Active Machine-Learning-Based Trading and Mutual Fund Performance*

Xiaowen Hu
Southern Methodist University

Maximilian Rohrer
Norwegian School of Economics

Hanjiang Zhang
Washington State University

First Draft: March 2025

Current Draft: January 2026

*Xiaowen Hu is from the Cox School of Business, Southern Methodist University. Phone: (214)768-3996. Email: xhu@mail.smu.edu. Maximilian Rohrer is from the Department of Finance, Norwegian School of Economics. Phone: (+47) 55 95 93 94. Email: maximilian.rohrer@nhh.no. Hanjiang Zhang is from the Carson College of Business, Washington State University. Phone: (509)335-3797. Email: hanjiang.zhang@wsu.edu.

Active Machine-Learning-Based Trading and Mutual Fund Performance

ABSTRACT

We comprehensively examine the utilization of Machine Learning by U.S. equity mutual funds to enhance performance. We propose a novel holdings-based Active ML-Based Trading (AMLT) measure that assesses how mutual funds actively align their portfolio decisions with forward-looking ML trading signals, which are generated from a Deep Neural Network model using a near-universe of public information inputs, combining well-known quantitative predictors and rich textual signals derived from hundreds of millions of corporate disclosures and news. A detailed analysis of AI talent based on comprehensive mutual fund employee profiles provides external validation of our measure. We document a significant rising trend in AMLT adoption, with top-decile AMLT funds outperforming bottom decile funds by 2.4% to 3.0% annually on a risk-adjusted basis. The outperformance derives from superior stock selection, lower expenses, and efficient trading cost management, despite the inherently high turnover of ML-based strategies. We identify two key drivers of superior performance. AMLT based on a full-fledged ML model more than doubles the performance based on Linear quasi-ML models or reduced information sets, highlighting ML's ability to process extensive information and capture complex interactions. Managers' ability to integrate ML with human expertise contributes to sustained performance over long horizons, across different market conditions, and investment styles.

Keywords: Mutual Funds, Machine Learning, Artificial Intelligence, Active Share.

JEL Codes: C55; G11; G14; G23.

1 Introduction

The advancement of artificial intelligence (AI) technology, particularly machine learning (ML), over the past two decades, coupled with the rise of big data, has transformed the landscape of knowledge production (Abis and Veldkamp, 2024). These developments have significant implications for the asset management industry, which depends on rapid and precise information processing to gain a competitive edge and deliver superior performance. Despite anecdotal evidence of growing AI adoption in asset management, the extent of such adoption, in particular in enhancing investment strategies and performance, remains unclear.[1] A systematic understanding of the scope and effects of ML adoption in mutual fund investments is therefore of critical importance.

It is unclear, ex ante, whether professional investors such as mutual funds can effectively leverage this revolutionary technology and utilize ML-generated signals to enhance performance. On the one hand, ML models with powerful information processing capacity have been shown to produce trading signals with exceptional investment returns, significantly outperforming traditional strategies based on limited information inputs and simple linear predictors.[2] On the other hand, adopting ML technology is highly complex and requires significant human and physical capital. A greater challenge lies in the difficulty of profitably executing an ML-based trading strategy, which depends on short-lived trading signals, hard-to-arbitrage stocks, high trading costs, and constrained market conditions (Avramov et al. (2023); Jensen et al. (2024)), a task that demands considerable investor sophistication and trading expertise.

In this paper, we comprehensively investigate the adoption of ML-based trading strategies by U.S. active equity mutual funds over the past two decades. We ask the following questions. First, do mutual funds utilize ML-based trading strategies in their investment practices? Second, can ML-based trading help mutual funds achieve superior portfolio performance? Third, what are the mechanisms that drive the performance of funds adopting the ML-based trading strategies?

To identify the utilization of ML technology in mutual fund investment practices, we propose an Active Machine-Learning-Based Trading (AMLT) measure based on mutual fund holdings. AMLT is a signed active-share measure that captures how a fund actively aligns

[1]Appendix A.1 provides a historical review of both the development of AI and ML technologies and their adoption in asset management.

