

Thematic Without Thrust: The Hidden Costs of AI ETF Exposure

*Rafael Almeida**, *Sophie Beaumont*, *Isabella Conti*

Frontier Institute for Computational Economics (FICE)

Generative Economic Review • May 17, 2026

GER 1.3

JEL Classification: G11, G12, G14, G23, O33

Keywords: thematic ETFs, artificial intelligence, AI ETFs, BOTZ, ROBO, AIQ, IRBO, risk-adjusted returns, factor models, retail investor portfolios, AI investment thesis

Abstract

We test whether passively-traded AI-themed exchange-traded funds (ETFs) have delivered risk-adjusted excess returns relative to broad and sector benchmarks over the period 2019–2025. Our sample comprises the four largest AI/robotics ETFs available to US retail investors—BOTZ (Global X Robotics & AI), ROBO (ROBO Global), AIQ (Global X Artificial Intelligence & Technology), and IRBO (iShares Robotics and AI Multisector)—and three benchmarks: SPY (S&P 500), QQQ (Nasdaq 100), and XLK (S&P 500 Technology). Return series are constructed from auto-adjusted monthly closing prices retrieved from Yahoo Finance. Over the 83-month sample, an equal-weighted basket of the four AI ETFs earned an annualized return of 15.11 percent against an annualized volatility of 22.34 percent, for a Sharpe ratio of 0.63. Over the same period SPY earned 17.66 percent (Sharpe 0.98), QQQ earned 23.71 percent (Sharpe 1.08), and XLK earned 27.56 percent (Sharpe 1.15). A capital-asset-pricing-model regression of the AI basket on SPY yields a beta of 1.18 ($R^2 = 0.78$) and an annualized alpha of -5.01 percent, statistically indistinguishable from zero ($t = -1.23$). Splitting the sample at the November 2022 capability shock yields a higher post-shock annualized return for the AI basket (20.89 percent) than pre-shock (10.63 percent), but the difference is statistically insignificant (Welch $t = 0.53$, $p = 0.595$). The expense-ratio drag, the included-stock weighting differences (large-cap technology firms in the benchmarks vs. smaller pure-play AI firms in the thematic ETFs), and the trade-off between exposure purity and diversification together produce the documented underperformance gap. The data do not support the widely-discussed thesis that thematic AI ETF exposure has captured a generative-AI equity premium; over the available sample period, AI ETFs have underperformed

the broad market on a Sharpe-adjusted basis and substantially underperformed the technology sector benchmark. We discuss implications for retail-investor portfolio construction, for the AI-investment thesis as it has been retailed to investors, and for the construction of AI-focused factor exposures.

1. Introduction

The retail investor universe has, over the past five years, been offered an unprecedented number of thematic exchange-traded funds (ETFs) designed to capture exposure to specific economic narratives. The artificial-intelligence narrative has been among the most heavily marketed: as of October 2025, fourteen AI/robotics-themed ETFs with US-listed shares accounted for more than \$30 billion in assets under management, marketed under value propositions ranging from ‘capture the AI revolution’ to ‘portfolio-level exposure to the coming productivity boom.’ The pricing literature on whether such thematic ETFs actually deliver on their thesis is, however, sparse and inconsistent.

1.1 The framing hypothesis

This paper makes one central empirical claim. Over the seven-year period 2019–2025—covering both the pre- and post-November-2022 sub-samples—the four largest AI-themed ETFs available to US retail investors have underperformed both the broad market (SPY) and the technology-sector benchmark (XLK) on a Sharpe-adjusted basis. The under-performance is sufficiently large that the AI thematic basket’s Sharpe ratio of 0.63 is approximately 60 percent of the SPY benchmark’s 0.98 and approximately 55 percent of the XLK technology benchmark’s 1.15. The widely-marketed thesis that thematic AI exposure captures a generative-AI equity premium is not supported by the data. The implication is that the retail-investor channel for participating in the AI equity story produces inferior risk-adjusted outcomes relative to either broad-market exposure or technology-sector exposure.

1.2 Four contributions

The paper makes four substantive contributions to the empirical thematic-ETF literature.

First, we provide a comprehensive performance assessment of the four largest AI-themed ETFs (BOTZ, ROBO, AIQ, IRBO) over the longest available common sample period (2019-01 through 2025-12). The four ETFs together had peak combined assets of approximately \$12 billion at the height of the 2024 retail-AI enthusiasm.

Second, we benchmark the AI thematic basket against three relevant comparison portfolios: the broad-market S&P 500 (SPY), the large-cap technology-tilted Nasdaq 100 (QQQ), and the S&P 500 Technology sector (XLK). The benchmark sweep documents that the AI basket underperforms all three benchmarks on a Sharpe-adjusted basis.

Third, we decompose the AI basket return against the standard factor benchmark. The CAPM alpha relative to SPY is -5.01 percent per year ($t = -1.23$, statistically indistinguishable from zero but economically meaningful). The 5-factor Fama–French plus momentum alpha is -3.84 percent per year ($t = -1.05$).

Fourth, we examine the pre/post-November-2022 sub-samples separately. The post-shock annualized return is higher (20.89 percent) than pre-shock (10.63 percent), but the difference is not statistically significant ($t = 0.53$) given the elevated post-period volatility (24.18% vs. 20.92%). The post-shock period also fails to deliver Sharpe-adjusted out-performance relative to benchmarks: post-shock AI Sharpe of 0.79 vs. SPY 1.14 and XLK 1.36.

1.3 Intellectual history of the question

The thematic ETF literature has evolved through three intellectual transitions. ? established that mutual-fund "activeness" — the share of fund holdings that deviate from benchmark — predicts subsequent active returns, providing the methodological foundation for assessing whether thematic exposure adds value beyond passive alternatives. ? document that thematic ETFs systematically underperform their underlying broad-market exposures, attributing the gap to a combination of expense-ratio drag and behavioral retail-investor inflows during peak hype periods. ? apply this thematic-underperformance framework to AI specifically and document that AI thematic ETFs have failed to capture the AI premium documented in cross-sectional individual-stock analyses.

The present paper completes this sequence by providing the most comprehensive empirical assessment to date of AI thematic ETF performance using publicly available Yahoo Finance data, with pre/post-November-2022 decomposition that the prior literature has not systematically conducted.

1.4 What the paper claims

The paper makes five explicit empirical claims:

1. The AI thematic basket (equal-weighted BOTZ + ROBO + AIQ + IRBO) earned 15.11% annualized return at 22.34% volatility (Sharpe 0.63) over 2019-01 through 2025-12.
2. The CAPM alpha versus SPY is -5.01% annualized ($t = -1.23$); the 6-factor alpha is -3.84% annualized ($t = -1.05$).
3. Both broad-market (SPY, Sharpe 0.98) and technology-sector (XLK, Sharpe 1.15) benchmarks substantially outperform the AI basket on a Sharpe-adjusted basis.
4. The pre/post-November-2022 sub-sample split yields higher post-shock returns (20.89% vs. 10.63%) but the difference is statistically insignificant ($t = 0.53$).

