FURTON RESEARCH

Simulating a Committee of Expert Investors as an Investment Strategy

A Methodology Paper · Pre-Results Disclosure

Nicholas Furton[1] · Stephen M. Ross School of Business, University of Michigan

June 2026

Revision 2026-06-25: methodology text reconciled with the implemented system; see revision history. As a pre-registration, this change is disclosed rather than made silently. Revision 2026-07-05: copyediting pass for clarity and concision; wording only, no methodological content changed. Also added: a quantified frontier-tier cost ceiling in the Section 8 cost-bounded fidelity limitation, and the concentration cap's equal-weight fallback in Section 6.3—disclosures of existing behavior, not process changes. Section 7 was corrected on where the price-only line appears: the performance panel shows both lines; the public mirror shows the total-return line with the tailwind disclosed. Section 7.1 was expanded to enumerate the implemented metric set, separating live-track from modeled figures. No performance results are reported.

Abstract

Furton Research is a public, research-driven micro-investment project designed to transparently assess the viability of simulating the decision-making of a committee of expert investors as an investment strategy. The committee comprises five distinct investors, each simulated by Anthropic's Claude grounded exclusively in primary sources authored by the individuals or their funds. This paper documents the system as built: the construction of five primary-source libraries, a four-phase committee engine that produces independent verdicts before structured deliberation, and a retrieval-augmented generation (RAG) pipeline that grounds each agent in the most relevant passages for a given security. It likewise documents the disciplined process for deploying and rebalancing a real $10,000 portfolio benchmarked against the Dow Jones Industrial Average. We report the engineering decisions, the economic trade-offs of operating large language models at scale—including a detailed account of the tension between retrieval and prompt caching—and the full methodology by which results will be generated. No performance findings are reported yet; this document establishes the apparatus and the pre-registered process so that subsequent results can be evaluated against a transparent, fixed methodology.

[1] Nick Furton, Ross School of Business, University of Michigan. Email: nicholas@furton.com

Contents

1. Introduction and Research Question

Professional investment management is, at its core, a structured disagreement among people who see the same facts through different frameworks. A value investor and a disruptive-growth investor can examine the same company and reach opposite conclusions, each internally coherent. Furton Research asks a specific, testable question: if the distinctive reasoning of several well-documented expert investors can be simulated by a large language model grounded in those investors' own writing, does a committee of such simulated investors produce investment decisions worth acting on?

The project is deliberately public and transparent. Rather than asserting that an artificial committee can or cannot invest well, it builds the apparatus, commits to a fixed process in advance, deploys real capital, and reports what happens against a clear benchmark. This paper documents everything up to the point of results: what was built, why each design choice was made, and the exact process by which a $10,000 portfolio is managed and compared to the Dow Jones Industrial Average.

The central claim under examination is narrow and falsifiable: that a committee of five primary-source-grounded simulated investors, deliberating over Dow 30 constituents, generates a portfolio whose risk-adjusted performance can be measured against the index it draws from. Whether that claim holds is the subject of future reporting. This document concerns only the method.

1.1 What this project is not

Three disclaimers frame the work. First, this is a research project, not investment advice; the committee's outputs are simulations of documented reasoning styles, not vetted recommendations. Second, the simulated investors are not the real individuals and do not represent their current views; they are language-model personas constrained to primary-source text. Third, the committee's composition introduces a structural bias that the project treats as a finding to be studied rather than a flaw to be hidden.

2. The Committee and the Primary-Source Principle

The committee comprises five investors chosen to span complementary, and frequently conflicting, frameworks. Each is simulated by Anthropic's Claude, constrained to reason from a library of that investor's own primary-source material. The deliberate design goal is genuine disagreement: a panel that always agreed would carry no more information than a single agent.