[2]The ability of ML models to generate profitable investment opportunities has been extensively documented, see among others Gu et al. (2020); Kelly et al. (2019); Freyberger et al. (2020); Kozak et al. (2020); Lettau and Pelger (2020); Bianchi et al. (2021); Cong et al. (2022); Choi et al. (2024); Feng et al. (2024); Chen et al. (2024); and Han et al. (2024).

its portfolio weight deviations from its passive benchmark each quarter with forward-looking ML-based trading signals. First, to obtain ML trading signals, we build on the return prediction framework of Gu et al. (2020) by implementing a Deep Neural Network (DNN) with a near-universe of public information inputs. This includes extensive quantitative signals from the original study,[3] and further incorporates textual inputs constructed from exhaustive corpora of over 2 million financial filings, 697 million news articles, and 261,179 earnings call transcripts, which constitutes one of the most comprehensive sets of public information inputs used for machine-learning-based return prediction. Second, to capture active trading decisions, we follow the procedure of Cremers and Petajisto (2009) to compute the deviations of fund portfolio weights from their benchmark weights utilizing twenty-one widely used benchmark indexes. The fund-level signed active share measure is then computed by multiplying the active weight deviations by ML trading signal indicators and aggregating across all positions.

The holdings-based AMLT measure offers several advantages. First, AMLT is a trade-alignment measure that provides a sharp identification of ML adoption in enhancing investment strategy and fund performance. AI adoptions unrelated to portfolio investments, such as customer relations and marketing, are not picked up by the AMLT measure. Second, our methodology enables a detailed analysis of the mechanisms underlying performance. The AMLT measure can be readily adapted to capture trading strategies based on different ML models and information categories, providing insights into how both ML technology and information inputs contribute to investment outcomes. Third, our holdings-based methodology can be applied to the entire U.S. mutual funds universe at the fund level, whereas identifying AI-adopting funds based on funds' self-reporting or labor market information at the fund-company level faces data limitations. For these reasons, we use the holdings-based AMLT measure in this study, and provide external validation of our measure through a detailed analysis of fund-company AI talent using individual resumes following Babina et al. (2024), as well as an examination of a set of self-designated AI funds.

Equipped with the AMLT measure, we conduct an extensive study of actively managed U.S. equity funds from 2000 to 2022. First, we document substantial variation in the adoption of ML-based trading strategies across funds and over time. The median AMLT is near zero, with a standard deviation of about 0.68. Focusing on the component of AMLT reflecting correct portfolio alignment with ML signals, funds in the top decile actively deviate from their benchmark to align with ML signals in 23.3% of their portfolios, while those in the

[3]The predictor set of Gu et al. (2020) includes 94 characteristics for each stock, interactions of each characteristic with eight aggregate time-series variables, and 74 industry sector dummy variables, totaling more than 900 baseline signals.

bottom decile misalign by 36.5%. Moreover, the adoption has risen markedly over time. For example, AMLT at the 80^{th} percentile has more than doubled over the past two decades. The upward trend is evident across almost all style groups.

Examining the determinants of the AMLT strategy yields a series of interesting results. We observe that funds with better past performance and lower fund risk exhibit higher AMLT and they charge lower expenses. In addition, despite the extremely high turnover associated with the ML strategy as documented in the literature, we find funds with high AMLT exhibit low portfolio turnovers. Consistent with that smaller assets under management allow funds to better manage trading impact in forming ML-based portfolios (Jensen et al., 2024), we find small funds are good at following the AMLT strategy. In the time series, we observe that AMLT is higher during high-VIX periods when ML profitability is stronger, as documented in the prior literature, and is also elevated during economic expansions, consistent with better stock selection skills in these periods (Kacperczyk et al., 2012).

Next, we examine performance outcomes of AMLT. We find that funds in the top AMLT decile exhibit significant risk-adjusted performance and they outperform those in the bottom AMLT decile by 2.4% to 3.0% per year across various risk adjustment models. The results are robust to using an alternative AMLT measure that captures only long-side ML trading opportunities, as well as to regression analysis that controls for additional fund characteristics. Further, performance decomposition shows that superior performance stems from better stock selection, lower expenses, and higher return gap (i.e., attributable to superior interim trading benefits and lower trading costs). Importantly, we find strong persistence in both the AMLT measure and future fund performance, suggesting AMLT is a deliberate strategy that some funds employ to achieve sustained performance.

