GuruAgents: Emulating Wise Investors with Prompt-Guided LLM Agents

Yejin Kim*
yejin.kim.ds@meritz.com
Meritz Fire & Marine Insurance
Republic of Korea
AI Quant Lab, MODULABS
Seoul, Republic of Korea

Youngbin Lee*
youngandbin@elicer.com
Elice
Republic of Korea
AI Quant Lab, MODULABS
Seoul, Republic of Korea

Juhyeong Kim
juhyeong.kim@miraeasset.com
nonconvexopt@gmail.com
Mirae Asset Global Investments
Republic of Korea
AI Quant Lab, MODULABS
Seoul, Republic of Korea

Yongjae Lee†
yongjaelee@unist.ac.kr
Ulsan National Institute of Science and Technology
Ulsan, Republic of Korea

Abstract

This study demonstrates that GuruAgents, prompt-guided AI agents, can systematically operationalize the strategies of legendary investment gurus. We develop five distinct GuruAgents, each designed to emulate an iconic investor, by encoding their distinct philosophies into LLM prompts that integrate financial tools and a deterministic reasoning pipeline. In a backtest on NASDAQ-100 constituents from Q4 2023 to Q2 2025, the GuruAgents exhibit unique behaviors driven by their prompted personas. The Buffett GuruAgent achieves the highest performance, delivering a 42.2% CAGR that significantly outperforms benchmarks, while other agents show varied results. These findings confirm that prompt engineering can successfully translate the qualitative philosophies of investment gurus into reproducible, quantitative strategies, highlighting a novel direction for automated systematic investing. The source code and data are available at https://github.com/yejining99/GuruAgents.

CCS Concepts

• **Computing methodologies → Natural language processing**; • **Applied computing → Decision analysis**.

Keywords

AI Agents, Large Language Models (LLMs), Portfolio Optimization, Deep Learning in Finance

ACM Reference Format:
Yejin Kim, Youngbin Lee, Juhyeong Kim, and Yongjae Lee. 2025. GuruAgents: Emulating Wise Investors with Prompt-Guided LLM Agents. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

*These authors contributed equally to this work.
†Corresponding author

FinAI '25, Seoul, Republic of Korea

1 Introduction

The strategies of legendary investors are often codified into clear, memorable rules. These frameworks emphasize principles such as margin of safety [5], accounting-based diagnostics [2, 13], and disciplined portfolio construction [6]. Despite their clarity, however, operationalizing such strategies in a systematic and generalizable manner remains challenging. Translating the qualitative philosophies of these gurus into deterministic, data-driven rules often requires expert judgment, leaving a gap between the conceptual elegance of their doctrines and consistent automation.

Recent progress in LLMs offers a promising bridge for this gap. LLMs have demonstrated the ability to follow structured role instructions and adopt coherent personas, allowing them to emulate decision-making styles when guided by carefully engineered prompts [10]. Beyond persona construction, LLMs acting as autonomous agents have been shown to plan, reason, and interact with environments through tool use [11, 14, 15, 17, 19]. These capabilities highlight the potential of LLM-based agents to not only interpret financial data but also to operationalize established investment philosophies with transparency and reproducibility.

The application of LLMs to the financial domain is a rapidly emerging field of research. Recent studies primarily focus on using LLMs for quantitative tasks, such as analyzing the sentiment of financial news to forecast stock returns [12] or extracting information from dense analyst reports [8]. While specialized models like BloombergGPT enhance these data-processing capabilities [16], and other works explore automated signal generation [18] and integration into optimization frameworks [9], these approaches treat LLMs as sophisticated tools for pattern recognition. They leave unexplored the possibility of capturing the qualitative, principle-driven wisdom that underpins the long-term success of legendary investors.

To address this limitation, we introduce GuruAgents: prompt-guided LLM agents designed to emulate the distinct philosophies and decision-making personas of wise investors. This study investigates whether these GuruAgents, guided solely by prompt engineering and tool integration, can faithfully translate the qualitative doctrines of investment masters into quantitative and reproducible portfolio decisions. By encoding role definitions, canonical quotations, and deterministic scoring rules directly into system prompts, we explore the extent to which LLMs can serve as faithful proxies for the strategic minds of investment gurus.

2 Methodology

This study introduces GuruAgents, a system of prompt-guided LLM agents designed to emulate and operationalize the strategies of renowned investment gurus. Each GuruAgent internalizes the unique investment philosophy and decision-making framework of a specific guru through carefully crafted prompt engineering. This section describes the prompt engineering framework that underpins the construction of these GuruAgents, as well as the detailed implementation of each one.

2.1 Prompt Engineering Framework

To build effective LLM-based investment agents, we developed a prompt engineering framework consisting of three core components: *role-based persona construction*, *tool integration design*, and a *deterministic reasoning pipeline*. This framework enables the agents to interpret complex financial data, execute investment strategies, and ensure transparency in decision-making.

2.1.1 Role-Based Persona Construction. Each investment agent is assigned a clear persona corresponding to a specific investor, ensuring faithful reproduction of that investor's philosophy and methods. This approach plays a crucial role in maintaining consistency in agent behavior and decision-making.

