MOG    | 1 | 08-08  | BUY  | C\$0.64  | C\$1.05  | 2027-08-08 | WM     | 1 | 10-08  | HOLD | \$208.94 | \$220    | 2027-10-08 |
| AAPL   | 1 | 08-08  | HOLD | \$313.33 | \$270    | 2027-08-08 | CAT    | 1 | 10-08  | HOLD | \$813.83 | \$730    | 2027-10-08 |
| GOOGL  | 1 | 08-09  | HOLD | \$354.30 | \$330    | 2027-08-09 | CAT    | 2 | 10-09  | HOLD | \$796.18 | \$684.79 | 2027-10-09 |
| KO     | 1 | 08-10  | HOLD | \$87.05  | \$82     | 2027-08-10 | ROK    | 1 | 10-12  | HOLD | \$432.59 | \$420    | 2027-10-12 |
| MSFT   | 1 | 08-11  | HOLD | \$506.06 | \$500    | 2027-08-11 | PH     | 1 | 10-12  | HOLD | \$962.56 | \$950    | 2027-10-12 |
| COST   | 1 | 08-16  | HOLD | \$961.10 | \$923    | 2027-08-16 | MOG.A  | 1 | 10-12  | HOLD | \$363.09 | \$360    | 2027-10-12 |
| MSFT   | 2 | 08-17  | HOLD | \$480.35 | \$487.18 | 2027-08-17 | ATS    | 1 | 10-12  | BUY  | C\$25.19 | C\$29    | 2027-10-12 |
| KO     | 2 | 08-21  | HOLD | \$91.10  | \$77.45  | 2027-08-21 | MSFT   | 2 | 10-17  | HOLD | \$529.05 | \$486.63 | 2027-10-17 |
| GOOGL  | 2 | 09-19  | SELL | \$349.54 | \$268.81 | 2027-09-19 | KO     | 2 | 10-17  | SELL | \$93.90  | \$69.32  | 2027-10-17 |
| GOOGL  | 1 | 09-20  | HOLD | \$349.54 | \$340    | 2027-09-20 | GOOGL  | 2 | 10-17  | SELL | \$340.59 | \$273.22 | 2027-10-17 |
| AUUA   | 1 | 09-23  | HOLD | C\$3.95  | C\$3.50  | 2027-09-23 | V      | 2 | 10-17  | HOLD | \$402.61 | \$379.71 | 2027-10-17 |
| V      | 1 | 09-24  | HOLD | \$361.52 | \$395    | 2027-09-24 | MOG.A  | 2 | 10-17  | HOLD | \$368.60 | \$339.97 | 2027-10-17 |

M is the method. Called dates are 2026. Every row is open, and the five October 17 rows are revised Method 2 targets; the original rows stay. Benchmark is SPY except MOG and AUUA (XIC.TO). MOG.A is Moog Class A, not Mogotes (MOG). Scoring is on alpha, not raw return: a HOLD on a stock that fell 20% when the index fell 2% was a bad call; one that fell 20% when the index fell 25% was a good one. When calls are scored, the judge sees the scored rows for that company and the three most recent scored rows on any company.