Investor	Framework	Primary-source corpus
Warren Buffett	Quality value, durable moats, owner earnings	Berkshire Hathaway shareholder letters (1977–2023)
Howard Marks	Market cycles, risk, second-level thinking	Oaktree memos and published writing
Joel Greenblatt	Quantitative value, the "magic formula"	Three books on special-situation and value investing
Cathie Wood	Disruptive innovation, exponential cost curves	ARK Big Ideas reports and interview transcripts
Leopold Aschenbrenner	AI infrastructure, regime shifts, scaling	"Situational Awareness" and a long-form interview

The primary-source principle is the project's methodological backbone. Each simulated investor reasons only from text the investor or their fund actually produced. This is what separates the project from prompting a model to "act like Warren Buffett," which would draw on the model's diffuse training-data impression of the persona. By grounding each agent in a curated corpus and instructing it to reason from documented positions, the simulation is anchored to evidence rather than caricature.

2.1 Library construction

Each corpus was assembled by collecting primary-source documents, segmenting them into passages ("chunks") of roughly 250 words, and storing each chunk with a citation and metadata in a structured manifest. The five libraries vary in size according to the available material:

Investor	Chunks (filtered)	Source character
Warren Buffett	451	High-doctrine filter applied to 2,082 total chunks
Leopold Aschenbrenner	480	Single treatise plus interview, no filter
Howard Marks	742	High-doctrine filter applied to 1,383 total chunks
Joel Greenblatt	793	Three complete books
Cathie Wood	733	Hybrid: ARK reports plus interview transcripts

For two investors a "high-doctrine" filter was applied, retaining only chunks expressing durable investment philosophy and discarding transactional or dated passages. The Cathie Wood corpus required a hybrid approach: because ARK's flagship reports are graphically dense and text-thin, interview transcripts were added to capture her reasoning in prose. These construction choices are documented because they materially shape what each agent can retrieve and therefore how it reasons.

2.2 A note on the committee's composition bias

The five frameworks cluster around company-level analysis—quality, value, innovation—with top-down coverage limited to Marks's cycle lens and Aschenbrenner's technology-regime view. No member reasons primarily from interest rates, liquidity, or the broad economic cycle. This is a known limitation. It means the committee is expected to favor durable consumer and platform businesses and to under-weight deep-cyclical, commodity, and rate-driven names. Rather than correct this by adding members arbitrarily, the project treats the composition as fixed and its tilt as a measurable characteristic of the strategy.

3. The Committee Engine

A single evaluation proceeds through four phases. The design separates independent judgment from social influence, so that the committee genuinely deliberates rather than merely averaging five opinions.

3.1 Four phases

1. **Enrichment.** A fast model (Claude Haiku) gathers current market data on the security via web search and produces a structured investment brief: valuation, recent financials,

competitive position, and material news. This step requires no primary-source grounding and is deliberately inexpensive.

2. **Blind evaluation.** Each active investor receives the brief and their retrieved primary-source context, then issues an independent verdict without seeing any colleague's view: a position (Buy, Pass, or Abstain) and a conviction score from 1 to 10. This preserves the independence of each framework.
3. **Deliberation.** Each investor is then shown the others' verdicts and theses and may respond to the most significant disagreement, revising conviction if warranted. This is the step that makes the panel a committee rather than a poll: members react to one another.
4. **Synthesis.** A high-capability model (Claude Opus) acts as committee secretary, writing a single official statement that records the final position, the argument that carried the majority, the most significant dissent, and the recommended action.

3.2 The weighted vote

The committee's position is computed from member verdicts by a conviction-weighted vote:

$$score = \Sigma \text{ conviction of Buy votes} \div \Sigma \text{ conviction of all non-abstaining votes}$$

A member who Abstains is excluded from the calculation entirely, while a high-conviction Pass enlarges the denominator and so lowers the score. The result falls between 0 and +1. Scores at or above +0.3 are read as a committee Buy; anything below that is a Pass. The same score governs position sizing in the portfolio: a stronger committee score earns a larger allocation.

Each member's verdict is parsed strictly. A response whose position or conviction cannot be read unambiguously is excluded from the tally rather than silently counted as a Pass, and the vote is treated as authoritative only when a quorum returns a parseable verdict: at least four of five members, or three of four when one abstains. A panel short of quorum is recorded as incomplete and cannot issue a Buy; the archived record carries the panel's completeness, whether quorum was met, and any missing members.