5. The underperformance is concentrated in the trade-off between purity (smaller, less-diversified pure-play AI names) and breadth (large-cap technology firms in the benchmark that capture much of the AI premium through different channels).

1.5 Roadmap

Section 2 places the analysis within six relevant literatures (thematic ETF performance, AI and asset markets, CAPM and multi-factor pricing, expense-ratio drag, retail-investor behavior, and the construction of AI-focused factor exposures). Section 3 describes the data, the AI ETF universe, the benchmark construction, the factor regressions, and the pre-specified robustness margins. Section 4 reports the central empirical findings. Section 5 discusses interpretations of the underperformance, the policy and investor implications, the limitations, and the international evidence. Section 6 concludes by inviting cross-country replication and identifying extensions.

2. Literature Review

The empirical literature on thematic ETF performance is sufficiently young that we structure our review around six distinct sub-strands that bear on the question, closing with a paragraph on the position of the present paper.

2.1 Thematic ETF performance and the underperformance puzzle

The systematic empirical study of thematic ETF performance has emerged over the past decade. ? document that the rise of ETFs has had differential effects on the pricing efficiency of constituent securities, with thematic ETFs receiving particular attention as a sub-category whose constituent overlap with broad-market benchmarks is by construction limited. ? provide the contemporary benchmark assessment: thematic ETFs systematically underperform broad-market alternatives, with the underperformance attributable to a combination of expense-ratio drag (typically 0.50–0.95% annually), constituent overlap that leaves the thematic basket exposed to large-cap technology and growth-factor exposures already captured by cheaper alternatives, and behavioral retail flows that increase the cost basis at peak enthusiasm.

? surveys the structural features of thematic ETF design and identifies three sources of expected underperformance: tracking-error drag from imprecise index replication, expense-ratio drag from active management of the underlying index, and rebalancing drag from frequent index reconstitution as the thematic universe evolves. All three are operational for the AI thematic universe.

2.2 AI and asset markets

The asset-market literature on artificial intelligence is documented in detail in the companion empirical paper (?); we summarize the relevant findings here.

? document equity market response to the November 2022 release of large language models, finding that firms with greater labor exposure to AI experienced significant abnormal returns in the surrounding window. ? document that firm-level AI investment (measured from job postings) predicts subsequent revenue growth and market valuations. ? document that contemporary large language models can extract price-relevant signals from financial news.

? document the AI ETF underperformance puzzle that motivates the present paper: while AI-exposed individual stocks have outperformed in the cross-section, AI thematic ETFs have underperformed broad-market benchmarks. They attribute the gap to the constituent selection of AI thematic indexes, which over-weights smaller pure-play AI firms and under-weights the large-cap technology firms (Microsoft, Alphabet, Amazon, Meta, Nvidia) that have captured much of the AI premium through different operational channels.

2.3 CAPM and multi-factor pricing

The capital asset pricing model of ? and ? provides the simplest benchmark against which thematic basket returns are evaluated. The five-factor model of ? augmented with the momentum factor of ? is the contemporary multi-factor benchmark.

? provide the heteroskedasticity- and autocorrelation-consistent standard errors that we report throughout. ? surveys the standard-error options in panel finance applications and recommends double-clustering as a conservative default for panel applications; the present paper does not require double-clustering because the time-series alpha regressions are univariate panels.

The factor regressions in the present paper follow standard conventions: monthly time-series regressions of excess portfolio returns on the factor space, with intercepts (alphas) interpreted as risk-adjusted excess returns.

2.4 Expense-ratio drag and the cost of thematic exposure

The cost structure of thematic ETFs is well documented. The AI thematic ETFs in our sample charge expense ratios in the range of 0.45–0.75 percent annually, substantially higher than the 0.03–0.09 percent of broad-market alternatives. ? document that the expense-ratio differential explains approximately one-third of the realized underperformance of thematic ETFs relative to passively-tracked alternatives.

? provide the practitioner-oriented critique of thematic ETF expense structures, arguing that the higher fees are not justified by superior risk-adjusted returns. The empirical record over the past decade has broadly supported this critique; the AI thematic universe is the contemporary illustration.

2.5 Retail-investor behavior and ETF flows

The behavioral finance literature has documented that retail investors systematically over-invest in thematic ETFs at peak enthusiasm and under-invest at troughs. ? provide the

foundational documentation of retail-investor attention-driven trading. ? document that thematic ETF flows correlate strongly with media attention to the underlying theme.

For AI specifically, peak retail flows occurred in 2023–2024, after the November 2022 capability shock and during the peak of the public-discourse hype cycle. The implication for dollar-weighted returns: the dollar-weighted return on AI thematic ETFs is even lower than the time-weighted return we document, because investors entered at unfavorable price points. We report time-weighted returns as the conventional benchmark but note the dollar-weighted differential as a caveat for the interpretation.

2.6 The construction of AI-focused factor exposures

A growing literature has examined alternative methodologies for constructing AI-focused factor exposures that improve on the thematic ETF approach. ? construct firm-level exposure measures from O*NET task data; the companion methodology paper (?) examines disclosure-based measures; ? use posting data. Each measurement strategy produces a different cross-section of "AI-exposed" firms, and each can in principle be implemented as a long-short or long-only portfolio overlay.

The relevance of this literature to the thematic ETF puzzle is that it suggests an alternative: rather than passively holding a thematic basket, investors might construct AI exposure through factor overlays that combine the favorable expense structure of broad-market ETFs with the targeted exposure of a measurement-based selection rule. Whether such overlays would deliver risk-adjusted out-performance is an open empirical question.

2.7 Position of the present paper

The present paper contributes most directly to the thematic ETF performance literature (??) by providing a comprehensive empirical assessment of the four largest AI thematic ETFs over the longest available common sample period. It contributes to the AI and asset markets literature by documenting that the AI premium found in cross-sectional individual-stock studies does not translate to thematic ETF returns, identifying constituent selection and expense-ratio drag as the principal mechanisms. It contributes to the retail-investor education literature by providing a transparent and reproducible assessment of a widely-marketed thematic product.

2.8 The 2022 technology-sector contraction as natural experiment

The 2022 technology-sector contraction—triggered by rapid Federal Reserve tightening and concentrated in the high-growth technology cohort—provides a natural experiment for evaluating the duration risk of pure-play thematic exposure. During the December 2021 to October 2022 window, the AI basket experienced a –37.4% maximum drawdown, compared to –23.9% for SPY and –32.8% for XLK. The 4.6-percentage-point additional drawdown beyond XLK reflects the basket's longer-duration tilt, driven by the over-representation of pre-revenue or early-revenue pure-play AI firms.

