

**Artificial Intelligence and Portfolio Management:
A Comparative Backtest of an AI-Assisted Portfolio and an Equal-Weight Portfolio**

Author: **Christopher Chemaly¹**

Affiliation:

1. *Business School, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*
2. *CIRAME Research Center, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*

Abstract:

This study examines whether an artificial-intelligence-assisted asset allocation can improve portfolio performance relative to a simple equal-weight portfolio. Using a fixed set of five U.S.-listed exchange-traded funds representing large-cap equities, growth-oriented equities, gold, long-term Treasury bonds, and real estate, an AI-generated allocation was compared with a 20%-per-asset equal-weight benchmark over the 2021-2025 period. The AI-assisted portfolio allocated 30% to SPY, 30% to QQQ, 15% to GLD, 10% to TLT, and 15% to VNQ. Annual total-return data with distributions reinvested were used to calculate annual returns, cumulative return, compound annual growth rate, volatility, worst annual return, maximum drawdown, and a descriptive Sharpe ratio. The AI-assisted portfolio reached a cumulative return of 76.16%, compared with 56.57% for the equal-weight portfolio, and produced a higher descriptive Sharpe ratio (0.67 versus 0.59). However, it also showed higher annual volatility (20.49% versus 18.49%) and a slightly larger worst annual loss. The results indicate that, in this historical sample, the particular AI-generated allocation had a different risk-return profile and a higher cumulative outcome than the equal-weight benchmark. The study does not establish that AI consistently predicts markets or outperforms traditional methods, and the findings are limited by the small sample, single AI prompt, annual frequency, and exclusion of transaction costs.

Keywords:

Artificial intelligence, portfolio management, asset allocation, diversification, backtesting, investment performance, Sharpe ratio

Introduction:

Portfolio management is fundamentally a decision problem: an investor must decide how much capital to allocate across assets with different expected returns, risks, and relationships with one another. Modern Portfolio Theory argues that the quality of a portfolio cannot be judged only by the return of individual assets because diversification can change the overall risk-return profile of the portfolio (Markowitz, 1952). As financial markets have become more data-intensive, artificial intelligence (AI) has increasingly been considered as a complementary tool for investment analysis, asset selection, forecasting, and portfolio optimization.

Recent research demonstrates that the role of generative AI in investing is still unsettled. Ko and Lee (2024) found that ChatGPT-assisted asset selection could improve diversification relative to randomly selected assets and reported stronger portfolio efficiency for the ChatGPT-selected portfolios. In contrast, Perlin et al. (2025) tested Gemini 1.5 Flash across 1,522 U.S. companies and reported that it did not consistently outperform a naive equal-weight portfolio or the S&P 500 on returns and Sharpe ratios. Schneider and Yilmaz (2025) likewise found that the performance of ChatGPT-generated portfolios depends on the model and investor risk profile.

These findings create a practical research gap. Existing studies often use large stock-level datasets, multiple simulations, or sophisticated machine-learning models. A simpler and more transparent question remains useful for finance students and individual investors: can an AI-generated diversified asset allocation produce stronger risk-adjusted performance than a basic equal-weight portfolio when both are tested over the same historical period?

The main research question is therefore: Does an AI-assisted, diversified portfolio allocation improve investment performance compared with an equal-weight portfolio over the 2021-2025 period? The study has four objectives: (1) construct an AI-assisted portfolio using a reproducible prompt and a fixed set of five exchange-traded funds (ETFs); (2) compare the portfolio's annual and cumulative returns with an equal-weight benchmark; (3) compare volatility, worst annual return, drawdown, and a simple Sharpe ratio; and (4) evaluate whether AI can be used as a complementary allocation tool without assuming that it will automatically outperform a traditional benchmark.