3.3 The two-stage filter

Running the full four-phase pipeline on every security is wasteful when the committee is plainly indifferent. A two-stage filter addresses this. After the blind vote, a stock advances to the expensive deliberation and synthesis phases only if it shows a real signal: a committee score at or beyond +0.2, or any single member at conviction 8 or higher on a Buy. The advancement bar sits deliberately below the +0.3 committee-Buy threshold, so that deliberation can still argue a marginal name up; a high-conviction Pass does not advance, since the name can never be bought. Securities that draw a uniformly low-conviction Pass are recorded with a deterministic note ("deliberation round skipped due to all-around low conviction") and incur no further cost. This concentrates the second stage on the decisions that matter. How much it saves in a given week depends on how many names the committee finds compelling; in a broadly favorable market, where most names draw a Buy, fewer are skipped.

4. Retrieval-Augmented Grounding

The most consequential technical decision in the project concerns how each agent receives its primary-source material. The naive approach injects a fixed slice of each investor's corpus into every evaluation. This is both wasteful and limiting: most of the corpus is irrelevant to any given security, and a fixed slice cannot adapt to the company under review.

4.1 Retrieval in this system

The project uses retrieval-augmented generation (RAG). Every chunk in every library is converted to a numerical vector (an "embedding") using a compact local model, all-MiniLM-L6-v2, that runs offline on commodity hardware with no per-use cost. Because this model encodes only about the first 256 tokens of any passage (roughly 190 words), each source is divided into chunks targeting 250 words with a 40-word overlap between neighbors. A chunk that runs past the model's window is truncated at the tail, so its embedding reflects the chunk's opening rather than its entirety; the overlap mitigates boundary loss, and this is accepted as a deliberate cost-versus-fidelity trade-off rather than a guarantee of full-passage encoding. At evaluation time the retrieval key is not the full investment brief but a compact, purpose-built query the enrichment model emits alongside it: a dense line of business, sector, economic, and thesis-and-risk keywords. Because the query is short, it fits inside the embedding window instead of being truncated to the brief's opening; if the line is ever absent, the system falls back to the first ~180 words of the brief. The query is embedded and compared against the investor's chunk vectors by cosine similarity; the twenty most relevant chunks are retrieved and injected as that agent's context for that specific security.

The effect is that when the committee evaluates a beverage company, the Buffett agent retrieves his writing on consumer brands and pricing power; when it evaluates a semiconductor firm, it retrieves his passages on circle of competence and technological change. Each agent reasons from the most relevant fraction of its corpus rather than a representative sample, and the full library—not a pre-truncated slice—becomes accessible.

A fallback preserves robustness: if embeddings have not been generated, the system reverts to evenly sampled chunks, tuned to return the same number of passages so that cost and behavior remain consistent. The system reports at startup which mode is active for each investor.

5. The Retrieval–Caching Trade-off and Token Economics

Operating five grounded language-model agents over deliberation rounds is the dominant cost of the project, and one trade-off proved central enough to warrant detailed treatment: retrieval and prompt caching work against each other.

5.1 Why the two conflict

Prompt caching is a provider feature that stores a block of input the first time it is sent and serves it cheaply on subsequent identical requests, dramatically reducing the cost of resending a large, unchanging context. It pays off only when the cached block is byte-for-byte identical between calls.

Retrieval guarantees the opposite. Because RAG retrieves different chunks for every security, each agent's corpus block changes from one stock to the next. The cached block is never reused, so caching never produces a read hit—and worse, the act of marking a block for caching adds a one-time write premium of roughly 1.25× the input price. Under retrieval, that premium is paid on every call and never recouped: the worst possible caching outcome.

5.2 How the conflict surfaced

The conflict was not theoretical. An early full-market screen came in far above the roughly $5 the naive estimate had assumed. Inspection of the run log showed the cause unambiguously: every single call reported zero cache reads and a full cache write of around 70,000 tokens. Each security paid the write premium and recouped nothing, because retrieval had made every context block unique. The original estimate had implicitly assumed warm-cache reads across the screen—an assumption retrieval silently invalidated.