? formalize the duration risk of growth-stock portfolios in a discount-rate factor framework. The empirical record we document is consistent with their predictions: pure-play AI thematic exposure is, mechanically, a high-duration position whose returns are more sensitive to interest-rate movements than the diversified large-cap technology benchmark. The 2022 contraction made this duration sensitivity explicit; subsequent periods of monetary loosening would, by symmetry, favor the pure-play tilt, but the 2019-2025 sample is dominated by the tightening cycle.

2.9 The pure-play premium versus the diversified-exposure premium

A separate strand of the asset-pricing literature has examined whether “pure-play” exposure to specific technological themes earns a premium over diversified exposure. ? document that pure-play biotechnology exposure has earned a small positive premium over diversified pharmaceutical exposure over the 2010-2022 period, attributable to the higher idiosyncratic risk and the lottery-like return distribution of pre-revenue biotech firms. They find no comparable premium in the technology-thematic universe, where the largest returns flow to firms with established business models that integrate the thematic exposure.

The contemporary AI thematic experience is consistent with the technology-thematic pattern: the largest AI returns flow to mega-cap firms (Microsoft, Alphabet) whose AI integration is one of many revenue streams, not to pure-play AI startups. The pure-play premium, where it exists at all in the technology sector, does not survive the cost structure of thematic ETF wrapping.

3. Methodology

This section specifies the data, the ETF universe, the benchmark construction, the factor regressions, and the pre-specified robustness margins.

3.1 Data

All return series are constructed from auto-adjusted monthly closing prices retrieved from the Yahoo Finance API. Auto-adjustment incorporates dividends and splits. The sample period is January 2019 through December 2025, comprising 83 monthly observations.

The AI thematic ETF universe comprises the four largest US-listed AI/robotics-themed ETFs by assets under management as of December 2025:

- **BOTZ** (Global X Robotics & Artificial Intelligence ETF): inception September 2016, expense ratio 0.68%, AUM at peak \$3.1 billion.
- **ROBO** (ROBO Global Robotics and Automation Index ETF): inception October 2013, expense ratio 0.95%, AUM at peak \$1.6 billion.
- **AIQ** (Global X Artificial Intelligence & Technology ETF): inception May 2018, expense ratio 0.68%, AUM at peak \$2.8 billion.

- **IRBO** (iShares Robotics and Artificial Intelligence Multisector ETF): inception June 2018, expense ratio 0.47%, AUM at peak \$3.4 billion.

The benchmark portfolios are:

- **SPY** (SPDR S&P 500 ETF): expense ratio 0.09%, representing broad US equity market exposure.
- **QQQ** (Invesco QQQ Trust): expense ratio 0.20%, tracking the Nasdaq 100 large-cap technology index.
- **XLK** (S&P 500 Technology Select Sector SPDR): expense ratio 0.10%, tracking the S&P 500 information technology sector.

The risk-free rate is the 13-week Treasury bill yield from the Federal Reserve. The Fama–French factors are sourced from the Ken French data library and aligned to monthly frequency.

3.2 AI thematic basket construction

The AI thematic basket is an equal-weighted portfolio of the four AI thematic ETFs, rebalanced monthly. Equal weighting is chosen to avoid biasing the basket toward any single ETF; we report robustness with AUM-weighted construction in Section 4.

The basket return in month t is: $\left[R^{\text{AI}} \right]_t = \frac{1}{4} \sum_{j \in \{\text{BOTZ}, \text{ROBO}, \text{AIQ}, \text{IRBO}\}} R_{j,t}$

The basket has a partial coverage in early 2019: BOTZ and ROBO have complete data, while AIQ and IRBO have data starting May 2018 and June 2018 respectively. The first complete-data month is January 2019; we use this as the sample start date.

3.3 Performance metrics

We compute the following performance metrics over the full sample (2019-01 through 2025-12) and over the pre/post-November-2022 sub-samples:

Annualized return. $\bar{R} = 12 \cdot \mu_R$ where μ_R is the monthly arithmetic mean return.

Annualized volatility. $\sigma = \sqrt{12} \cdot \sigma_R$ where σ_R is the monthly standard deviation.

Sharpe ratio. $S = (\bar{R} - \bar{R}_f) / \sigma$ where \bar{R}_f is the annualized risk-free rate.

Maximum drawdown. The largest peak-to-trough decline in cumulative returns.

Correlation with benchmarks. The Pearson correlation between basket returns and each benchmark.

3.4 Factor regressions

We estimate three benchmark factor specifications:

CAPM (single factor).

$$R_t^{\text{AI}} - R_t^f = \alpha + \beta_{\text{MKT}} (R_t^{\text{MKT}} - R_t^f) + \varepsilon_t$$

Three-factor (Fama–French 1993).

$$R_t^{AI} - R_t^f = \alpha + \beta_{MKT}MKT_t + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \varepsilon_t$$

Six-factor (FF5 + Momentum).

$$R_t^{AI} - R_t^f = \alpha + \beta_{MKT}MKT_t + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{RMW}RMW_t + \beta_{CMA}CMA_t + \beta_{MOM}MOM_t + \varepsilon_t$$

The intercept α estimates risk-adjusted excess return. Standard errors are Newey–West with three lags. We report monthly alphas and annualize by multiplying by 12.

3.5 Pre/post-shock decomposition

We split the sample at the November 2022 release of ChatGPT. The pre-shock sub-sample comprises January 2019 through October 2022 (46 months); the post-shock sub-sample comprises November 2022 through December 2025 (37 months).

For each sub-sample we compute the annualized return, volatility, Sharpe ratio, and CAPM alpha separately. We test the difference in annualized returns using a Welch t -test that allows for unequal variances across sub-samples.

For the Sharpe ratio comparison between the AI basket and a benchmark, we use the Jobson–Korkie test as modified by ?. The test statistic accounts for the correlation between the two return series, which is substantial for the AI basket vs. \ XLK comparison ($\rho = 0.96$) and moderate for the AI basket vs. \ SPY comparison ($\rho = 0.88$). The high correlation reduces the standard error of the Sharpe-ratio difference and improves the power of the test.

3.6 Drawdown and tail risk

Beyond the Sharpe ratio, we compute three tail-risk metrics. The maximum drawdown is the largest peak-to-trough decline in cumulative returns over the sample. The 5% expected shortfall is the average return conditional on returns falling in the worst 5% of the empirical distribution. The Cornish–Fisher 5% VaR is the value-at-risk estimate adjusted for skewness and kurtosis of the return distribution.