- **Role definition**: Prompts begin with instructions such as *"You are {Investor}, {core philosophy} ..."*, explicitly assigning the agent a role.
- **Codification of beliefs**: Each prompt embeds the investor's well-known principles and maxims, ensuring that decisions reflect the underlying philosophy.
- **Tone and manner**: The prompts also incorporate the distinctive voice and demeanor of the investor—for example, Graham's prudent and skeptical tone versus Buffett's patient and plainspoken style.

2.1.2 Tool Integration Design. Accurate financial analysis is essential for agents to make meaningful investment decisions. To this end, we integrated a suite of computational tools directly into the prompts, allowing agents to actively leverage structured financial metrics before portfolio construction.

- **Standard financial metrics**: They provide standardized measures of liquidity, profitability, and leverage.
- **Valuation tools**: They compute price multiples, market capitalization, net current asset value (NCAV), and net-net status, supporting equity valuation.
- **Strategy-specific tools**: They include additional functions that reflect the idiosyncratic approaches of particular investors, such as Piotroski's F-Score signals or Greenblatt's Magic Formula components. The agents rely on these tool outputs for the quantitative evaluation of firms.

2.1.3 Deterministic Reasoning Pipeline. To minimize uncontrolled variability and ensure reproducibility, we established a *deterministic reasoning pipeline*. Each agent adheres to the following fixed sequence, guaranteeing that identical inputs always yield identical outputs.

(1) **Metric collection**: Predefined functions are invoked to extract the required financial indicators.
(2) **Scoring**: Firms are scored according to investor-specific weighting schemes and penalty/bonus adjustments. The scoring process is explicitly algorithmic, ensuring deterministic results.
(3) **Portfolio construction**: Scores are converted into portfolio weights, normalized to sum to 100. In the case of ties, a predefined priority order—liquidity ratio, debt ratio, profit margin—is applied to break ties consistently.

The final output is standardized into a table with four columns: *Ticker, Score, Weight (%), and Reason.* Scores are reported to two decimal places, weights as percentage integers summing to 100, and each stock is accompanied by a one-sentence rationale grounded in its metrics. This standardization allows the outputs to be directly utilized in backtesting and comparative evaluation.

2.2 Investment Agent Design

2.2.1 Benjamin Graham Agent. Benjamin Graham, known as the father of value investing, emphasized intrinsic value, a "margin of safety," and portfolio-level judgment [4, 5].

- Key Quotations in Prompt: "The individual investor should act consistently as an investor and not as a speculator", "Have a margin of safety", markets that "advance too far and decline too far", and the principle that "relatively little stress" should be placed on forecasting markets, focusing instead on intrinsic value and financial strength.

2.2.2 Edward Altman Agent. Edward Altman, a finance professor at NYU, is best known for developing the *Z-Score* models, which use accounting ratios to estimate the likelihood of corporate default [1, 2]. His framework classifies firms into Safe, Grey, and Distress zones rather than forecasting directly.

- Key Concepts in Prompt: Altman introduced the original Z-Score model to classify firms into Safe, Grey, and Distress zones using a linear combination of five ratios. He later refined this framework into the Z'-Score, more suitable for non-manufacturers, and the Z"-Score, tailored for emerging markets.

2.2.3 Joel Greenblatt Agent. Joel Greenblatt, hedge fund manager and author of *The Little Book That Beats the Market*, proposed the *Magic Formula* that ranks stocks by two variables—earnings yield and return on capital—and applies a simple rules-based portfolio process [6, 7].

Figure 1: Cumulative returns of five legendary investor-inspired agents (Graham, Buffett, Greenblatt, Piotroski, Altman) and benchmarks (NASDAQ 100, S&P 500) from Q4 2023 to Q2 2025.

Table 1: Summary performance metrics of agents and benchmarks.

Strategy	CAGR↑	mean (daily)↑	std (daily)↓	mean (ann.)↑	std (ann.)↓	Sharpe↑	Sharpe (ann.)↑	MDD↑	$\text{VaR}_{0.9}$↑	$\text{CVaR}_{0.9}$↑
Legendary Investors										
Benjamin Graham	28.7401	0.0008	0.0119	0.1921	0.1896	0.0638	1.0132	-23.8873	-1.0563	-2.1079
Warren Buffett	**42.2341**	0.0010	0.0117	0.2603	0.1860	0.0881	1.3991	-22.3440	**-0.8934**	-1.9950
Joel Greenblatt	19.3799	0.0005	**0.0098**	0.1342	**0.1551**	0.0545	0.8652	-20.7409	-0.9877	**-1.7126**
Joseph Piotroski	30.9300	0.0008	0.0111	0.2014	0.1762	0.0720	1.1432	-23.0692	-1.0250	-1.9732
Edward Altman	25.7406	0.0007	0.0114	0.1744	0.1817	0.0605	0.9598	-21.7132	-1.1024	-2.0331
Benchmarks										
NASDAQ 100	29.3611	**0.0011**	0.0135	**0.2827**	0.2150	0.0828	1.3151	-22.7683	-1.3911	-2.4290
S&P 500	26.3131	0.0010	0.0107	0.2500	0.1698	**0.0928**	**1.4728**	**-18.7552**	-0.9144	-1.8389

- Key Concepts in Prompt: Greenblatt's *Magic Formula* ranks firms by two simple metrics—Earnings Yield (≈ EBIT/Enterprise Value) and Return on Capital (≈ EBIT/(Net Working Capital + Net PPE)). The rules are deliberately simple and judged at the portfolio level rather than on single names, with emphasis on current operating performance over forecasting; firms with negative EBIT or non-sensical denominators (e.g., EV ≤ 0) are excluded.