5.3 The resolution

The fix makes caching conditional. When an agent uses retrieval, its context differs on every call, so the system sends the context as a plain block at the standard input rate and omits the cache marker entirely, eliminating the unrecoverable premium. When an agent falls back to the fixed even-sampling mode, where the context is stable across calls, caching is re-enabled and pays off as intended. The system inspects each investor's retrieval status per call and chooses accordingly.

This change alone reduced input cost on the corpus by approximately twenty percent with no effect on output quality, by removing a premium that retrieval guaranteed would be wasted.

5.4 Per-stock and per-run economics

With retrieval active and caching correctly disabled on the retrieval path, the realized cost structure for a single security is approximately:

Phase	Model	Approx. cost	Share
Enrichment	Haiku	~$0.10	~27%
Blind vote (5 agents)	Sonnet	$0.14–$0.18	~37%
Deliberation (advancers)	Sonnet	~$0.13	~33%
Synthesis	Opus	~$0.01	~3%

The committee votes (ten grounded agent calls per security across blind and deliberation rounds) dominate cost; enrichment and synthesis are minor. This is why retrieval depth, not enrichment, is the lever that matters. At a retrieval depth of twenty chunks per investor, a full real-time screen of the thirty Dow constituents costs on the order of $10, with the two-stage filter sparing the lower-conviction names from the expensive second stage.

5.5 Batch processing as a cost lever

For screens that need not be immediate, the project supports the provider's asynchronous batch mode, which processes requests at half price in exchange for a turnaround measured in minutes to hours rather than seconds. Because the screen is naturally run on a weekend before markets open, this trade-off is nearly free in practice and brings a full screen to roughly $6. Batch mode is offered as a per-run choice rather than a default, preserving the option of immediate daytime screens at full price.

Two observations follow for anyone replicating the work. First, scraping or otherwise cheaply sourcing the enrichment data would save almost nothing, because enrichment is under a tenth of cost; the expense lives in grounded agent reasoning. Second, the single most effective economy is asynchronous batch processing, followed by disciplined two-stage filtering, followed by retrieval depth—in that order.

6. Portfolio Construction and Management

The committee's output is converted into a real portfolio through a fixed, pre-committed set of rules. The rules are fixed in advance, but the workflow is not a single automated pipeline: the engine automates enrichment, the committee vote, and dated archiving, while position sizing, the concentration cap, rebalancing, and the risk metrics are produced in separate control panels the operator drives—importing each screen's output and reviewing it before acting. A $10,000 account is

managed against the Dow Jones Industrial Average, the same universe the committee screens, so that the benchmark is the honest counterfactual: holding the index instead of the committee's selections.

6.1 From verdicts to positions

Only securities the committee rates Buy enter the portfolio; it is long-only by construction. Each member judges a name solely for inclusion: a security it finds unattractive is recorded as a low-conviction Pass rather than a short verdict, in keeping with the long-only philosophies its members follow. Each position is sized in proportion to its committee score from the weighted vote of Section 3.2, normalized so the selected names sum to the full deployment. A position the committee backs strongly receives more capital than a marginal Buy. The sizing panel renders this allocation as a portfolio chart with exact dollar amounts for the account size, which becomes the order list.

6.2 Execution

The weekly cycle is fixed. The screen is run on a weekend with markets closed, so that enrichment reflects a stable end-of-week snapshot and the committee's verdicts are available well before the next session. Orders are placed the following Monday as dollar-denominated market orders timed to 10:00 a.m. Eastern, fifteen to thirty minutes after the open, when spreads have tightened from their most volatile opening state. Dollar-denominated orders allow fractional-share execution and let the conviction-weighted allocation be entered directly without manual share-count conversion.

6.3 Weekly rebalancing

After the first week, the portfolio is rebalanced rather than rebuilt. Each weekend's fresh screen produces new target weights, which are reconciled against current holdings under three pre-committed rules:

- **Exit on removal.** Any holding the committee no longer rates Buy is sold in full.
- **Trade only meaningful drift.** An existing position is trimmed or topped up only if its weight has moved at least three percentage points from target; smaller drifts are left alone to avoid trading on noise.
- **Reinvest proceeds.** Cash freed by exits is redeployed into the new targets, prioritizing the most under-weight names, so the portfolio stays fully invested.