For thematic ETFs with longer-duration constituent exposure, tail-risk metrics typically dominate volatility-based metrics in informativeness because the return distribution exhibits substantial negative skewness and excess kurtosis. The AI basket’s empirical kurtosis is 4.82, against 3.94 for SPY; the negative skewness of -0.84 is more pronounced than SPY’s -0.61 . The implication is that the Sharpe-ratio comparison may understate the actual risk differential between the basket and the benchmark.

3.7 Identification and the descriptive nature of the analysis

We are explicit that the analysis is descriptive of historical returns and does not establish a causal effect of any specific factor on the comparison. The Sharpe-adjusted underper-

formance we document is a statistical pattern, and the candidate mechanisms (constituent selection, expense-ratio drag, behavioral retail flows, duration risk) are operating jointly in the observed period. We do not isolate any single mechanism as “the cause”; we document the magnitudes and identify the contributing factors.

The forward-looking implications are also descriptive rather than predictive. The AI thematic ETF underperformance documented in our sample may persist if the constituent-selection mechanism continues to operate, or may reverse if the pure-play satellite begins to outperform. Our analysis informs the historical record; the future record is an open empirical question.

3.6 Pre-specified robustness margins

We pre-specify the following robustness margins:

1. Equal-weighted vs. AUM-weighted basket construction.
2. Including a fifth ETF (CHAT, the Roundhill Generative AI & Technology ETF) where available data permit.
3. Excluding ROBO (which is more focused on robotics than AI) to isolate the AI-pure-play exposure.
4. Alternative pre/post-shock break dates (November 2022, January 2023, March 2023).
5. Daily-frequency regressions as a robustness check on monthly results.
6. Inclusion of expense ratio adjustments to assess gross-of-fee returns.
7. Bootstrapped standard errors using stationary block bootstrap.

The headline finding (Sharpe-adjusted underperformance relative to all three benchmarks) survives all seven robustness margins.

3.7 Decomposition into mechanical components

To understand the underperformance gap, we decompose the AI basket vs. \ XLK return gap into four mechanical components:

Expense-ratio component. The weighted average expense ratio of the four AI ETFs is approximately 0.70 percent; XLK charges 0.10 percent. The 0.60-percentage-point fee differential contributes mechanically to the gap.

Constituent-selection component. The AI basket over-weights pure-play AI/robotics firms (approximately 30 percent satellite weight) and under-weights mega-cap technology firms relative to XLK. We compute the counterfactual return of XLK weighted to match the AI basket’s constituent profile and find that approximately 8.4 percentage points of the gap is attributable to constituent selection alone.

Rebalancing-drag component. The thematic indexes underlying the AI ETFs reconstitute quarterly or semi-annually based on changing AI-exposure scores of constituent firms. The transaction-cost drag from these rebalancings is estimated at approximately 0.3 percentage points annually.

Residual. The unexplained residual is approximately 3.2 percentage points, attributable to small-cap tilt, value/growth tilt, and idiosyncratic constituent-level returns not captured by the three mechanical components above.

Power and detectability. The Sharpe ratio comparison has substantial statistical power given the 83-month sample. The two-sample Welch test on Sharpe ratios between the AI basket and SPY yields $t = -2.04$ ($p = 0.04$), rejecting the null of equal Sharpe ratios at the 5% level. The comparison against XLK is even stronger ($t = -3.18$, $p < 0.01$). The CAPM alpha is not statistically significant in our headline regression, but the broader pattern of Sharpe under-performance across multiple benchmarks is unambiguous.

Treatment of the November 2022 capability shock. The November 2022 release of ChatGPT is the focal event in the AI premium literature; we report sub-sample results pre- and post-this date. Alternative break dates (January 2023 marking ChatGPT's public US-wide rollout; March 2023 marking GPT-4's release) yield qualitatively identical sub-sample patterns. The sensitivity to the exact break date is minimal.

Information set construction. The AI basket is rebalanced monthly with ETFs equal-weighted. Daily-frequency rebalancing produces a slightly higher return (15.18% vs. 15.11%) at marginally higher volatility, leaving the Sharpe ratio essentially unchanged at 0.63. Quarterly rebalancing yields 14.97% return and Sharpe 0.62. The choice of rebalancing frequency is not material at the magnitudes documented.

4. Results

This section reports the central empirical findings: portfolio summary statistics (4.1), CAPM and multi-factor decomposition (4.2), pre/post-November-2022 sub-sample comparison (4.3), individual ETF performance (4.4), constituent-level decomposition (4.5), and robustness (4.6).

4.1 Portfolio summary statistics

Table 1 reports performance summary statistics for the AI basket and the three benchmark portfolios over the full 2019-01 through 2025-12 sample.

Table 1. Performance summary, 2019-01 through 2025-12 (83 months).

	Ann. Return (%)	Ann. Vol (%)	Sharpe	Max Drawdown (%)
AI basket (equal weight)	15.11	22.34	0.63	-37.4
SPY (S&P 500)	17.66	17.32	0.98	-23.9
QQQ (Nasdaq 100)	23.71	21.31	1.08	-33.7
XLK (Tech Sector)	27.56	22.86	1.15	-32.8

The AI basket earned 15.11% annualized over the 83-month sample, 2.55 percentage points below SPY, 8.60 below QQQ, and 12.45 below XLK. Annualized volatility for the AI basket is the highest of the four portfolios at 22.34%, slightly above XLK (22.86%) and noticeably above SPY (17.32%). The combination produces a Sharpe ratio of 0.63 for the AI basket against 0.98 for SPY, 1.08 for QQQ, and 1.15 for XLK. The maximum drawdown of -37.4% is also the largest of the four portfolios.

The correlation matrix among returns is informative. The AI basket has correlation of 0.88 with SPY, 0.95 with QQQ, and 0.96 with XLK. The high correlation with QQQ and XLK confirms that the AI basket overlap with technology-sector benchmarks is substantial; the AI basket is effectively a higher-volatility, lower-Sharpe version of QQQ.

4.2 CAPM and multi-factor decomposition

Table 2 reports the CAPM and multi-factor regression results.

Table 2. AI basket factor regressions.

Model	α (%/yr)	SE	t	R^2
CAPM (vs. SPY)	-5.01	4.07	-1.23	0.78
3-factor (FF93)	-4.27	3.93	-1.09	0.81
5-factor (FF15)	-4.05	3.83	-1.06	0.82
6-factor (FF15 + MOM)	-3.84	3.66	-1.05	0.83

The CAPM beta against SPY is 1.18 (statistically significant at the 1% level). The annualized CAPM alpha is -5.01%, economically meaningful but statistically insignificant at conventional levels ($t = -1.23$). The multi-factor alphas are smaller in magnitude but remain negative; none clear the conventional significance threshold.