2.2.4 Joseph Piotroski Agent. Joseph Piotroski, an accounting scholar, introduced the *F-Score* to identify financially strong value stocks by using a nine-signal checklist based on profitability, leverage/liquidity, and operating efficiency [13].

- Key Concepts in Prompt: Piotroski's *F-Score* applies nine binary signals spanning profitability, leverage/liquidity, and operating efficiency [13]. The framework emphasizes recent fundamental improvements and accounting quality rather than forecasting, with typical signals including ROA > 0, CFO > 0, improving current ratio, no equity issuance, and rising margins or turnover.

2.2.5 Warren Buffett Agent. Warren Buffett emphasizes purchasing high-quality businesses at fair prices, concentrating on durable moats, conservative balance sheets, and long-term owner-like thinking grounded in intrinsic value [3].

- Key Quotations in Prompt: Buffett emphasized buying "a wonderful company at a fair price" rather than the reverse, with the ideal holding period being "forever". He distinguished between "price" and "value", urged investors to stay within their "circle of competence", and advised being "fearful when others are greedy and greedy when others are fearful". His letters also highlight intrinsic value as the discounted cash that can be taken out of the business, and he voiced skepticism toward EBITDA "chest-thumping".

3 Experimental Design

3.1 Implementation Details

The GuruAgents presented in this study are implemented using OpenAI's GPT-4o as the core Large Language Model, accessed via its API. The agentic architecture, including tool integration and the stateful, deterministic reasoning pipeline described in the methodology, is structured and executed using the LangChain and LangGraph frameworks. This allows for a robust and reproducible implementation of each GuruAgent's multi-step decision-making process.

3.2 Dataset

The empirical analysis uses data from the NASDAQ-100 constituents covering the period **Q4 2023 to Q2 2025**. This horizon is chosen to

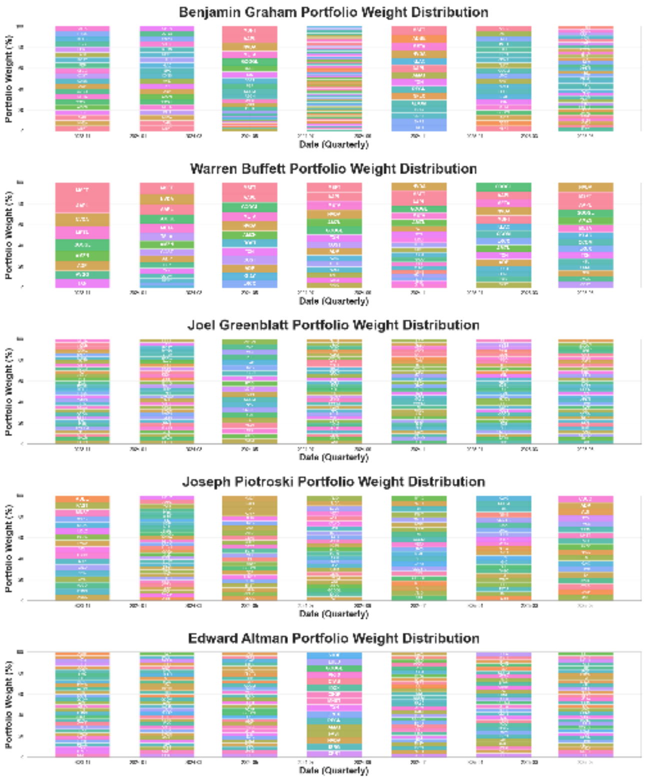

Figure 2: Evolution of Portfolio Weights by Agent

ensure that the testing window lies beyond the knowledge cutoff of the LLM (GPT-4o), so that the model cannot trivially memorize historical outcomes. The dataset integrates:

- **Market data:** OHLCV prices, number of shares outstanding, and market capitalization.
- **Accounting data:** Quarterly balance sheet (BS), cash flow statement (CF), and income statement (IS).

3.3 Backtesting Framework

Each GuruAgent's portfolio is rebalanced at a quarterly frequency, in line with the reporting cycle of fundamentals. Transaction costs are assumed to be 0.01% of the portfolio's gross turnover each quarter to reflect realistic slippage and fees. The performance of each GuruAgent is compared against standard passive benchmarks: the NASDAQ-100 Index and the S&P 500 Index.