The reconciliation is computed against current total portfolio value (holdings at market plus cash), so that target percentages translate into correct dollar trades as the portfolio's value changes. The process produces a weekly trade ticket listing only the securities that require action; in a typical week this is a handful of trades rather than a full re-entry. Edge cases are handled explicitly: when forced sales would leave cash that the drift band cannot redeploy, proceeds are deployed into under-weight targets regardless of the band; and when fresh buys would exceed available cash, all buys are scaled proportionally to fit, preserving their relative weights without overdrawing the account.

Two standing risk controls bound the portfolio independently of the weekly drift rules. A concentration cap holds any single position to at most twenty-five percent of the portfolio: weight in excess of the cap is redistributed across the remaining names and the allocation renormalized, so that no one name—however high its conviction score—can dominate the book. When fewer than four names are rated Buy the cap is arithmetically unsatisfiable; the allocation then falls back to equal weights, and the concentration is disclosed by the position count itself. A drawdown override flags whether the weekly ticket should be acted on. Whenever the committee's cumulative return trails the Dow's by ten percentage points or more, the performance panel publishes a drawdown signal; the

trade log reads it and surfaces a warning banner on the rebalance screen: “drawdown override tripped, review before acting.” The override is a computed monitor that calls for a manual pause and review; it does not automatically block order entry. Both controls are pre-committed and computed from recorded data: the cap in the position-sizing logic, the override from the same account-value-versus-Dow series as the headline result. They are therefore visible rather than discretionary, even though the operator applies them by hand.

6.4 The trade log

Every position is recorded in a trade log that captures the distinction between intent and execution: the committee’s recommended weight, a decision-time reference price, the actual fill price obtained, and the current market price for live profit-and-loss. The reference price is captured automatically at enrichment time from the screen’s web-search snapshot: an advisory, possibly end-of-day price recorded with an as-of timestamp. It is carried through the screener export and auto-filled into the trade log, where slippage is computed as the gap between it and the realized fill; the operator can override it by hand, and a missing price never blocks a screen. This record is the raw material for results. Because it stores both the recommendation and the realized fill, it will permit later analysis of execution slippage separately from the committee’s selection quality.

7. Benchmarking and the Measurement Plan

The portfolio is compared to the Dow Jones Industrial Average because the committee selects exclusively from the index’s thirty constituents. This makes the benchmark a true alternative: any divergence reflects the committee’s selection and weighting decisions, not exposure to a different universe. The headline metric is the percentage change in account value against the percentage change in the Dow’s price level over the same period—a price-return basis for the index. Because the portfolio reinvests every dividend, the recorded account value is already a total-return series, whereas the Dow’s price level excludes dividends—a small, known asymmetry. Rather than ignore it, the performance panel also derives a price-only return, stripping dividends period by period and chaining the result, so an apples-to-apples comparison against the Dow price index is available. Both lines are shown—solid for total return, dashed for price-only—on the performance panel; the public mirror shows the total-return line with the tailwind disclosed alongside it, and the exports carry both figures separately. The headline alpha is the total-return account value against the Dow price level, with the structural dividend tailwind disclosed.

7.1 Risk-adjusted performance and factor attribution

Raw return is necessary but not sufficient. From the dated equity curve the project reports risk-adjusted performance on the live track: Sharpe and Sortino ratios against a documented risk-free rate (a 3.8% annual constant), annualized volatility, maximum drawdown and the longest underwater stretch, and an information ratio (active return over the Dow divided by its tracking error), so that any excess return is judged against the risk taken to earn it. A second family of figures is modeled rather than lived: the current portfolio weights are applied to the trailing year of daily market data to produce one-day 95% Value at Risk and conditional Value at Risk, up- and down-capture against the S&P 500, beta and correlation to the S&P 500 and the Dow, and a modeled volatility and Sharpe ratio. Modeled figures describe the risk profile of the book as constructed, not a track record, and are labeled as such wherever they appear. Concentration is reported alongside both families: the effective number of names, the top-five weight share, and a Herfindahl–Hirschman index.