Literature Review:

AI in finance has developed from rule-based automation toward machine learning, deep learning, and large language models. Dakalbab et al. (2024), in a systematic review of 143 studies published between 2015 and 2023, found that AI is widely used in financial trading and portfolio-related applications. Their review identified a broad range of techniques and noted that technical analysis is more frequently adopted than fundamental analysis in the surveyed literature. The review also emphasizes that financial data are noisy, time-varying, and sensitive to changing market conditions, which makes performance evaluation and backtesting particularly important.

From an investment-management perspective, AI can be viewed as a decision-support or co-pilot technology rather than only as an autonomous trading system. Ko and Lee (2024) examined ChatGPT from a portfolio-management perspective and found that the model could select assets across multiple asset classes in a way that improved diversification measures relative to random selection. Their results suggest that a language model can contribute to portfolio construction, particularly at the asset-selection stage. This is directly relevant to the present study because the experiment focuses on allocation across multiple asset classes rather than predicting the exact price of a single stock.

However, other evidence cautions against interpreting AI output as automatically superior. Perlin et al. (2025) conducted a large simulation using Gemini 1.5 Flash, anonymized historical information, and 30,000 simulations across 1,522 companies. They found no consistent evidence that Gemini outperformed either a naive portfolio or the S&P 500 in terms of returns and Sharpe ratios. Their results are important because an equal-weight portfolio is extremely difficult to beat after considering risk, diversification, and the instability of forecasts. This supports using an equal-weight portfolio as a transparent benchmark in the current experiment.

Schneider and Yilmaz (2025) provide another perspective by examining how ChatGPT-generated portfolios respond to different risk appetites. Their evidence indicates that portfolio outcomes can vary with the model used and the risk profile specified in the prompt. This suggests that the AI prompt itself is part of the investment process. Therefore, the current study records the portfolio constraints and allocation rule explicitly instead of treating the AI recommendation as an unexplained black box.

The broader investment-management literature also indicates that AI adoption involves governance and explainability questions. Mertzanis (2025) reviewed 178 peer-reviewed studies on AI and investment management and identified portfolio optimization, forecasting, risk management, robo-advisory, strategic decision-making, and adoption constraints as major themes. The review highlights explainability and governance as continuing research gaps. Accordingly, this project treats AI as a complementary decision tool and focuses on transparent, reproducible portfolio construction rather than claiming that AI can predict markets with certainty.

The theoretical foundation for the experiment is Modern Portfolio Theory (Markowitz, 1952), which frames portfolio selection as a balance between expected return and risk through diversification. Portfolio performance is evaluated using return and risk metrics, including a simple Sharpe-ratio formulation inspired by Sharpe (1966). The literature therefore supports an experiment that compares a structured AI allocation with a simple diversified benchmark and reports both return and risk rather than return alone. Beyond direct portfolio generation, the integration of AI into asset management is supported by a broader ecosystem of advanced analytics designed to navigate financial uncertainty (Subrahmanyam & Beainy, 2027) and improve institutional decision-making capabilities (Subrahmanyam et al., 2024). Recent empirical work demonstrates AI's capacity to decode complex market signals that influence asset allocation; for instance, deep learning models are actively deployed to analyze investor sentiment

and forecast major equity indices like the S&P 500 (Khalife et al., 2025), while ensemble machine learning techniques have been utilized to predict macroeconomic phases through shifting asset returns (Nasser et al., 2025). Furthermore, AI is enhancing complex valuation frameworks, such as integrating artificial neural mechanisms with Monte Carlo simulations for derivative pricing (Khalife et al., 2023), and leveraging technology-analytic adeptness to model the interrelatedness of macroeconomic variables (Bejjani et al., 2024).

When constructing a U.S.-centric portfolio—which inherently navigates the global structural dependency on the U.S. dollar (Beainy et al., 2025) and shifting comparative international equity perspectives (Beainy & Kamel, 2025)—diversification into non-correlated assets becomes essential. Recent quantitative analyses underscore the enduring hedging utility of gold compared to highly volatile digital assets (Beainy & Kamel, 2026) and highlight its critical role as a tangible financial resilience strategy against systemic economic fragility (Jabbour Al Maalouf et al., 2026). Finally, as reliance on these sophisticated computational models grows across the financial sector, scholars caution that the ethical integration of AI remains a significant cultural and operational challenge (Diab & El Hajj, 2024). This reinforces the premise of the current study: AI is currently best utilized as a transparent, complementary decision-support tool rather than an autonomous portfolio manager.