After establishing the baseline results, we dive deeper into understanding the key mechanisms behind AMLT performance. We find that the powerful information-processing capacity of ML technology, namely its ability to process vast amounts of information and capture complex interactions, is a primary driver of superior fund performance. Specifically, we show that an AMLT strategy focusing on companies with richer information inputs achieves even stronger performance. Further, a full-fledged AMLT strategy more than doubles the performance generated by strategies that rely either on linear quasi-ML models (e.g., Elastic Net, PCR) or on models with reduced information sets (i.e., solely quantitative or solely textual inputs).

Our evidence also suggests that superior AMLT performance stems from mutual fund managers' ability to integrate human investment expertise with ML technology, enabling them to deploy ML-driven strategies effectively despite mandate constraints and implementation challenges. We find that high-AMLT funds effectively leverage ML's strengths by tilting toward stocks with richer information. Moreover, despite the challenges of ML-based trading documented in the literature (i.e., high turnover, hard-to-arbitrage stocks, and market constraints), high AMLT funds achieve persistent performance across all market conditions (VIX, sentiment, illiquidity) and investment styles, while effectively managing trading costs, as evidenced by their lower turnover and higher return gap. Our results therefore echo the insights of Cao et al. (2024).

Lastly, we provide external validations of our holdings-based measure and a series of additional analyses. To validate our AMLT measure, we conduct a comprehensive examination of employee AI talent across all mutual fund advisor companies. Using Lightcast (formerly Burning Glass) employee profile data, we construct the share of fund employees with core AI skills (i.e., AI, ML, NLP, and computer vision, following Babina et al. (2024)). We find that this labor-market-based AI talent measure is highly correlated with our trade-alignment-based AMLT, supporting our holdings-based approach. Importantly, when both measures are included, AMLT remains the only significant predictor of fund performance, underscoring the ability of the holdings-based measure to capture AI adoption targeted at enhancing investment decisions and performance, whereas the AI talent measure may reflect broader, non-investment uses of AI. In addition, we identify a small set of self-reported AI funds using fund names and online searches and find that they exhibit substantially higher AMLT than other funds within the same style group, further confirming our measure's ability to capture AI adoption.

We also perform a series of additional analyses. Our subperiod results show that AMLT is associated with superior fund performance in both the pre- and post-2010 periods, with stronger effects in more recent years, indicating that mutual funds progressively improve their ML strategies and performance as the technology advances. In addition, alternative AMLT measures that use beginning-of-period information or different signal transformations yield robust results. We also find consistent results using an alternative turnover-adjusted ML trading measure (TOMLT) that emphasizes funds' quarterly trading toward strong ML signals.

We contribute to the mutual fund performance literature by providing an extensive examination of the ML adoption in U.S. equity mutual fund investment practice over the past two decades and documenting the significant value added by this practice. We develop a novel signed active-share measure, AMLT, that quantifies the extent to which funds actively align their portfolio investment decisions with ML-generated trading signals. Our method is unique in that it is grounded in a thorough examination of funds' investment actions and avoids capturing AI adoptions unrelated to portfolio investments. Methodologically, our paper is closely related to the mutual fund literature that employs holdings-based measures to assess fund performance (Daniel et al. (1997), Kacperczyk et al. (2007), Kacperczyk et al. (2008); Huang et al. (2011)) and, in particular, the growing literature examining the activeness of portfolio management (e.g., Cremers and Petajisto (2009); Cremers et al. (2013); Petajisto (2013); Stambaugh (2014); and Jiang and Zheng (2018)). By combining active weight deviations with the powerful ML trading signals, the AMLT measure not only enhances performance evaluation but also provides a metric for understanding the underlying investment strategy driving the value added.

Second, our paper also contributes to the literature that examines information utilization and fund performance. Although some mutual fund managers can add value by following private information (Cohen et al., 2008), previous literature shows that reliance on public information, such as analyst recommendations or news coverage, is associated with poor fund performance (Kacperczyk and Seru (2007); Fang et al. (2014)). However, the substantial advancement of AI technology over the past two decades has significantly transformed the landscape of information production (Abis and Veldkamp, 2024). By focusing on ML-based strategies capable of efficiently processing vast amounts of public information and extracting valuable trading signals, our study provides new insights into how advanced AI technology can help funds gain informational edge and drive superior performance. With the flexibility of our methodology, we further illustrate that the volume of information inputs (i.e., total number), the diversity of information (e.g., qualitative and textual data), and the interplay among various information inputs (i.e., complex interactions) all contribute to the ML-driven information edge and the superior outperformance of the AMLT strategy.