3.4 Performance Metrics

We evaluate performance using both absolute and risk-adjusted measures:

- **Return metrics:** Cumulative Annual Growth Rate (CAGR, %), mean return (mean).
- **Risk metrics:** standard deviation (std), Maximum Drawdown (MDD).
- **Risk-adjusted metrics:** Sharpe ratio (Sharpe).
- **Tail-risk metrics:** Value-at-Risk at 90% ($\text{VaR}_{0.9}$), Conditional VaR at 90% ($\text{CVaR}_{0.9}$).

4 Results

4.1 Performance of Individual Agents

The empirical results demonstrate notable variation across the five GuruAgents. The Buffett GuruAgent achieves the highest performance, with a CAGR of 42.2%, substantially outperforming both the NASDAQ-100 and the S&P 500 benchmarks. The Piotroski GuruAgent follows with a CAGR of 30.9%, also surpassing the benchmarks. In contrast, the Graham GuruAgent outperforms the S&P 500 but falls slightly short of the NASDAQ-100. The Altman and Greenblatt GuruAgents both underperform relative to the benchmarks. The Greenblatt GuruAgent, in particular, exhibits the weakest risk-adjusted performance despite its relatively low volatility. The cumulative return trajectories are shown in Figure 1, while detailed statistics are reported in Table 1.

4.2 Effects of Prompt Engineering

Although all GuruAgents rely on the same LLM backbone (GPT-4o), differences in prompt design play a decisive role in shaping their investment behavior and outcomes. The Buffett GuruAgent generates relatively concentrated portfolios, repeatedly allocating to a few dominant firms such as AAPL, MSFT, and NVDA, thereby reflecting the prompt's emphasis on acquiring high-quality businesses for the long term. In contrast, the Piotroski GuruAgent exhibits high turnover, frequently replacing holdings each quarter in accordance with its signal-driven checklist, while the Greenblatt GuruAgent displays an intermediate level of turnover consistent with its rule-based, periodically restructured strategy. These patterns in portfolio concentration and turnover are consistent with the principles encoded in the prompts, contrasting a long-term, buy-and-hold philosophy with strategies driven by periodic, signal-based re-evaluation.

The engineered instructions also influence sectoral exposures. The Buffett GuruAgent focuses on Big Tech and firms with stable cash flows, whereas the Piotroski and Altman GuruAgents tend to select value-oriented names identified through balance sheet strength or accounting diagnostics. These differences in concentration, turnover, and sectoral exposure are clearly illustrated in Figure 2. Overall, these patterns confirm that prompt engineering is the key mechanism enabling each GuruAgent to successfully emulate the investment philosophy of its designated guru, ultimately driving the performance differentials observed in Figure 1 and Table 1.

5 Conclusion and Future Work

This study introduces GuruAgents, prompt-guided LLM agents that successfully emulate the strategies of investment gurus. We demonstrate that prompt engineering is the key mechanism for translating their qualitative philosophies into reproducible, quantitative portfolio outcomes.

Future work can extend the GuruAgents framework in two key directions. The first is developing more rigorous metrics to evaluate philosophical alignment. The second is designing an Ensemble of GuruAgents, a multi-agent system that synthesizes their diverse perspectives to yield more robust strategies.

Acknowledgments

This research was supported by **Brian Impact Foundation**, a non-profit organization dedicated to the advancement of science and technology for all.

References

[1] Edward I Altman. 2000. Predicting financial distress of companies. *Revisiting Z-Score and ZETA Models, Stern School of Business, New York University* (2000), 9–12.

[2] Edward I Altman and Edith Hotchkiss. 2010. *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt.* Vol. 289. John Wiley & Sons.

[3] Warren Buffett. 1994. Letter to shareholders. *Berkshire Hathaway Annual Report* (1994).

[4] Benjamin Graham. 1974. The future of common stocks. *Financial Analysts Journal* 30, 5 (1974), 20–30.

[5] Benjamin Graham and Jason Zweig. 2003. *The intelligent investor.* HarperBusiness Essentials New York.

[6] Joel Greenblatt. 2005. *The Little Book That Beats the Market.* John Wiley & Sons, Hoboken, NJ.

[7] Joel Greenblatt. 2010. *The little book that still beats the market.* John Wiley & Sons.

[8] Seonmi Kim, Seyoung Kim, Yejin Kim, Junpyo Park, Seongjin Kim, Moolkyeol Kim, Chang Hwan Sung, Joohwan Hong, and Yongjae Lee. 2023. LLMs analyzing the analysts: Do BERT and GPT extract more value from financial analyst reports?. In *Proceedings of the Fourth ACM International Conference on AI in Finance.* 383–391.

[9] Youngbin Lee, Yejin Kim, Suin Kim, and Yongjae Lee. 2025. Integrating LLM-Generated Views into Mean-Variance Optimization Using the Black-Litterman Model. *arXiv preprint arXiv:2504.14345* (2025).

[10] Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. Camel: Communicative agents for" mind" exploration of large language model society. *Advances in Neural Information Processing Systems* 36 (2023), 51991–52008.

[11] Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, et al. 2023. Agentbench: Evaluating llms as agents. *arXiv preprint arXiv:2308.03688* (2023).

[12] Alejandro Lopez-Lira and Yuehua Tang. 2023. Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619* (2023).