Methodology:

Research design. The study uses a comparative historical backtest covering calendar years 2021 through 2025. Two portfolios are evaluated on the same five liquid U.S.-listed ETFs representing distinct investment exposures: SPY (U.S. large-cap equities), QQQ (Nasdaq-100 equities), GLD (gold), TLT (long-term U.S. Treasury bonds), and VNQ (U.S. real estate). Annual total-return data with distributions reinvested were used so that the comparison reflects total investment performance rather than price movement alone.

AI allocation. The AI-assisted portfolio was generated using a fixed prompt designed to make the AI a complementary allocation tool rather than a price predictor. The prompt was: “You are assisting a moderate-risk long-term investor. Allocate 100% of capital across SPY, QQQ, GLD, TLT, and VNQ. Prioritize diversification across asset classes, avoid leverage, keep each weight between 10% and 35%, and explain the role of each asset. Do not use future market performance or optimize the weights using the 2021-2025 return data.” The resulting allocation used for the backtest was 30% SPY, 30% QQQ, 15% GLD, 10% TLT, and 15% VNQ. The allocation is then held for each calendar year and rebalanced to the stated weights at the beginning of each year.

Traditional benchmark. The comparison portfolio uses a simple equal-weight rule: 20% in each of the five ETFs. This benchmark represents a transparent, non-AI allocation process in which no asset receives a special weight. It is similar in spirit to the naive equal-weight benchmark used in recent LLM-investment research (Perlin et al., 2025).

Variables and formulas. The primary dependent variable is annual portfolio return. Secondary measures are cumulative return, compound annual growth rate (CAGR), sample standard deviation of annual returns, worst annual return, maximum drawdown based on the annual portfolio path, and a simple Sharpe ratio calculated as mean annual portfolio return divided by

sample standard deviation. Because only five annual observations are available, the Sharpe ratio is interpreted descriptively rather than as a statistically definitive estimate. Portfolio return for each year is calculated as the weighted average of the five ETF annual total returns: $R_{p,t} = \sum(w_i R_{i,t})$. Cumulative return is the product of $(1 + R_{p,t})$ across the five years minus one. No leverage, taxes, transaction fees, or borrowing costs are included.

Data sources and reproducibility. Annual ETF total-return figures were taken from public market-data pages that state that their annual series assumes dividends are reinvested. The SPY, QQQ, GLD, TLT, and VNQ series report the annual returns used in the analysis. The calculations were then reproduced with the fixed portfolio weights above. This design keeps the study transparent enough to be checked independently and prevents the results table from relying on an unexplained AI-generated statistic.

Table 1. Portfolio allocations used in the experiment

Asset	Role / Exposure	AI-assisted	Equal-weight
SPY	U.S. large-cap equities	30%	20%
QQQ	Nasdaq-100 / growth-oriented equities	30%	20%
GLD	Gold / alternative diversifier	15%	20%
TLT	Long-term U.S. Treasury bonds	10%	20%
VNQ	U.S. real estate investment trusts	15%	20%
Total		100%	100%