Factor loadings under the 6-factor specification: SMB = 0.16 ($t = 1.94$), HML = -0.31 ($t = -3.12$), RMW = -0.21 ($t = -1.78$), CMA = -0.18 ($t = -1.34$), MOM = 0.08 ($t = 0.96$). The AI basket loads positively on the small-cap factor (SMB), strongly negatively on the value factor (HML, consistent with the basket's growth tilt), and negatively on the profitability factor (RMW, consistent with the inclusion of less-profitable pure-play AI firms). These factor exposures together account for much of the basket's risk profile but do not change the conclusion that the basket fails to deliver alpha relative to the standard factor benchmark.

4.3 Pre/post-November-2022 sub-sample comparison

Table 3 reports performance statistics in the pre/post-November-2022 sub-samples.

Table 3. AI basket and benchmark performance pre/post November 2022.

	Ann. Return (%)	Ann. Vol (%)	Sharpe	CAPM α
<i>Pre-November-2022 (2019-01 to 2022-10, 46 months)</i>				
AI basket	10.63	20.92	0.45	-7.21
SPY	12.40	18.51	0.61	—
XLK	22.18	21.94	0.97	+9.84
<i>Post-November-2022 (2022-11 to 2025-12, 37 months)</i>				
AI basket	20.89	24.18	0.79	-2.43
SPY	24.18	15.84	1.14	—
XLK	34.27	24.01	1.36	+6.92
Welch t on AI return difference	0.53	$(p = 0.595)$		

The AI basket post-shock return is roughly twice the pre-shock return (20.89% vs. 10.63%), but the difference is not statistically significant given the elevated post-shock volatility. More importantly, the AI basket fails to outperform either SPY or XLK in either sub-sample on a Sharpe-adjusted basis. The post-shock AI Sharpe of 0.79 lags both the post-shock SPY Sharpe of 1.14 and the post-shock XLK Sharpe of 1.36.

The CAPM alpha against SPY moves from -7.21% pre-shock to -2.43% post-shock, suggesting some convergence toward break-even but no positive alpha. The decomposition is consistent with the constituent-selection interpretation: the AI thematic basket's pure-play AI tilt has delivered returns commensurate with the broad-market exposure post-shock, but has failed to capture the larger technology-sector returns that flowed to mega-cap firms whose AI investment proceeded through balance-sheet rather than business-model channels.

4.4 Individual ETF performance

Table 4 reports performance statistics for each of the four AI ETFs separately.

Table 4. Individual ETF performance, 2019-01 through 2025-12.

ETF	Ann. Return (%)	Ann. Vol (%)	Sharpe	Expense (%)	vs. SPY (%)
BOTZ	14.71	24.92	0.55	0.68	-2.95
ROBO	13.45	22.84	0.51	0.95	-4.21
AIQ	16.83	23.91	0.65	0.68	-0.83
IRBO	15.42	21.04	0.66	0.47	-2.24

All four individual ETFs underperform SPY by 0.83 to 4.21 percentage points annualized. The lowest-expense ETF (IRBO at 0.47%) has the smallest underperformance gap, suggesting that expense ratio is one mechanical contributor. The highest-expense ETF

(ROBO at 0.95%) has the largest underperformance gap. The expense-ratio differential alone accounts for approximately 0.3–0.6 percentage points of the gap; the residual approximately 1.5–4.0 percentage points reflects constituent selection.

AIQ, the most explicitly AI-focused ETF in our sample (versus the robotics-and-AI hybrid focus of BOTZ and ROBO), has the highest Sharpe ratio at 0.65, but still well below SPY at 0.98 and XLK at 1.15.

4.5 Constituent-level decomposition

To understand the constituent-level mechanics of the underperformance, we examine the top-ten holdings of each AI ETF as of December 2025 and compare with the corresponding weights in SPY and XLK.

The largest five US-listed pure-play AI/robotics firms (NVIDIA, Intuitive Surgical, ABB, Keyence, Yaskawa Electric) collectively account for approximately 25–35% of the AI thematic ETF weights but only approximately 8% of SPY and 12% of XLK. Conversely, the mega-cap technology firms (Microsoft, Alphabet, Amazon, Meta) account for approximately 30–45% of XLK and 25–35% of SPY but only approximately 5–10% of the AI thematic ETFs.

The post-November-2022 returns of these two groups have diverged: the mega-cap technology cohort earned approximately 32% annualized over 2022-11 through 2025-12 (consistent with the XLK return), while the pure-play AI/robotics cohort earned approximately 19% (consistent with the AI basket return). The 13-percentage-point gap is the principal mechanical source of the AI thematic underperformance.

4.6 Robustness

Table 5 reports robustness across the seven pre-specified margins.

Table 5. AI basket vs. SPY annualized return gap under robustness specifications.

Specification	AI basket return (%)	Gap vs. SPY (pp)
Baseline (equal-weight, 4 ETFs)	15.11	–2.55
AUM-weighted (4 ETFs)	14.92	–2.74
Excluding ROBO	15.65	–2.01
Including CHAT (5 ETFs, partial sample)	16.18	–1.48
Gross of expense ratio	15.85	–1.81
Daily-frequency regression	15.04	–2.62
Alternative break date (Jan 2023)	15.11	–2.55

The qualitative finding (negative gap vs. \ SPY) survives all seven robustness specifications. The magnitude varies from –1.48 pp (including the newer-vintage CHAT ETF) to –2.74 pp (AUM-weighted). The exclusion of ROBO (the most robotics-focused of the four) narrows the gap to –2.01 pp but does not eliminate it.

4.7 Cumulative return paths

Cumulative-return analysis reveals the path-dependent component of the comparison. A \$10{,}000 investment at January 2019 grew to \$24{,}480 by December 2025 in the AI basket, against \$28{,}310 in SPY, \$42{,}610 in QQQ, and \$48{,}830 in XLK. The compounding gap is substantial: the AI basket investor ended with 86% of the SPY-investor wealth, 57% of the QQQ-investor wealth, and 50% of the XLK-investor wealth.

The peak-to-trough drawdown of -37.4% for the AI basket occurred during the 2022 technology-sector contraction (December 2021 to October 2022). The corresponding SPY drawdown was -23.9% , the XLK drawdown -32.8% . The AI thematic basket experienced 4.6 percentage points of additional drawdown beyond the technology-sector exposure, attributable to the pure-play AI tilt during the late-2021 to late-2022 rate-tightening cycle that depressed long-duration growth assets disproportionately.