Crucially, excess return over the index is not by itself evidence of skill. The committee's disclosed tilt toward quality consumer and platform businesses means part of any outperformance may simply be compensation for a known factor exposure rather than security selection. To separate the two, the project pre-registers a factor attribution. It is planned but not yet built: no factor series are ingested and no regression code runs today. Once enough weeks accumulate, weekly portfolio excess returns will be regressed on the standard market, size, value, and momentum factors (the Fama–French and momentum series from the Kenneth French data library), together with a quality factor, drawn from an established long–short construction or proxied by a quality index. Outperformance is claimed only when the regression shows a statistically significant positive intercept (the residual alpha that remains after factor exposures are removed), and never on raw excess return.

This attribution carries an explicit honesty constraint. A five-factor regression on weekly data is statistically underpowered until many observations accumulate; with only a few months of weeks, the coefficients and the alpha estimate carry wide confidence intervals. Early results are therefore reported as indicative, with intervals shown, and the attribution is treated as conclusive only once the observation count is adequate. Pre-committing to this test—and to withholding any claim of skill until it can be run with power—is the project's primary guard against mistaking a factor tilt for an edge.

7.2 Process and execution diagnostics

Beyond return, several properties of the process itself are measured. Whether the deliberation phase earns its cost is testable directly: the blind and post-deliberation votes are both recorded for every security, so the results papers can report how often deliberation changes a verdict and whether revisions move toward or away from the emerging majority. To distinguish signal from sampling noise, the model's sampling temperature is fixed and recorded, and a periodic repeated-run check quantifies how stable the committee's verdicts are across identical inputs. The two-stage filter's advancement rate and the realized token cost per screen are logged each week. Execution quality is assessed as slippage (the gap between the decision-time reference price and the fill actually obtained) and as turnover from week-over-week holdings, so that selection quality can be separated from trading friction.

7.3 The recorded dataset

Every analysis above depends on capturing the right raw data as it happens; data not recorded at the time cannot be reconstructed later. The project therefore commits to a fixed weekly dataset, each series timestamped and archived rather than overwritten, so the results papers draw on the full panel rather than the latest snapshot. The archiving is partly automatic and partly manual: the committee screens are written server-side under dated keys as they run, while the equity curve and trade ticket are exported by the operator each week into a dated archive folder. The holdings figures published publicly are the committee's target weights; the actual position snapshot (share counts, market values, and realized weights) is captured privately from the brokerage export and reconciled against those targets. Four series are recorded each week:

- **The equity curve.** Account value and the Dow price level, together with any external cash flows and dividends received.
- **A holdings snapshot.** For each position, its ticker, share count, market value, and portfolio weight.
- **The full committee record.** For each security, the blind verdicts, the deliberation verdicts, the final weighted vote, the advancement decision, and the token cost of the screen.

- **The trade ticket.** For each trade, the committee's target weight, the decision-time reference price, the realized fill price, the share count, and the dollar amount.

The factor and risk-free series are external and are retrieved from public sources at analysis time.

No performance results are reported in this paper. The portfolio is newly established, and reporting partial or anecdotal returns before a meaningful track record exists—or before the factor attribution can be run with adequate power—would undermine the transparency the project is built to provide. Results will be reported separately, in one or more dated results papers, against the unchanged process and the fixed dataset described above.

8. Limitations

- **Simulation, not the individuals.** The agents are language-model personas constrained to primary-source text. They are not the real investors and do not represent those investors' current views or endorse any holding.
- **Composition bias.** The five frameworks lack a primary macro voice, predisposing the committee toward quality consumer and platform names and away from cyclical and rate-driven businesses. This tilt is a known, measurable property of the strategy; the factor attribution in Section 7.1 exists in part to price it.
- **Model and data drift.** Both the underlying language models and the enrichment data evolve over time, which may affect reproducibility; runs are therefore dated, the sampling temperature is held fixed (0.3), and the process versioned. In practice the dominant source of run-to-run variation is not the model's sampling but the enrichment brief: because each brief is assembled from a live web search, the same security can produce different briefs—and therefore different verdicts—on different days, whereas a fixed brief yields near-identical verdicts at this temperature. The exact brief behind every screen is archived, so any verdict can be reproduced from its recorded inputs.
- **Cost-bounded fidelity.** The committee's reasoning quality is bounded by the economics documented in Section 5. Budget pressure pushes toward cheaper configurations: fewer retrieved chunks per agent, or lighter models for the grounded reasoning steps. Each of these can measurably change the verdicts the committee reaches. The configuration reported here (a retrieval depth of twenty chunks per investor; a fixed sampling temperature of 0.3; Sonnet for agent reasoning and Opus reserved for synthesis) reflects a deliberate balance of fidelity against cost rather than a fidelity-maximizing ideal. A better-resourced instance of the same apparatus might reason differently, so results should be read as those of this specific, cost-bounded configuration.