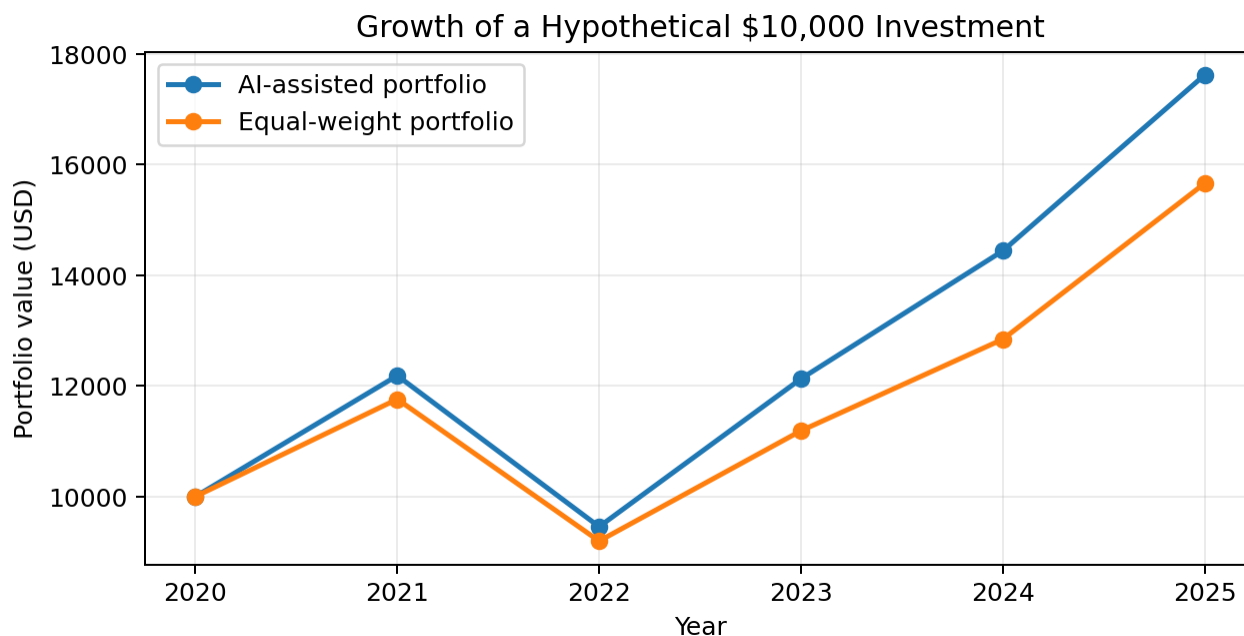
Results:

The backtest produced different outcomes across the five calendar years. The AI-assisted portfolio generated annual returns of 21.84% in 2021, -22.40% in 2022, 28.27% in 2023, 19.06% in 2024, and 22.01% in 2025. The equal-weight portfolio generated 17.59%, -21.80%, 21.67%, 14.78%, and 21.93%, respectively.

The AI-assisted portfolio therefore produced a higher annual return in four of the five years. The largest relative difference occurred in 2023, when the AI-assisted portfolio returned 28.27% compared with 21.67% for the equal-weight portfolio. In 2022, both portfolios experienced a negative year, and the AI-assisted portfolio's loss was slightly larger. In 2025, the two portfolios were nearly identical because the strong GLD contribution partly offset the lower weight assigned to that asset in the AI allocation.

Table 2. Annual portfolio returns, 2021-2025

Year	AI-assisted portfolio	Equal-weight portfolio	Difference (AI - Equal)
2021	21.8%	17.6%	4.3%
2022	-22.4%	-21.8%	-0.6%
2023	28.3%	21.7%	6.6%
2024	19.1%	14.8%	4.3%
2025	22.0%	21.9%	0.1%

Figure 1. Growth of a hypothetical \$10,000 investment, 2020-2025

Note. Portfolio returns are based on annual total returns with distributions reinvested; weights are rebalanced annually to the stated target allocation.

Table 3. Summary performance metrics

Metric	AI-assisted portfolio	Equal-weight portfolio	Observation
Cumulative return	76.2%	56.6%	AI portfolio ended with the larger cumulative gain in this sample.
CAGR	12.0%	9.4%	AI portfolio had the higher compounded annual rate.
Mean annual return	13.8%	10.8%	AI portfolio had the higher arithmetic average.
Annual volatility (sample SD)	20.5%	18.5%	AI portfolio was more volatile.
Worst annual return	-22.4%	-21.8%	AI portfolio had a slightly larger worst-year loss.
Maximum drawdown (annual path)	-22.4%	-21.8%	Drawdown was slightly larger for the AI portfolio.
Simple Sharpe ratio	0.67	0.59	Descriptive only; five annual observations.