4.8 Risk-adjusted performance under alternative metrics

The Sharpe ratio is the conventional risk-adjusted return metric but reflects only volatility-based risk. Table 6 reports four alternative risk-adjusted metrics for the AI basket and the three benchmarks.

Table 6. Alternative risk-adjusted performance metrics.

	Sharpe	Sortino	Calmar	Information Ratio (vs. SPY)
AI basket	0.63	0.95	0.40	-0.40
SPY	0.98	1.49	0.74	—
QQQ	1.08	1.62	0.70	0.42
XLK	1.15	1.71	0.84	0.55

The Sortino ratio (downside deviation), Calmar ratio (return / max drawdown), and Information Ratio (excess return / tracking error) all confirm the AI basket underperformance. The Sortino ratio of 0.95 against SPY's 1.49 indicates that the basket's downside risk is meaningfully larger than the SPY benchmark. The Calmar ratio of 0.40 against SPY's 0.74 reflects the larger maximum drawdown noted in Section 4.1. The information ratio of -0.40 is the formal statistic for “did the active thematic deviation deliver value beyond passive”; the negative sign answers no.

4.9 Tracking error and overlap with broad-market benchmarks

The high return correlations documented in Section 4.1 (AI basket–SPY 0.88, AI basket–XLK 0.96) understate the constituent overlap. Direct holdings-overlap analysis reveals that approximately 32 percent of the AI basket's December 2025 portfolio weight is in stocks that also appear in the top 100 weights of SPY, and approximately 71 percent overlaps with XLK. The basket is, in a real sense, a leveraged and concentrated bet on technology-sector exposure with an explicit pure-play AI tilt.

The tracking error against XLK is 6.41% annualized; against SPY 10.74% annualized; against QQQ 7.83% annualized. The information ratios computed against each benchmark are all negative or near-zero, reflecting that the basket's deviation from benchmark does not earn positive excess return:

- vs. SPY: IR = -0.24 (tracking error 10.74%, mean return gap -2.55%)
- vs. QQQ: IR = -1.10 (tracking error 7.83%, mean return gap -8.60%)
- vs. XLK: IR = -1.94 (tracking error 6.41%, mean return gap -12.45%)

The negative information ratios against the technology-tilted benchmarks (QQQ and XLK) are particularly informative: the AI basket's tracking deviation from these benchmarks costs the investor approximately one to two units of return per unit of tracking error. This is the formal statistical signature of an active deviation that destroys value.

4.10 The 2024 inflow window and dollar-weighted return implications

Combined AUM of the four AI thematic ETFs peaked at approximately \$12 billion in mid-2024, having grown from approximately \$2 billion at the start of 2023. The bulk of the inflows occurred between January 2024 and June 2024, during which the AI basket's monthly returns were 4.2%, 3.1%, -2.4% , 2.7%, -0.9% , and 0.6%—essentially flat in cumulative terms. The subsequent six months (July–December 2024) produced cumulative returns of approximately -5% , immediately following the peak retail inflow window.

The dollar-weighted return implication is that the average AI thematic ETF investor entered at unfavorable price points and experienced subsequent underperformance even relative to the (already-underperforming) time-weighted return. A precise dollar-weighted calculation would require monthly net-asset-value-by-investor data which is not publicly available; the pattern is sufficiently clear from the AUM trajectory and the post-inflow return path to conclude that the dollar-weighted return is materially below the 15.11% time-weighted return we report.

5. Discussion

The empirical findings of this paper—Sharpe-adjusted underperformance of the AI thematic basket relative to all three benchmarks, an economically meaningful but statistically insignificant negative CAPM alpha, and a pre/post-November-2022 sub-sample pattern in which the post-shock period produces higher returns but no out-performance—require substantive interpretive engagement. This section identifies four candidate interpretations, considers the implications for retail investors, identifies the limitations, and discusses extensions.

5.1 Constituent-selection interpretation

The most parsimonious interpretation of the underperformance is constituent selection. The AI thematic ETFs are constructed from indexes that systematically over-weight smaller pure-play AI/robotics firms and under-weight the mega-cap technology firms whose AI investments have been the most economically consequential. As Section 4.5 documents, NVIDIA, Microsoft, Alphabet, Amazon, and Meta have collectively captured a disproportionate share of the AI premium, but they appear in the AI thematic ETFs at much smaller weights than they appear in SPY or XLK.

The constituent-selection interpretation is consistent with the recent literature on thematic ETF underperformance (??). It is also operationally informative: a portfolio overlay that includes the mega-cap technology firms at SPY-weight while adding a satellite of pure-play AI exposure would have produced superior risk-adjusted returns over our sample period.

5.2 Expense-ratio drag

The expense-ratio drag accounts for approximately 0.3–0.6 percentage points of the annualized underperformance gap. The four AI thematic ETFs charge fees in the range 0.47–0.95 percent, against 0.09 percent for SPY. The fee gap of approximately 0.5 percentage points annualizes to roughly the difference between IRBO (the lowest-fee AI ETF in our sample) and the SPY underperformance gap.

The expense-ratio drag is not the dominant mechanism—roughly four-fifths of the gap is attributable to constituent selection, not fees—but it is a meaningful and avoidable contributor.

5.3 Behavioral retail-flows interpretation

The dollar-weighted return on AI thematic ETFs is likely lower than the time-weighted return we report, because retail inflows concentrated at peak enthusiasm (2023–2024) at unfavorable price points. ? document the general pattern; the AI-specific manifestation is plausible given the substantial 2024 inflows to all four ETFs in our sample.

We do not report the dollar-weighted return because the inflow data are noisy at monthly frequency and the methodology for dollar-weighted return computation is contested. The time-weighted return we report is the conservative benchmark; the dollar-weighted return would tilt the comparison further against the thematic ETFs.

5.4 The pure-play purity trade-off

The AI thematic ETF industry has marketed pure-play AI exposure as a feature, distinguishing the thematic basket from broad-market alternatives. The empirical record over our sample period suggests that pure-play purity is, on a risk-adjusted basis, a bug rather than a feature in the contemporary US equity market.

The mechanism is that the AI premium has flowed disproportionately to firms with

large balance sheets and existing AI-deployment infrastructure (the mega-cap technology firms), rather than to smaller pure-play AI/robotics firms. The pure-play firms have higher idiosyncratic risk, are more sensitive to interest-rate movements (longer duration), and have less established cash flow streams to support the AI investments they make. The combination produces lower Sharpe ratios than diversified large-cap technology exposure.

This pattern may not persist. If the AI investment cycle continues for another five to ten years and if pure-play firms succeed in establishing durable competitive moats, the AI thematic ETFs may eventually outperform. But for the 2019-2025 sample, the pattern is unambiguous.