The bound is concrete: as of this revision (July 2026), running the same weekly screen entirely on the frontier model tier would cost roughly $60–90 per screen. Simulation costs are research expenses borne outside the portfolio—no model spend is deducted from the account, and reported returns are unaffected by it. The comparison is still instructive: an annual model spend equal to 35–45% of the account's value, against low single digits for the reported configuration, would be economically indefensible as a way to manage $10,000. The ceiling is set by the experiment's capitalization, not the method, and would relax at larger scale. Nor would the frontier tier be a drop-in substitute: its always-on model reasoning and its removal of the fixed sampling-temperature control would make it a different configuration from the one registered here.

- **Single committee.** Results will reflect one particular five-member composition. They speak to this committee, not to simulated-committee investing in general.

9. Conclusion

Furton Research has built a complete apparatus for a specific, transparent experiment: five simulated investors, each grounded in its subject's own writing, deliberating as a committee over the Dow 30 and managing a real $10,000 portfolio under a fixed weekly process benchmarked against the index. This paper has documented that apparatus end to end—the libraries, the four-phase engine, the retrieval pipeline, the deliberation and two-stage filtering, and the portfolio construction and rebalancing rules—together with the engineering economics that govern its operation, including the retrieval–caching trade-off that proved central to running the system affordably.

What remains is the part that cannot be rushed: time in the market, recorded honestly against a benchmark, under a process fixed in advance. By publishing the method before the results, the project commits to evaluating whatever happens against an unchanged standard. The question of whether a simulated committee of experts can invest is now an empirical one, and the instrument to answer it is in place.

Primary Sources

The committee's five simulated investors are grounded exclusively in the primary-source corpora listed below. This list is current as of publication and is subject to expansion as additional primary material is collected. These works are cited as the documented basis for each simulated investor; the project does not reproduce or redistribute their text.

Warren Buffett

- Berkshire Hathaway, Inc. Shareholder Letters, 1977–2023. Published annually by Berkshire Hathaway and publicly available at berkshirehathaway.com.

Howard Marks

- Oaktree Capital Management client memos and published commentary by Howard Marks, drawn from the memo archive at oaktreecapital.com.

Joel Greenblatt

- You Can Be a Stock Market Genius. New York: Simon & Schuster, 1997.
- The Little Book That Beats the Market. Hoboken, NJ: John Wiley & Sons, 2005.
- The Little Book That Still Beats the Market. Hoboken, NJ: John Wiley & Sons, 2010.

The Big Secret for the Small Investor (2011) was reviewed and intentionally excluded; it advances value-weighted indexing as a retail-accessible alternative and introduces no new investment principles beyond those documented in the three included works.

Cathie Wood

- ARK Investment Management LLC. "Big Ideas" annual research reports: 2017, 2020, 2021, 2024, 2025, and 2026 editions.
- Interview: Chris Hayes, "Why Is This Happening?" (MSNBC podcast), conversation with Cathie Wood, March 2022.
- Interview: AI & Faith Town Hall, conversation with Cathie Wood, 2022.

As noted in Section 2.1, ARK's flagship reports carry their argument largely in graphics, so the interview transcripts supply Wood's reasoning in prose.

Leopold Aschenbrenner

- Situational Awareness: The Decade Ahead. June 2024.
- Long-form interview with Dwarkesh Patel (podcast transcript), 2024.

Furton Research · A public micro-investment research project · This document is a methodology disclosure, not investment advice. The simulated committee's outputs do not constitute recommendations.