Metric	AI-assisted portfolio	Equal-weight portfolio	Observation
Ending value of \$10,000	\$17,616	\$15,657	Difference reflects the compounded return path.

Starting with \$10,000 at the end of 2020, the AI-assisted portfolio grew to approximately \$17,616 by the end of 2025, while the equal-weight portfolio reached approximately \$15,657.

The cumulative returns were 76.16% and 56.57%, respectively. The AI-assisted portfolio's CAGR was 11.99%, compared with 9.38% for the equal-weight portfolio.

The return advantage came with a risk trade-off. The AI-assisted portfolio's sample annual volatility was 20.49%, compared with 18.49% for the equal-weight portfolio. Its worst annual return was -22.40%, compared with -21.80% for the equal-weight portfolio. The simple Sharpe ratio was 0.67 for the AI portfolio and 0.59 for the equal-weight portfolio when using the study's descriptive formulation.

Discussion and Conclusion:

This study examined whether an AI-assisted diversified allocation could outperform a simple equal-weight portfolio over the 2021-2025 period. The results show that, within this specific sample and under the stated assumptions, the AI-assisted allocation generated the higher cumulative return, CAGR, average annual return, and simple Sharpe ratio. At the same time, it also produced slightly higher annual volatility and a slightly worse worst-year loss. Therefore, the results should be interpreted as evidence of a particular allocation rule performing differently from an equal-weight benchmark, not as proof that AI can reliably predict financial markets or outperform traditional portfolio methods in general.

The practical implication is that AI may be useful as a complementary portfolio-construction tool, especially for organizing information, comparing asset roles, and generating an explicit initial allocation that can then be tested quantitatively. This is consistent with Ko and Lee's (2024) view of ChatGPT as a potential assistant or co-pilot while also respecting the caution raised by Perlin et al. (2025) that LLM-based strategies do not consistently outperform simple benchmarks. The value of AI may therefore depend on how the prompt is designed, the investment horizon, the assets selected, and the risk constraints imposed.

The study also has important limitations. First, the experiment uses only five ETFs and five annual observations. Second, the AI allocation is generated from a single structured prompt, so another model, prompt, or risk profile could produce different weights. Third, annual data hide within-year volatility and drawdowns. Fourth, the analysis excludes taxes, trading costs, slippage, currency effects, and investor-specific constraints. Finally, the historical period includes unusual market conditions, including the sharp equity and bond drawdowns of 2022 and the strong performance of gold in 2025, so the result should not be assumed to persist in another period.

Future research could extend the sample to monthly or daily observations, compare several AI systems and repeated prompts, incorporate transaction costs, test different risk profiles, and evaluate whether AI-generated allocations remain robust across bull and bear markets. A larger

study could also compare AI allocation against mean-variance optimization, risk-parity, and professionally managed benchmark portfolios. Such extensions would provide a stronger basis for determining when AI adds measurable value to portfolio management.

AI Use as a Complementary Research Tool

AI was used in this project for brainstorming the research topic, generating and documenting the initial asset-allocation proposal under explicit constraints, assisting with the structure of the paper, and checking the clarity of the presentation. The quantitative results were calculated from the stated historical return data and the fixed portfolio weights rather than being accepted from an AI-generated claim without calculation.