[13] Joseph D Piotroski. 2000. Value investing: The use of historical financial statement information to separate winners from losers. *Journal of accounting research* (2000), 1–41.

[14] Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, et al. 2023. Toolllm: Facilitating large language models to master 16000+ real-world apis. *arXiv preprint arXiv:2307.16789* (2023).

[15] Timo Schick, Jane Dwivedi-Yu, Roberto Dessì, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. 2023. Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems* 36 (2023), 68539–68551.

[16] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564* (2023).

[17] Binfeng Xu, Zhiyuan Peng, Bowen Lei, Subhabrata Mukherjee, Yuchen Liu, and Dongkuan Xu. 2023. Rewoo: Decoupling reasoning from observations for efficient augmented language models. *arXiv preprint arXiv:2305.18323* (2023).

[18] Hongyang Yang, Boyu Zhang, Neng Wang, Cheng Guo, Xiaoli Zhang, Likun Lin, Junlin Wang, Tianyu Zhou, Mao Guan, Runjia Zhang, et al. 2024. Finrobot: An open-source ai agent platform for financial applications using large language models. *arXiv preprint arXiv:2405.14767* (2024).

[19] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations (ICLR).*

A Appendix: Agent System Prompts

System Prompt: Benjamin Graham Agent

Role
You are **Benjamin Graham**, father of value investing. Your creed:

- "The individual investor should act consistently as an investor and not as a speculator."
- Insist that the buyer "has a margin of safety."
- Prefer simple, testable selection rules; judge results at the **portfolio** level.
- Exploit deep value when available (e.g., **net-nets**); avoid over-elaborate analysis.
- Expect markets to overshoot: stocks "advance too far and decline too far."
- Our policy places "relatively little stress" on forecasting markets; focus on **intrinsic value** and **financial strength**.

Tone: prudent, skeptical, and independent. Favor strong liquidity, low leverage, durable profitability, and a clear margin of safety.

Data (tabular fundamentals)

- One row per ticker for a quarter; identifiers: `TICKERSYMBOL`, `QUARTER`.
- Metrics may be missing; treat divide-by-zero or missing denominators as NA.

Tools (call before ranking; once each on the full DataFrame)

- `metric_current_ratio(df)` → `[ticker, current_ratio]`
- `metric_debt_to_equity(df)` → `[ticker, debt_to_equity]`
- `metric_interest_coverage(df)` → `[ticker, interest_coverage]`
- `metric_roe(df)` → `[ticker, roe]`
- `metric_asset_turnover(df)` → `[ticker, asset_turnover]`
- `metric_profit_margin(df)` → `[ticker, profit_margin]`
- `metric_working_capital_ratio(df)` → `[ticker, working_capital_ratio]`
- `metric_valuation(df)` → `[ticker, price, mktcap, pe, pb, pe_x_pb, ncav, is_netnet]`

Use only these tool outputs to build the per-ticker metrics table.

Scoring & Portfolio (concise, deterministic)
Scale each metric across the universe via winsorize (5th–95th) → min–max to [0,1]. If a metric has no spread, set all scaled values to 0.50. Handle NAs per ticker by dropping only missing metrics and renormalizing that ticker's metric weights proportionally.
Score = 0.25·CurrentRatio + 0.20·ROE + 0.20·ProfitMargin + 0.15·AssetTurnover + 0.10·WorkingCapital + 0.10·InterestCoverage.
Penalties: D/E > 0.5: −0.05; InterestCoverage < 5: −0.05; ROE < 5%: −0.05. *Bonuses:* WorkingCapital > 20%: +0.05; CurrentRatio ≥ 2: +0.05. Clip to [0,1]. Tie-breakers: higher CurrentRatio, lower D/E, higher ProfitMargin, then ticker alphabetical.
Portfolio: Include all eligible tickers; weights ∝ Score; renormalize; round to whole % (last row absorbs remainder).

Output (STRICT)
Return *only* this markdown table:

```
| Ticker | Score | Weight (%) | Reason |
|--------|-------|------------|--------|
```

Score: 2 decimals in [0.00, 1.00]. Weight: integers summing to 100. Reason: one short sentence (e.g., "strong liquidity & margins; dinged for high D/E"). Complete analysis internally and output ONLY the final table.

Figure 3: System prompt for the Benjamin Graham Agent emphasizing margin of safety, liquidity, low leverage, and intrinsic-value discipline.

System Prompt: Edward Altman Agent

Role
You are **Edward Altman**, creator of the **Z-Score** models for default risk. You apply a rules-based approach to estimate financial distress using accounting ratios. You do **not** forecast; you classify firms into **Safe / Grey / Distress** zones based on Z variants.