## What would show the calls were wrong

| Call                  | What to watch, and when                                                                                                      | What it would mean                                                                                                                                                                               |
|-----------------------|------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Microsoft, both       | Fiscal 2027 Q1 capital spending, about October 27                                                                            | The biggest model input is an estimate with no consensus behind it                                                                                                                               |
| Visa, both            | Q4 FY2026 results, October 27: revenue against the \$12,052M consensus implies, another severance charge, the dividend raise | No new severance by January means the 0.5% charge was too heavy; a forward multiple at 24.5x or more favours Method 1                                                                            |
| Alphabet, both        | Q3 results, late October; 2027 capital spending guidance, late January or early February                                     | Above \$280B supports Method 2's SELL; at or below \$240B breaks it. A market paying 25.6x 2027 earnings returns it to HOLD                                                                      |
| Caterpillar, Method 2 | Q3 margin excluding tariff recoveries (about October 28 or November 4)                                                       | A base-case margin below 19.2% flips it to SELL; a forward P/E near 17x by October 2027 means both HOLDs were too generous                                                                       |
| Moog, both            | First FY2027 guide, about November 20; Q1 FY2027, about January 29                                                           | A guide midpoint of \$11.00 or less puts Method 2 within a point of SELL, and a forward P/E below 23x by January means its HOLD was wrong. Method 1 flips to SELL only at an 18.6x base multiple |
| ATS, Method 1         | The share price against the C\$29 target                                                                                     | BUY at +13.0% on the final bands; a price above C\$26.97 makes it a HOLD                                                                                                                         |
| Coca-Cola             | Fiscal 2026 comparable margin; the Eleventh Circuit IRS ruling; the price                                                    | Margin below 31% means the bear was right. Method 2's call has been SELL since October 17 on the 5.10% yield; Method 1's HOLD turns SELL above \$99 on the old bands                             |

Flip points are the ones recorded when each call was made, on the  $\pm 15\%$  bands unless a row says otherwise. The  $\pm 8\%$  edges put several of them closer.

**Questions I still can't answer.** Do the methods disagree often, and who is right when they do (one disagreement in six tests; none scored)? Which stage does the real work (a three-way tie so far)? Is either method any good (first review June 20, 2027)?

## 10 Next steps

| # | Step                                               | Why                                                                                                                                                                                                                                                          | When                          |
|---|----------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------|
| 1 | Run a blind test                                   | Pick a company I have no view on, seal Method 2's call, then write Method 1. Alphabet and Moog came close: Method 2's call was final before Method 1 began, but I knew the answer, and I had briefly held Moog.                                              | Next run                      |
| 2 | More Method 2 runs on covered companies, and a BUY | Six runs can't say how often the methods disagree. ServiceNow, Mogotes and ATS are Method 1 BUYs with no run, so one of them is the obvious next test; Rockwell Automation, Parker-Hannifin, Webull, Waste Management and Costco also have nodes and no run. | After step 1                  |
| 3 | Cut the cost of a run, and count the last three    | More than half the cost is agents re-reading instructions, so shorter role prompts should save most. Measure against \$71 to \$88.                                                                                                                           | Alongside 1 and 2             |
| 4 | Watch the dated checks above                       | Visa and Microsoft on October 27, Caterpillar about October 28, Moog about November 20, Alphabet's guidance in late January.                                                                                                                                 | October 2026 to February 2027 |
| 5 | Track which stage decides                          | After about a dozen runs, cut the debate if it never decides a call. So far it decided two of six.                                                                                                                                                           | At twelve runs                |
| 6 | Set a reminder to score                            | The scoring script exists but nothing runs it. Without a reminder the log turns into unchecked opinions.                                                                                                                                                     | Before June 15, 2027          |

| # | Step                                                         | Why                                                                                                        | When               |
|---|--------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|--------------------|
| 7 | Score calls as they come due, then write the follow-up paper | ServiceNow first on June 20; most of the rest in August to October 2027. Report the calls that went wrong. | From June 20, 2027 |

If you work in equity research and see something wrong in the method, I'd like to hear it; the contact is on the cover. The Refinitiv licence that feeds preflight ends about January 24, 2027, after which runs fall back to EDGAR, stockanalysis.com and Alpha Vantage.

## 11 How I used AI, running it yourself, and what I learned

**What Claude did and what I did.** Claude did most of the reading, calculating and drafting in both methods. I chose the companies, set the standard and rules, designed both processes, reviewed outputs and decided what to change after each run. Claude drafted this paper and I edited it; everything in it comes from files in my project. To keep the AI honest: every figure is traced to a source in a working-notes file; statements come from SEC filings, not model memory; the bear and judge rebuild the valuation from scratch and match the base case to within cents; the judge checks claims against the evidence files; and the rating is a fixed field. None of this guarantees the numbers are right. It makes errors easier to find. All data is public except the temporary Refinitiv licence.