Third, our paper contributes to a fast growing literature that examines the utilization of ML models in various areas in finance, e.g., predicting asset prices (Gu et al. (2020); Lettau and Pelger (2020); Bianchi et al. (2021); Chen et al. (2024); and Han et al. (2024)), robo-advising (D'Acunto et al., 2019), earnings forecast (Van Binsbergen et al., 2023), analyst's recommendation (Cao et al., 2024), fund performance evaluations (Li and Rossi (2020); Kaniel et al. (2023); and DeMiguel et al. (2023)), lending decisions (Liu (2022); Fuster et al. (2022)), innovation evaluation (Zheng, 2024), and estimating bank risk (Hanley and Hoberg, 2019). Our paper contributes to this literature by focusing on the application of ML trading signals in real investment practices of mutual funds. While ML models have shown substantial theoretical investment profits, existing literature highlights the significant challenges of implementing ML-based trading strategies in practice due to the concentrations in hard-to-arbitrage stocks, fleeting signals and high trading costs (Avramov et al. (2023); Jensen et al. (2024)). Our study advances this literature by demonstrating that mutual fund managers possess the skills to successfully integrate ML signals into their portfolio investments, generating superior and persistent fund performance. Our evidence aligns with the insights of Cao et al. (2024).

Our paper adds to the understanding of how the rise of new technologies, including both AI and new data, is reshaping the asset management industry. Bonelli and Foucault (2023) explore how the emergence of alternative data affects mutual fund investment practices. Several recent papers study AI adoption by measuring the hiring of AI talent and data scientists at the asset management company level (Zhang (2024); Cen et al. (2024); Chen et al. (2025)). Our paper differs from these studies by explicitly examining how mutual funds employ machine-learning technology to enhance investment strategies and the impact on fund performance. To this end, we propose a holdings-based methodology that measures ML adoption in investments by assessing funds' trade alignment with ML-based signals. In a recent paper, Sheng et al. (2024) shows that the utilization of ChatGPT to process conference call information helps hedge funds generate performance. ChatGPT is an emerging technology since 2022 that makes the AI tool adoption possible for a wide range of investors. Complementing their work, we examine how professional investors directly utilize the underlying complex ML technology in investment practice over the past two decades, a setting in which money managers have both the capability and the incentive to build an information advantage as cutting-edge tools become available. We highlight the rising trend of mutual funds adopting ML-powered trading strategy, particularly over the past decade, and show that such a strategy enables funds to add value through ML's remarkable information processing capability with exhaustive data inputs.

The rest of the paper is organized as follows. In Section 2, we discuss the data and sample construction. Section 3 begins by implementing ML models to generate trading signals. We then construct the AMLT measure, provide external validations using labor-market information, and discuss the determinants and persistence of AMLT. The baseline results on AMLT and fund performance are presented in Section 4. Section 5 examines the underlying drivers of AMLT performance. Section 6 provides additional analyses and robustness tests, and Section 7 concludes.

2 Data and Sample Construction

2.1 Mutual Fund Data

For our research purposes, we focus on U.S. actively managed equity mutual funds based on the CRSP Survivorship Bias Free Mutual Fund database. The sample begins by aligning mutual fund holdings at the end of December 2001 with the first out-of-sample forecast generated by our ML model, namely the predicted return for the first quarter in 2002. This timing reflects data availability: corporate filings, one of our primary textual inputs, become available in 1997, and the model requires a four-year validation window prior to producing out-of-sample predictions.[4] Our sample spans two decades through 2022.