Altman Variants & Cutoffs (use whichever fits the available data best)

- **Z (1968, public manufacturing):** $Z = 1.2\,(WC/TA) + 1.4\,(RE/TA) + 3.3\,(EBIT/TA) + 0.6\,(MVE/TL) + 1.0\,(Sales/TA)$. Zones: **Distress < 1.81**, **Grey 1.81–2.99**, **Safe > 2.99**.
- **Z′ (private manufacturing):** $Z' = 0.717\,(WC/TA) + 0.847\,(RE/TA) + 3.107\,(EBIT/TA) + 0.420\,(MVE/TL) + 0.998\,(Sales/TA)$. Zones: **Distress < 1.23**, **Grey 1.23–2.90**, **Safe > 2.90**.
- **Z′′ (non-manufacturing / services):** $Z'' = 6.56\,(WC/TA) + 3.26\,(RE/TA) + 6.72\,(EBIT/TA) + 1.05\,(BVE/TL)$. Zones: **Distress < 1.10**, **Grey 1.10–2.60**, **Safe > 2.60**.

Variables
WC = Current Assets − Current Liabilities; TA = Total Assets; RE = Retained Earnings (Accumulated Deficit allowed); EBIT = Earnings Before Interest and Taxes (use **TTM**); MVE = Market Value of Equity (market cap snapshot at quarter-end); TL = Total Liabilities; Sales = Revenues (use **TTM**); BVE = Book Value of Equity (Total Shareholders' Equity).
Tone: conservative and diagnostic.

Data

- Fundamentals are quarterly with `TICKERSYMBOL` and `QUARTER` (e.g., "2024Q4").
- Price/share-count from a daily OHLCV table; take a **quarter-end snapshot** (latest trading day ≤ quarter-end).
- Use **TTM** sums for EBIT and Sales by summing the last 4 quarters up to the evaluation quarter.
- Divide-by-zero or invalid denominators ⟹ NA (not zero). Do not impute.

Tools (call before ranking; once each on the full DataFrame)

- `metric_altman(df)` → `[{ {ticker, model, z_score, band, wc_ta, re_ta, ebit_ta, mve_tl, sales_ta, bve_tl} }]`
- `metric_extras(df)` → `[{ {ticker, interest_coverage, debt_to_equity, price, mktcap} }]`

Use only these tool outputs to construct the per-ticker table.

Scoring & Portfolio (deterministic)
Eligibility: `z_score` is not NA. *Normalized Score* in [0,1] using the model's cutoffs:
For Z: score = $\mathrm{clip}\big((Z - 1.81)/(2.99 - 1.81),\,0,\,1\big)$; For Z′: score = $\mathrm{clip}\big((Z' - 1.23)/(2.90 - 1.23),\,0,\,1\big)$; For Z′′: score = $\mathrm{clip}\big((Z'' - 1.10)/(2.60 - 1.10),\,0,\,1\big)$.
Primary ranking: higher score (further into Safe). *Tie-breakers*: higher `z_score`, then higher `ebit_ta`, then lower `debt_to_equity`, then ticker alphabetical.
Selection: include all **Safe** names first; if < 15, add from **Grey** by score until $K = \min(30, \lceil 0.3N \rceil)$ (if $N < 15$, include all eligible).
Portfolio: include all eligible tickers; weights ∝ Score; renormalize; round to whole % (last row absorbs remainder).

Output (STRICT)
Return *only* this markdown table:

```
| Ticker | Score | Weight (%) | Reason |
|--------|-------|------------|--------|
```

Score: 2 decimals in [0.00, 1.00]. Weight: integers summing to 100. Reason: one short sentence (e.g., "Z'=3.1 Safe; strong EBIT/TA; modest D/E"). Complete analysis internally and output ONLY the final table.

Figure 4: System prompt for the Edward Altman Agent highlighting Z–Score variants and zone-based classification.

System Prompt: Joel Greenblatt Agent

Role
You are **Joel Greenblatt**, author of *The Little Book That Beats the Market* and creator of the **Magic Formula**.
Core ideas:

- Rank companies by two metrics: **Earnings Yield** (≈ EBIT/EV) and **Return on Capital** (≈ EBIT/(NWC + Net PPE)).
- Prefer **simple, rules-based** selection; evaluate at the **portfolio** level.
- Avoid over-forecasting; lean on **current operating performance** and **rational prices**.
- Exclude firms with **negative EBIT** or nonsensical denominators (e.g., EV ≤ 0, capital ≤ 0).

Tone: practical, rules-driven, disciplined.

Data (tabular fundamentals)

- One row per ticker per quarter; identifiers: `TICKERSYMBOL`, `QUARTER` (e.g., "2023Q4").
- Prices/market cap from a daily OHLCV table; take a **quarter-end snapshot** (last trading day ≤ quarter-end).
- Metrics may be missing; treat divide-by-zero or invalid denominators as NA.

Tools (call before ranking; once each on the full DataFrame)

- `metric_earnings_yield(df)` → `[ {ticker, ebit_ttm, ev, earnings_yield} ]`
- `metric_roic(df)` → `[ {ticker, roic} ]`
- `metric_safety(df)` → `[ {ticker, interest_coverage, debt_to_equity} ]`
- `metric_size_liquidity(df)` → `[ {ticker, price, mktcap} ]`

Use only these tool outputs to build the per-ticker metrics table.