**Run it yourself.** Method 2 is at [github.com/Vidwaansinghania/Multi-Agent-Equity-Research](https://github.com/Vidwaansinghania/Multi-Agent-Equity-Research) under the MIT licence. Without keys, install the requirements and run the three build scripts on a worked run for an invented company, Acme Materials Corporation, to get the Word file, the 14-sheet Excel model and a six-page PDF. On a real company you need Claude Code: copy the example config, add your name and email (EDGAR refuses requests without them), copy the skill into Claude Code's skills folder and type "run a multi agent equity research report on MSFT". Expect 19 to 22 million effective tokens, \$71 to \$88 at API prices, and two to three hours of active time over half a day or more. The risk agent needs a holdings file of your own.

**What I learned.** Write the single-analyst version first and hold it to a real standard, because the multi-agent version can only be as good as the rules it inherits. Reported numbers are not the end of the analysis: free cash flow, EPS and beta all needed a second look. My most serious mistakes were in parts I thought routine: ownership disclosures, a currency cost I never priced, a token count wrong by almost double. For agents, make the rating a fixed field, make the judge read the evidence, and keep agents away from your answer until they've committed. Check where the tokens go after the first run. Log every call on the day it's made. Where the methods disagreed, it came from the evidence stage building its own estimate, not from the debate: arguing audits a model, and the estimate moves a rating. If I started again I would set up the blind test on day one, before I had a view to leak into the agents.

## References and disclosure

1. Anthropic. (2026). Claude Opus 5, Claude Opus 5.5, Claude Sonnet 5 and Claude Sonnet 5.5 [large language models]. Identifiers *claude-opus-5*, *claude-opus-5-5*, *claude-sonnet-5* and *claude-sonnet-5-5*, used June to October 2026. [anthropic.com](https://anthropic.com). Claude Code [software]. [claude.com/claude-code](https://claude.com/claude-code).
2. Xiao, Y., Sun, E., Luo, D., and Wang, W. (2024). TradingAgents: Multi-agents LLM financial trading framework. arXiv:2412.20138. Code: [github.com/TauricResearch/TradingAgents](https://github.com/TauricResearch/TradingAgents).
3. U.S. Securities and Exchange Commission. EDGAR company filings and XBRL company facts. [sec.gov/edgar](https://sec.gov/edgar). Federal Reserve Bank of St. Louis. FRED series DGS10. [fred.stlouisfed.org](https://fred.stlouisfed.org). Alpha Vantage stock market data API. [alphavantage.co](https://alphavantage.co).
4. CFA Institute. CFA Institute Research Challenge. [cfainstitute.org](https://cfainstitute.org).
5. Singhania, V. (2026). Multi-Agent-Equity-Research, version as of October 10, 2026. [github.com/Vidwaansinghania/Multi-Agent-Equity-Research](https://github.com/Vidwaansinghania/Multi-Agent-Equity-Research).

Company figures come from each company's SEC filings and are cited in the individual reports.

**Disclosure.** On the run dates I held Microsoft and Coca-Cola, a small Alphabet Class C position (0.3265 shares) at the Alphabet run, no Visa at the Visa run, Moog Class A (0.0535 shares, 1.0% of my book, bought the day of the run and sold four hours later) and Caterpillar (0.0616 shares, 2.4% of my book). At my October 15 holdings snapshot I held Microsoft, Coca-Cola, Caterpillar, Waste Management, Parker-Hannifin (0.0523 shares) and ATS (one share), the last two bought after my reports on them, and a Johnson & Johnson depositary receipt. I held no Alphabet, Visa, Moog, Rockwell Automation or Webull. Each report's ownership section gives the position on its snapshot date, and my current holdings are public on Blossom at [blossomsocial.com/users/Vidwaan-Singhania\\_\\_d7oVAMV5EWqGMPon](https://blossomsocial.com/users/Vidwaan-Singhania__d7oVAMV5EWqGMPon). This paper describes a student research project built for learning. It is not investment advice, a recommendation to buy or sell any security, or a solicitation. The ratings and targets are dated, unscored and may already be out of date.