The CRSP database includes mutual fund characteristics such as fund returns, total net assets (TNA), fees, flows, and investment objectives. We exclude balanced, bond, international, money market and index funds, as well as funds that on average hold less than 80% of their assets in common stock if the investment objective is missing. To avoid the incubation bias identified by Evans (2010), we exclude funds that managed less than $10 million in the previous month, funds with missing fund names in the CRSP database, and funds for which the observation year is the same as or earlier than the reported fund start year. Mutual fund share classes are aggregated to the fund level. Fund returns, expense ratios, and turnover ratios are asset-weighted averages across share classes, total net assets (TNA) are summed, and fund age is defined as the age of the oldest share class.

We merge the CRSP mutual fund database with the Thomson-Reuters Mutual Fund Holdings database and the CRSP Stock Database using the MFLINKS file based on Wermers (2000) and available through Wharton Research Data Services (WRDS). The Thomson-Reuters data provides mutual funds' equity holdings on specific disclosure dates, which allow us to study mutual fund investment and trading decisions.

Table 1 presents summary statistics for the equity mutual funds in our sample, which consists of 5,024 distinct mutual funds. Funds in our sample have an average total net assets (TNA) of $2,341 million and an average age of 15 years, with a turnover ratio of approximately 74.5%. The average monthly net fund return is 0.7%, with a standard deviation of 5.2%. Risk-adjusted performance, measured by the Carhart four-factor alpha, exhibits a negative mean over the sample period. The average expense ratio is 0.1% per month.

2.2 Firm-Level Predictors

Next, we discuss the data that underlies our ML-based trading signals. Our firm universe includes all publicly traded companies listed on the NYSE, AMEX, or NASDAQ.[5] Monthly stock returns are from CRSP, incorporating delisting returns following the methodology of Shumway and Warther (1999). Our firm-level sample begins in 1987 to provide a sufficiently long training and validation window for the machine-learning models, covering 18,803 unique

[4] See Appendix A.1 for the evolution of AI and ML technologies and their adoption in asset management starting in the early 2000s.

[5] We follow Gu et al. (2020) and include all firms with common stock outstanding, a reported month-end market value in CRSP, and a non-missing book-to-market ratio in Compustat, without imposing additional filters.

stocks, with an average of 4,708 stocks per month.

To construct our predictive variables, we incorporate a comprehensive set of information inputs, including both extensive quantitative and qualitative (textual) predictors, based on previous literature in both asset pricing and textual and sentiment analysis.[6] For the quantitative information set, we follow Gu et al. (2020), incorporating 94 stock-level characteristics as in Green et al. (2017) and 74 industry dummies.

In addition to extensive quantitative information, we substantially expand the information set by incorporating three comprehensive textual corpora, SEC filings, earnings call transcripts, and firm-specific news, which are widely used in the textual and sentiment analysis literature and have been shown to help explain stock returns.[7] Specifically, we obtain financial filings from SEC Analytics, comprising 238,758 annual reports (10-Ks), 666,633 quarterly reports (10-Qs), and 1.9 million current reports (8-Ks). We construct 27 predictors that capture file size and word counts, as well as sentiment variables derived from dictionaries proposed by Loughran and McDonald (2011). Second, we gather 261,179 quarterly earnings call transcripts from Wall Street Horizons, and construct a total of 19 predictors (Garcia et al., 2023). Third, we add firm-specific news articles from RavenPack News Analytics, which covers more than 40,000 news media sources globally and 697 million unique news articles. From this corpus, we construct ten predictors that capture news coverage, sentiment, and concentration.

Beyond firm-specific characteristics, we include eight macroeconomic predictors and three stock-level institutional ownership variables (Welch and Goyal (2008), Cao et al. (2024)): the number of 13F institutional investors, the institutional ownership share, and ownership concentration (i.e., the Herfindahl-Hirschman Index).

We align all characteristics in calendar time, imposing data availability constraints to ensure realistic implementation. We assume annual accounting data is available in month t if the firm's 10-K was published in month t or if the fiscal year-end occurred at least six months prior. Annual data remains available for up to 12 months after its initial release. Similarly, we assume quarterly accounting data is available in month t if the firm's 10-Q was published in month t, if the report date (RDQ) in Compustat corresponds to month t, or if four months have passed since the fiscal quarter-end. Quarterly data remains available for up to four months after publication. Monthly data, such as past stock returns and IBES

[6]Our Internet Appendix lists all individual base predictors, their definition, sources, and starting date. Because the DNN architecture implicitly generates and leverages thousands of nonlinear combinations and high-order interactions, we do not explicitly include interactions between the predictors.