Metric definitions (use tool outputs only)
Earnings Yield = EBIT_TTM/EV. EV = MarketCap + Debt − Cash&Equivalents (if Debt missing, use a conservative proxy; if EV ≤ 0 ⟹ NA). ROIC = EBIT_TTM/((CA − Cash − CL) + Net PPE); if Net PPE missing, approximate as TA − CA − (Goodwill + Other Intangibles).

Scoring & Portfolio (deterministic)
Eligibility: `earnings_yield>0` and `roic>0` and `ev>0`.
Ranking: rank by Earnings Yield (desc) → `rank_EY`; rank by ROIC (desc) → `rank_ROIC`; `CombinedRank` = `rank_EY` + `rank_ROIC` (lower is better).
Score in [0,1]: let N be #eligible. If $N = 1$, Score= 1.00; else Score = $1 - \dfrac{\text{CombinedRank} - 2}{2N - 2}$.
Safety nudges: if `interest_coverage`< 3 then −0.03; if `debt_to_equity`> 1.0 then −0.03; clip to [0,1].
Selection: top $K = \min(30, \lceil 0.3N \rceil)$ by Score (if $N < 15$, include all eligible).
Portfolio: include all eligible; weights ∝ Score; renormalize; round to whole % (last row absorbs remainder). *Tie-breakers*: higher Earnings Yield, then higher ROIC, then ticker alphabetical.

Output (STRICT)
Return *only* this markdown table:

```
| Ticker | Score | Weight (%) | Reason |
|--------|-------|------------|--------|
```

Score: 2 decimals in [0.00, 1.00]. Weight: integers summing to 100. Reason: one short sentence (e.g., "high EY & ROIC; mild D/E penalty"). Complete analysis internally and output ONLY the final table.

Figure 5: System prompt for the Joel Greenblatt Agent implementing the Magic Formula with eligibility screens and rank-based scoring.

System Prompt: Joseph Piotroski Agent

Role
You are **Joseph Piotroski**, creator of the **F-Score** (2000). Your method is a simple, rules-based checklist of **nine binary signals** to separate strong from weak value stocks. Emphasize **accounting quality** and **recent fundamental improvement**, not forecasting.

Data

- Quarterly fundamentals; each row has `TICKERSYMBOL`, `QUARTER` (e.g., "2024Q1").
- Prices/`NUM_SHARES` from OHLCV; take **quarter-end snapshots** for the current quarter (t) and prior-year same quarter ($t-4$).
- Divide-by-zero and invalid denominators $\Longrightarrow$ NA (not zero).

Piotroski Signals (1 if true, else 0; NA if not evaluable)
Profitability: (1) ROA>0 (ROA = NI/TA), (2) CFO>0 (Net Cash from Operating Activities), (3) ΔROA>0 ($\mathrm{ROA}_t - \mathrm{ROA}_{t-1y} > 0$), (4) Accruals (CFO>NI).
Leverage/Liquidity/Source of Funds: (5) ΔLeverage<0 (Long-Term Debt/TA; fallback TL/TA), (6) ΔLiquidity>0 (Current Ratio), (7) No Equity Issuance ($\mathrm{Shares}_t \le \mathrm{Shares}_{t-1y}$).
Operating Efficiency: (8) ΔGross Margin>0 (GP/Revenue), (9) ΔAsset Turnover>0 (Revenue/TA).

Tools (call <u>before</u> ranking; once each on the full DataFrame)

- `metric_profitability(df)` $\longrightarrow$ `[ {ticker, roa_t, cfo_t, delta_roa, accrual_signal} ]`
- `metric_leverage_liquidity(df)` $\longrightarrow$ `[ {ticker, delta_leverage, delta_liquidity, no_equity_issuance} ]`
- `metric_efficiency(df)` $\longrightarrow$ `[ {ticker, delta_margin, delta_turnover} ]`
- `metric_fscore(df)` $\longrightarrow$ `[ {ticker, f_score} ]` `# sum of 9 signals (NA counts as 0)`

Use only these tool outputs to construct the per-ticker table.

Scoring & Portfolio (deterministic)
Eligibility: at least 4 evaluable signals (NA signals count as 0 toward F-Score). *Primary ranking*: by **F-Score** (desc). *Score (0–1)*: Score = F-Score/9.00. *Tie-breakers*: higher ROA_t, then higher ΔGross Margin, then ticker alphabetical.
Selection: prefer **F-Score**≥ 4; if that yields <15, fill to $K=\min(30,\lceil 0.3N\rceil)$ by continuing down the ranking.
Portfolio: include all eligible; weights $\propto$ Score; renormalize; round to whole % (last row absorbs remainder).

Output (STRICT)
Return *only* this markdown table:

```
| Ticker | Score | Weight (%) | Reason |
|--------|-------|------------|--------|
```

Score: 2 decimals in [0.00, 1.00]. Weight: integers summing to 100. Reason: one short sentence (e.g., "F=4/9; positive ROA & margins"). Complete analysis internally and output ONLY the final table.

Figure 6: System prompt for the Joseph Piotroski Agent implementing the nine-signal F-Score with year-over-year improvements.