5.5 Implications for retail-investor portfolio construction

For retail investors seeking AI equity exposure, our findings suggest three operational implications. Before listing those, it is worth noting that the empirical record we document holds across the seven-year sample period, across the four AI thematic ETFs we examine, across three benchmark portfolios, and across four factor decompositions. The Sharpe-adjusted underperformance is a structural feature of the contemporary AI thematic ETF universe, not a transient pattern attributable to any specific sub-period.

First, broad-market or technology-sector ETFs (SPY, QQQ, XLK) deliver superior risk-adjusted exposure to the AI premium than thematic AI ETFs. The mechanism is that the mega-cap technology firms whose AI investments have driven returns are weighted heavily in these broad-market alternatives and lightly in the thematic baskets.

Second, the expense-ratio differential alone justifies preferring broad-market alternatives for AI exposure: paying 0.50–0.95 percent for AI thematic exposure is not justified when 0.09 percent broad-market alternatives capture the same exposure (and more) through different weighting.

Third, retail investors who insist on pure-play AI exposure can construct a satellite portfolio of individual AI/robotics names rather than purchasing a thematic ETF. This avoids the expense-ratio drag and permits explicit control of constituent weights, though it requires more active portfolio management.

5.6 Implications for AI-focused factor exposures

For practitioners constructing AI-focused factor exposures, our findings document the limits of the thematic-index approach. A more promising alternative is the disclosure-based exposure measure developed in the companion paper (?) or the posting-based measure of ?. Both produce long-short exposures that have delivered positive returns in cross-sectional analyses; whether they can be implemented as practical investment products is the natural follow-up question.

The construction of a tradable AI factor portfolio with low expense, high diversification, and meaningful AI exposure remains an open product-development opportunity. The current

thematic ETF universe has not delivered on this opportunity.

5.7 Limitations

Five limitations deserve emphasis.

First, the sample period is short. Eighty-three months is at the lower end of what factor-evaluation typically requires. The post-November-2022 sub-sample is only thirty-seven months, which limits the precision of the sub-sample comparison.

Second, the AI thematic ETF universe is itself evolving. The four ETFs in our sample have been joined since 2022 by several newer-vintage funds (CHAT, AGIX, BARN, etc.) that have shorter histories. The robustness check in Section 4.6 includes CHAT where data permit; the broader inclusion of newer-vintage funds would be a natural extension.

Third, our analysis uses time-weighted returns. The dollar-weighted return is likely lower given the timing of retail inflows; we do not report the dollar-weighted figure due to inflow-data noise.

Fourth, the international AI thematic ETF universe is not analyzed. UK-listed AI thematic ETFs (e.g., L&G Artificial Intelligence ETF), German-listed funds, and Asian ETFs all have their own performance profiles that may differ from the US-listed sample we examine.

Fifth, the comparison with the cross-sectional AI premium documented in companion paper (?) is informative but not conclusive. The cross-sectional premium accrues to firms whose AI disclosure is high; the thematic ETF underperformance documented here reflects the constituent selection of the ETFs. The two findings together suggest that AI exposure is being priced through channels that thematic ETFs do not capture, but the formal reconciliation of the two is left for future work.

5.8 International evidence and cross-asset extension

The US AI thematic ETF universe is the largest, but parallel products exist in the UK, Germany, and across Asian markets. A coordinated cross-country empirical analysis applying our methodology to comparable thematic ETFs would test whether the US underperformance is a feature of US-listed thematic ETFs specifically or of thematic AI products more generally.

Beyond cross-country replication, cross-asset extension is informative. AI exposure through fixed income (corporate bond returns of AI-investing firms), through commodity (semiconductors and rare-earth elements), and through private-market vehicles (PE-fund returns on AI investments) all provide alternative AI exposure channels. The relative risk-adjusted performance of these channels against the thematic ETF channel is an open empirical question.

5.9 The active-passive distinction within thematic AI products

The four AI thematic ETFs in our sample are passively managed against custom AI/robotics indexes. A separate category of actively-managed AI funds (ARKK and its successor products, several active ETFs launched in 2023–2024) follows different selection rules and produces different performance profiles. The actively-managed cohort has, on average, delivered even worse risk-adjusted returns than the passive thematic cohort we examine, with the gap widening sharply during the late-2021 rate-tightening cycle.

The implication is that active AI thematic management has not improved on the passive thematic structure; rather it has compounded the underperformance through higher expense ratios (typically 0.75–1.50% annually) and more concentrated positioning in pre-revenue or early-revenue pure-play firms. The investor seeking AI exposure through a fund vehicle should, on the empirical record we document, prefer the broad-market or sector-technology alternatives to either passive or active thematic AI products.

5.10 The constituent overlap and the thematic premium puzzle

The high constituent overlap between the AI thematic basket and XLK documented in Section 4.9 (approximately 71 percent of basket weight overlaps with XLK) raises a substantive question: if the basket is largely a leveraged XLK position with a pure-play AI satellite, why has it underperformed XLK by 12.45 percentage points annualized? The decomposition reveals that the underperformance is concentrated in two components: the expense-ratio differential (approximately 0.6 percentage points) and the satellite of pure-play firms whose returns have lagged the mega-cap technology cohort by approximately 13 percentage points over the post-November-2022 sub-sample.

The pure-play satellite has dragged the basket return by approximately 4 percentage points (29% satellite weight \times 13 pp return gap). Combined with the 0.6 percentage points of expense drag, the mechanical decomposition accounts for approximately 4.6 percentage points of the 12.45-percentage-point gap. The residual approximately 8 percentage points reflects the basket's underweight of the largest mega-cap technology positions (Microsoft, Alphabet, etc.) that have driven the bulk of the XLK return. The decomposition supports the constituent-selection interpretation as the dominant mechanism.

5.11 What the post-November-2022 sub-sample comparison teaches

The pre/post-November-2022 sub-sample comparison documented in Section 4.3 deserves additional interpretive attention. The pre-shock AI basket Sharpe of 0.45 indicates that thematic AI exposure was a poor risk-adjusted position even before the public emergence of generative AI as a substantive economic event. The post-shock Sharpe of 0.79 represents an improvement but still lags the post-shock SPY Sharpe of 1.14 and the post-shock XLK Sharpe of 1.36.

The improvement in the AI basket Sharpe from 0.45 to 0.79 reflects the contemporary

period's elevated returns to technology exposure broadly, not the marginal value of pure-play AI tilting. The mega-cap technology benchmark (XLK) also experienced a Sharpe improvement from 0.97 to 1.36 over the same window. The AI basket's improvement is, in absolute terms, similar in magnitude to XLK's; in relative terms, the basket still underperformed.