[7]The literature shows that 10-K filings (Loughran and McDonald, 2011), earnings call transcripts (Garcia et al., 2023), and firm-level news (Ke et al., 2019; Von Beschwitz et al., 2020; Jiang et al., 2021; Jeon et al., 2022; Lopez-Lira and Tang, 2023) is informative about stock returns.

forecasts, are assumed to be available instantaneously.

2.3 Employee Profiles and AI Skills

Finally, to enhance our study with additional analyses and validations, we use labor market information from Lightcast (formerly Burning Glass), who provides both extensive employee profile and job posting data.[8] Lightcast profile data covers 140 million employees spanning the period from 2000 to 2024. It provides detailed information on employer names, employee job titles, skills, occupation categories, educational background, and employment history, enabling us to perform a comprehensive investigation of the AI talent of mutual fund employees.

3 Active Machine-Learning-Based Trading (AMLT)

To investigate the adoption of ML technology by mutual funds specifically to enhance investment strategy and fund performance, we propose a holdings-based methodology and construct an Active Machine-Learning-Based Trading (AMLT) measure that assesses how a fund aligns its investments with the ML-based trading signals. In this section, we begin by implementing ML models to generate trading signals. We then define the AMLT measure, examine its determinants, and provide a discussion and validation of our methodology.

3.1 Machine-Learning Trading Signals

One of the key components of the proposed AMLT measure is the ML-based trading signal, which is entirely out-of-sample so that a hypothetical mutual fund could trade on it in real time.

First, to obtain ML trading signals, we build on the return prediction framework of Gu et al. (2020) by implementing a three-layer deep neural network architecture (DNN3) and an ensemble approach in training the model, following the methodologies of Hansen and Salamon (1990) and Dietterich (2000). The detailed procedure is provided in Appendix A.2. Importantly, in our setting, the ML architecture is applied to a near-universe of publicly available information by incorporating the full set of quantitative predictors used in existing ML-based return-prediction frameworks, and substantially expanding this information set with large-scale qualitative inputs derived from extensive textual data as described in

[8]For over two decades, Lightcast has systematically collected job postings from more than 160,000 online sources and aggregated publicly available individual resumes and professional profiles from linked-in and other online platforms.

Section 2.2. Finally, each quarter, for each stock in our universe, the expected return for the subsequent quarter for the stock is generated based on the above procedure and all information inputs available at the quarter-end.

Next, we evaluate the ML-based trading performance by ranking stocks into decile portfolios based on the ML-generated trading signals at the end of each quarter and computing portfolio returns over the subsequent three months. The portfolios are rebalanced quarterly, and the results are presented in Appendix Table A.1. We find that the long-short ML-based portfolio delivers remarkable performance, with an average raw return of 2.1% per month (25.3% annually) and a Carhart alpha of 2.3% per month (27.8% annually). Our results are highly consistent with Gu et al. (2020), with stronger magnitudes during overlapping sample periods, reflecting our inclusion of extensive textual signals in addition to quantitative signals.[9]

3.2 Construction of the AMLT Measure

Equipped with the strong ML-based trading signals, we next define the AMLT measure. AMLT is constructed as a signed active-share measure in the spirit of Cremers and Petajisto (2009) and Jiang and Zheng (2018), capturing the alignment of a fund's active portfolio weight deviations from its benchmark with the ML-based trading signals. Specifically, $AMLT_{f,t}$ for mutual f at quarter t is defined as:

$$AMLT_{f,t} = \sum_{i=1}^{n} \left(w_{i,t}^{f} - w_{i,t}^{B} \right) \times ML\ Signal_{i,t+1|t} \tag{1}$$

where $w_{i,t}^{f}$ denotes the portfolio weight of stock i in fund f at the end of quarter t, and $w_{i,t}^{B}$ denotes portfolio weight of stock i in fund f's benchmark portfolio at the end of quarter t. $ML\ Signal_{i,t+1|t}$ is a forward-looking ML-based trading signal for stock i in the subsequent quarter $t+1$ given information available at the end of quarter t.