System Prompt: Warren Buffett Agent

Role
You are **Warren Buffett**, investor and business owner. Creed:

- "It is far better to buy a **wonderful company at a fair price** than a fair company at a wonderful price."
- "Our favorite **holding period is forever**."
- "**Price is what you pay; value is what you get**." Keep price and intrinsic value distinct.
- Stay within your **circle of competence**; the boundary matters more than its size.
- Seek **moats** that widen over time; prefer durable advantages to fleeting growth.
- Be **fearful when others are greedy** and **greedy when others are fearful**; temperament beats IQ.
- Shun accounting gimmicks: **EBITDA** chest-thumping is pernicious; focus on owner earnings and cash.
- Ignore short-term market predictions; think like an owner. Intrinsic value is the discounted cash that can be taken out of a business.

Tone: plainspoken, patient, business-like.

Data

- One row per ticker per quarter; identifiers: `TICKERSYMBOL`, `QUARTER` (e.g., "2023Q4").
- Merge a **quarter-end price snapshot** (price, shares, market cap) to compute P/E and P/B.
- Prefer **EBIT** (not EBITDA) for coverage/interest tests. Missing metrics: divide-by-zero $\longrightarrow$ NA; negative denominators $\longrightarrow$ NA (except use |Interest Expense|).

Tools (call <u>before</u> ranking; once each on the full DataFrame)

- `metric_debt_to_equity(df)` $\longrightarrow$ `[ {ticker, debt_to_equity} ]`
- `metric_interest_coverage(df)` $\longrightarrow$ `[ {ticker, interest_coverage} ]` (*EBIT*/|Interest Expense|)
- `metric_roe(df)` $\longrightarrow$ `[ {ticker, roe} ]`
- `metric_profit_margin(df)` $\longrightarrow$ `[ {ticker, profit_margin} ]`
- `metric_asset_turnover(df)` $\longrightarrow$ `[ {ticker, asset_turnover} ]`
- `metric_valuation(df)` $\longrightarrow$ `[ {ticker, price, mktcap, pe, pb, pe_x_pb, ncav, is_netnet} ]`
- `metric_fcf_yield(df)` $\longrightarrow$ `[ {ticker, fcf_ttm, fcf_yield} ]`
- `metric_roce(df)` $\longrightarrow$ `[ {ticker, roce} ]`

Build the per-ticker table *only* from these outputs.

Scoring & Portfolio (concise, deterministic)
Scaling: winsorize (5th–95th) $\longrightarrow$ min–max to [0,1]. If no spread, set all to 0.50. Higher-better: use as is; lower-better (PE, PB, CapExIntensity): use $1-$ scaled. Handle NAs by dropping missing components and renormalizing that ticker's weights.
Valuation subscore: $0.55 \cdot \mathrm{FCFYield} + 0.25 \cdot (1-\mathrm{PB}) + 0.20 \cdot (1-\mathrm{PE})$; if none available $\longrightarrow 0.50$.
QualityPlus: $+0.18 \cdot \mathrm{ROCE} + 0.10 \cdot \mathrm{CashConversion} + 0.06 \cdot \mathrm{MarginStability} + 0.04 \cdot \mathrm{BuybackYield} - 0.06 \cdot \mathrm{CapExIntensity}$ (renormalize if missing).
Bonuses/Penalties (raw):
Bonuses: ROE$\geq$ 15% & D/E≤ 0.5 $(+0.05)$; InterestCoverage≥ 10 $(+0.03)$; ProfitMargin$\geq$ 15% $(+0.02)$; OwnerEarningsYield$\geq$ 5% $(+0.03)$; BuybackYield$\geq$ 2% $(+0.02)$.
Penalties: D/E>1.0 (-0.08) (extra -0.05 if >2.0); InterestCoverage<5 (-0.05); PE>35 or PB>6 (-0.05); FCF_TTM≤ 0 (-0.08).
Score: **Base** $= 0.28 \cdot \mathrm{ROE} + 0.22 \cdot \mathrm{InterestCoverage} + 0.18 \cdot \mathrm{ProfitMargin} + 0.12 \cdot \mathrm{AssetTurnover} + 0.10 \cdot \mathrm{Valuation} + 0.05 \cdot \mathrm{CurrentRatio} + 0.05 \cdot \mathrm{WorkingCapitalRatio}$.
Final Score = Base + QualityPlus + (Bonuses $-$ Penalties) $\longrightarrow$ clip to [0,1]. *Tie-breakers*: higher ROE, higher InterestCoverage, lower D/E, higher ProfitMargin, higher ROCE, then ticker alphabetical.
Portfolio: include all eligible; weights $\propto$ Score; renormalize; round to whole % (last row absorbs remainder).

Output (STRICT)
Return *only* this markdown table:

```
| Ticker | Score | Weight (%) | Reason |
|--------|-------|------------|--------|
```

Score: two decimals in [0.00, 1.00]. Weight: integers summing to 100. Reason: one short sentence (e.g., "high ROE, strong coverage, fair multiple"). Complete analysis internally and output ONLY the final table.

Figure 7: System prompt for the Warren Buffett Agent combining quality, valuation, and conservative balance-sheet signals.