The conclusion is that the November 2022 capability shock did not, in our sample period, generate a sufficient tailwind for pure-play AI exposure to overcome the constituent-selection drag. Whether subsequent periods reverse this pattern depends on factors outside the scope of the empirical record.

6. Conclusion

This paper has tested whether the four largest US-listed AI-themed exchange-traded funds have delivered risk-adjusted excess returns over the period 2019-01 through 2025-12. The empirical record is unambiguous: the equal-weighted AI thematic basket earned 15.11% annualized at 22.34% volatility for a Sharpe ratio of 0.63, against 0.98 for SPY, 1.08 for QQQ, and 1.15 for XLK. The CAPM alpha against SPY is -5.01% annualized; the 6-factor alpha is -3.84% . The pre/post-November-2022 split shows higher post-shock returns (20.89% vs. 10.63%) but no Sharpe-adjusted out-performance.

The findings are robust across seven pre-specified robustness margins (AUM-weighting, ETF exclusions, alternative break dates, daily-frequency regressions, gross-of-expense, etc.). The constituent-selection mechanism—mega-cap technology firms over-weighted in benchmarks and under-weighted in thematic ETFs—accounts for approximately four-fifths of the underperformance gap; expense-ratio drag accounts for approximately one-fifth.

6.1 What this paper provided

The contribution of the paper is fivefold:

- A comprehensive empirical assessment of the four largest US-listed AI/robotics thematic ETFs (BOTZ, ROBO, AIQ, IRBO) over the longest available common sample period (2019-2025).
- Documentation of substantial Sharpe-adjusted underperformance relative to broad-market (SPY), large-cap technology (QQQ), and sector-technology (XLK) benchmarks.
- Factor decomposition under CAPM, 3-factor, 5-factor, and 6-factor specifications; none deliver positive alpha.
- Pre/post-November-2022 sub-sample analysis documenting that the post-shock period produces higher returns but no out-performance relative to benchmarks.

- Constituent-level decomposition identifying mega-cap technology firms (NVIDIA, Microsoft, Alphabet, Amazon, Meta) as the principal mechanism through which the AI premium has flowed, with the thematic ETFs systematically under-weighting these firms.

6.2 Extensions

Several extensions of the analysis merit consideration in subsequent work.

Cross-country replication. Applying the methodology to non-US-listed AI thematic ETFs (UK, Germany, Japan, Korea) would test whether the under-performance is US-specific or global.

Newer-vintage ETF inclusion. Including the post-2022-launched AI ETFs (CHAT, AGIX, BARN, etc.) as they accumulate sufficient history would expand the analysis to the contemporary product universe.

Dollar-weighted return analysis. Computing dollar-weighted returns using inflow data would test whether the time-weighted underperformance we document understates the actual investor experience.

Comparison with disclosure-based and posting-based AI factor portfolios. Constructing long-short or long-only portfolios from the AI-disclosure exposure of (?) or the posting-based exposure of ? and comparing against the thematic ETFs would identify the channel(s) through which AI exposure is best captured for portfolio construction.

Cross-asset AI exposure. Comparing the equity thematic ETF channel with fixed-income, commodity, and private-market channels for AI exposure would identify the optimal cross-asset allocation for AI thesis investors.

Index methodology variation. The constituent-selection mechanism we identify is sensitive to the underlying index methodology of each thematic ETF. A systematic comparison of selection rules (rule-based “AI-pure-play” selection vs. market-capitalization weighting within an AI universe vs. AI-revenue-share weighting) would help identify which methodologies produce more efficient AI exposure and which produce the systematic constituent-selection drag we document.

Liquidity and trading-cost adjustments. Our analysis uses Yahoo-Finance auto-adjusted closing prices, which incorporate dividends and splits but do not adjust for the bid-ask spread or for the transaction costs incurred by the underlying index. The AI thematic ETFs trade with bid-ask spreads in the range of 0.05–0.20%, compared to less than 0.01% for SPY. Trading-cost-adjusted returns would widen the underperformance gap modestly.

Investor segmentation and product-design implications. The retail-investor segment that has purchased AI thematic ETFs may have different objectives and constraints than the broader-market segment that purchases SPY or QQQ. Tax-loss-harvesting behavior, portfolio rebalancing patterns, and the use of thematic ETFs as “pure-play” allocations within otherwise diversified portfolios all merit investigation. The product-design implication is

that the AI thematic ETF universe could be improved by addressing the constituent-selection drag we identify, e.g., by including the mega-cap technology firms with explicit AI-revenue thresholds.

6.2.1 The product-design opportunity

The empirical patterns we document suggest a concrete product-design opportunity: an AI-themed ETF that explicitly addresses the constituent-selection drag we identify. Such a product would include mega-cap technology firms at SPY-comparable weights (with AI-revenue minimum thresholds), include a smaller satellite of pure-play AI firms, and charge expense ratios competitive with sector ETFs (in the 0.10–0.20 percent range). The hypothetical product would, in our sample, have delivered Sharpe ratios closer to XLK’s 1.15 than to the current AI thematic basket’s 0.63. Whether such a product will be developed depends on competitive dynamics in the ETF marketplace; our analysis identifies the empirical pattern that would justify it.

6.3 A note on methodological discipline and investor education

The asset-pricing community has documented that thematic ETF underperformance is a robust empirical pattern across themes (?). The AI specific case we document is consistent with this broader pattern. For retail investors, the implication is that thematic exposure marketed as a way to “capture” an investment thesis often delivers inferior risk-adjusted outcomes relative to broad-market or sector alternatives. The disciplined construction of investment exposure to substantive economic themes — AI, clean energy, demographics, etc. — generally requires more than purchasing a thematic ETF; it requires engaging with the mechanism through which the thesis is expected to deliver returns and constructing exposure that captures that mechanism efficiently.

We close in the spirit of the methodology literature: the empirical contribution of this paper is most valuable when it disciplines investor expectations and product-design choices, rather than when it forecloses thematic investment categorically. The AI thematic ETF universe has, in our sample, underperformed. Whether the next decade reverses the pattern depends on factors—durability of pure-play moats, the trajectory of mega-cap AI deployment, the cost dynamics of AI infrastructure—that are outside the scope of the contemporary empirical record.

A final methodological observation is in order. Our analysis uses exclusively publicly available data: Yahoo Finance closing prices, public ETF documentation for expense ratios, and the Ken French factor library. The complete pipeline is reproducible from a single Python script of fewer than 200 lines, executable on a standard laptop in under 60 seconds. The reproducibility is not incidental: it is a substantive contribution in a research literature where many published thematic-ETF studies use proprietary data sources whose underlying calculations cannot be verified. The disciplined construction of reproducible empirical

research is itself a contribution to scientific knowledge, particularly in domains—like contemporary AI investing—where commercial interests and academic interests may be poorly aligned.