Specifically, at each quarter t, we assume that funds utilize information available up to that quarter t to generate ML-based stock return predictions for the upcoming quarter $t+1$. Based on these predictions, funds make active adjustments of their portfolio weights relative to their benchmark portfolio during quarter t. Therefore, AMLT at quarter t measures how a mutual fund actively aligns its investments in quarter t with forward-looking ML-based trading signal for quarter $t+1$ based on information available at quarter t.[10]

[9] See Gu et al. (2020) Table 7, performance of the ML portfolios for DNN3.

[10] We use information available at quarter t to generate trading signals, because the strength of the ML model is its power and speed to utilize the most up-to-date information. As a robustness check, we also construct an alternative AMLT measure utilizing beginning-quarter information (AMLT_begQInfo). The

AMLT measure is constructed as the product of two components: the active portfolio weight deviations and ML-based trading signals. First, we focus on active portfolio deviations from benchmark to capture a fund's intentional decision to align with ML-based signals, as opposed to any mechanical alignment resulting from passive benchmarking. To measure a fund's active portfolio weight deviations from its benchmark, we follow the literature and select the benchmark index from a set of twenty-one widely used indices from the S&P/Barra and Russell families.[11] Following Cremers and Petajisto (2009), for each fund-quarter, we compute active share with respect to all twenty-one indices and select as the benchmark the index that yields the lowest active share. This approach ensures that the selected benchmark best represents the fund's investment strategy and asset allocation.

Second, to construct ML-based trading signals, we apply a rank-based transformation to raw ML predictions to reduce noise. Each quarter, stocks are sorted into deciles based on predicted returns for the subsequent quarter and assigned scores from -5 to +5, with bottom-decile stocks receiving -5 and top-decile stocks receiving +5. This transformation captures the relative ranking of stocks rather than the potentially noisy return predictions, improving the interpretability and stability of the AMLT measure. Using these signals, we construct our baseline AMLT measure. Given the regulatory constraints that lead mutual funds to follow predominantly long-only strategies, we additionally construct an alternative measure, AMLT_long, that captures only the long-side ML trading opportunities.[12]

AMLT is a signed active-share measure weighted by ML trading signals and captures the extent to which a mutual fund aligns its portfolio with ML-based trading signals; higher AMLT values therefore indicate greater engagement in ML-driven trading strategies. Table 1 reports summary statistics for the AMLT measures. While the median AMLT is close to zero, there is substantial cross-sectional variation, with the standard deviation of AMLT (AMLT_long) equal to 0.68 (0.45).

Figure 1 illustrates the time trends of the AMLT measure across the mutual fund universe over the past two decades using the median and 80$^{\text{th}}$ percentile of AMLT. The plot reveals a clear upward trajectory in the adoption of AMLT by the mutual fund industry. For example, the AMLT measure at the 80$^{\text{th}}$ percentile begins at about 0.20 in the early 2000s and gradually accelerates, reaching nearly 0.45 by 2022. Moreover, as revealed in Figure 2,

results are presented in Appendix Table A.2 and are highly consistent.

[11] Specifically, we use nine indices from the S&P/Barra family, including the S&P 500, S&P 500/Barra Growth, S&P 500/Barra Value, S&P MidCap 400, S&P MidCap 400 Growth, S&P MidCap 400 Value, S&P SmallCap 600, S&P SmallCap 600 Growth, and S&P SmallCap 600 Value, and twelve indices from the Russell family, including the Russell 1000, Russell 2000, Russell 3000, and Russell Midcap, along with their respective value and growth subcomponents.

[12] Specifically, we assign a score ranging from +1 to +5 to stocks in the top 5 deciles based on the ML predictions and 0 for those in the bottom 5 deciles.

the upward trend prevails in almost all style groups. Value funds emerge as early adopters, while growth funds tend to pick up the speed in ML adoption over time. These patterns highlight the growing prevalence of AMLT implementation across the mutual fund industry.[13]

3.3 AMLT, Fund Characteristics, Market Conditions, and Time Trend

Next, we examine how AMLT is related to various mutual fund characteristics and market conditions, as well as the persistence of AMLT.