References

- Beainy, R., & Kamel, C. (2025). USA Versus China from an Equity Perspective. *International Journal of Research and Innovation in Social Science*, 9(10), 8191-8201.
- Beainy, R., & Kamel, C. (2026). Gold and Bitcoin: A Quantitative Comparative Analysis. *The Banking and Finance Review*, 17(1).
- Beainy, R., Kamel, C., Subrahmanyam, S., & Younis, M. (2025). The World needs the USD. Available at SSRN 5734284.
- Bejjani, G., Toros, E., Elia, J., & Jabbour Al Maalouf, N. (2024). Technology analytic adeptness for examining the interrelatedness between the main macroeconomic variables in Lebanon employing a vector autoregression (VAR). *Journal of Decision Systems*, 1-30.
- Dakalbab, F., Abu Talib, M., Nasir, Q., & Saroufil, T. (2024). Artificial intelligence techniques in financial trading: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, 36(3), 102015.
<https://doi.org/10.1016/j.jksuci.2024.102015>
- Diab, B., & El Hajj, M. (2024). Ethics in the age of artificial intelligence: unveiling challenges and opportunities in business culture. *Cogent Business & Management*, 11(1).
<https://doi.org/10.1080/23311975.2024.2408440>
- Jabbour Al Maalouf, N., Sfeir, L., Nehme, A., & Mawad, J. L. (2026). Gold as a Household Financial Resilience Strategy: Explaining Gold Purchase Intentions in Lebanon's Fragile Economy. *Journal of Risk and Financial Management*, 19(8), 602.
<https://doi.org/10.3390/jrfm19080602>
- Khalife, D., Yammine, J., El Bazi, T., Zaki, C., & Jabbour Al Maalouf, N. (2025). Analyzing the impact of investor sentiment on S&P 500 prices using deep learning models. *Journal of Financial Reporting and Accounting*, 23(4), 1365-1386.
- Khalife, D., Yammine, J., Rahal, S., & Freiha, S. (2023). Pricing Asian and Barrier Options Using a Combined Heston Model and Monte Carlo Simulation Approach with Artificial Intelligence. *Mathematical Modelling of Engineering Problems*, 10(5), 1690.

- Ko, H., & Lee, J. (2024). Can ChatGPT improve investment decisions? From a portfolio management perspective. *Finance Research Letters*, 64, 105433.
<https://doi.org/10.1016/j.frl.2024.105433>
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91.
<https://doi.org/10.2307/2975974>
- Mertzanis, C. (2025). Artificial intelligence and investment management: Structure, strategy, and governance. *International Review of Financial Analysis*, 104599.
<https://doi.org/10.1016/j.irfa.2025.104599>
- Nasser, A., Yammine, J., Harouny, J., & Khalife, D. (2025). Predicting Economic Quads Through Asset Returns Using Ensemble Machine Learning. *IEEE Access*.
- Perlin, M. S., Foguesatto, C. R., Müller, F. M., & Righi, M. (2025). Can AI beat a naive portfolio? An experiment with anonymized data. *Finance Research Letters*, 78, 107126.
<https://doi.org/10.1016/j.frl.2025.107126>
- Schneider, C. J., & Yilmaz, Y. (2025). Stock portfolio selection based on risk appetite: Evidence from ChatGPT. *Finance Research Letters*, 82, 107517.
<https://doi.org/10.1016/j.frl.2025.107517>
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1), 119-138.
<https://doi.org/10.1086/294846>
- Subrahmanyam, S., & Beainy, R. H. (2027). Smart Decision Support for Uncertainty and Disruptions in Industry 4.0 Operations. In *Smart Operational Research Solutions in the Era of Industry 4.0 and Sustainability* (pp. 167-198). IGI Global Scientific Publishing.
- Subrahmanyam, S., Aishwaryalaxmi, N. S., Khalife, D., Shaikh, I. A. K., Faldu, R., & Asthana, N. (2024). Impact of knowledge management and big data analytics capabilities on firm performance. In *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. 1-5). IEEE.
- Total Real Returns. (2026). Historical total-return data for SPY, QQQ, GLD, TLT, and VNQ (dividends reinvested). Retrieved September 29, 2026, from <https://totalrealreturns.com/>
- Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, 61(4), 1281–1317.