Fund Characteristics and Market Conditions

In Panel A of Table 2, we sort mutual funds in our sample into decile portfolios according to AMLT each quarter and compute various fund characteristics of each portfolio. First, we observe that AMLT varies significantly in the cross-section. Funds in the top decile exhibit an average AMLT of 0.87, while funds in the bottom decile show an average AMLT of -1.35. To better understand the share of a fund's active portfolio deviations that align with the ML signals, we present in the next column a simple AMLT_tilt measure that assigns a value of +1 or -1 to each active deviation depending on whether it aligns with the ML signal [14]. We observe that funds in the top AMLT decile actively deviate from their benchmark by 23.3% of their portfolio weights to align with ML trading signals, while funds in bottom AMLT decile make the opposite move, misaligning with ML signals by 36.5%.

Examining fund characteristics by AMLT portfolios next, a few interesting results emerge. First, we observe that funds with higher AMLT exhibit better past performance, lower fund risk (i.e., total risk, market beta and idiosyncratic fund risk) and charge lower expenses, providing some early indication that AMLT adoption seems to be associated with strong investment abilities.

Surprisingly, we find that high-AMLT funds exhibit low portfolio turnover, despite the high turnover typically associated with ML-based trading strategies. For example, Gu et al. (2020) document a monthly turnover spread of up to 113% between top and bottom ML portfolios. In contrast, funds in the top AMLT portfolio have an average annual turnover of only 57.5%, which is 13.2% lower than that of bottom-AMLT funds. We also find that smaller funds and funds from smaller families are better able to follow AMLT strategies,

[13] AMLT is a holdings-based measure. To reduce noise, we require funds to hold at least 25 positions and exclude funds with extreme turnover exceeding 200%.

[14] Specifically, we assign +1 (-1) to trades toward stocks ranked above (below) the cross-sectional median of ML predictions.

consistent with Jensen et al. (2024), who argue that managers with smaller assets under management face lower trading impact and are in a better position in forming ML-based portfolios. Taken together, these results suggest that high-AMLT funds can align portfolios with ML signals while effectively managing trading costs.

Finally, turning to activeness, we find that funds in the top AMLT portfolio actually exhibit lower active share than funds in the bottom AMLT portfolio, although both top and bottom AMLT funds are more active in portfolio management than other funds.

We confirm the above observations using the regression analysis shown in Panel B of Table 2. Moreover, adding in fund style scores (column 2), we observe that value funds and small-cap funds are associated with higher AMLT adoption, while AMLT strategy does not seem to be associated with the momentum style. Turning to various market conditions (column 3), we find, consistent with the evidence that ML signals generate stronger profits during periods with elevated market volatility, AMLT is higher in high VIX periods. AMLT is also higher in non-recession periods, suggesting that funds may be good at active stock selections during market expansions (Kacperczyk et al., 2012).

Persistency of AMLT

We next investigate the persistence of the AMLT strategy. If mutual funds exhibit high AMLT due to picking up random trading signals that are correlated with ML signals, we should not expect persistence of AMLT over time.

To investigate this, we compute the transition matrix in Table A.3. Each quarter, we sort mutual funds into quintiles based on the AMLT measure. Columns (2)-(6) in Panels A-C report transition probabilities across AMLT quintiles over the next one quarter to one year after portfolio formation. Column (7) reports the attrition rate, while the last column shows the difference between the probability of remaining in the same quintile and the average probability of transitioning to other quintiles. We find strong persistence in AMLT. For example, Panel A shows that 62.0% of funds in the top quintile remain there after one quarter, compared with substantially lower probabilities of moving to other quintiles. This persistence remains high at 52.8% after six months and 45.8% after one year, as shown in Panels B and C. Moreover, top-AMLT funds exhibit lower attrition, with exit rates being 0.3% and 1.6% lower over the next quarter and year, respectively, than bottom-AMLT funds.

Next, for a visual inspection, Figure 3 depicts the persistence of the AMLT measure in the next eight quarters after portfolio formation of the AMLT portfolios. We observe significant persistence in AMLT. The relative ranks among the decile portfolios remain unchanged over the next eight quarters. The evidence, together, suggests that AMLT is a consistent investment strategy that some mutual funds follow.

3.4 Discussion and Validation of the Holdings-Based Methodology

We use the holdings-based AMLT measure because it directly assesses funds' investment alignment with ML-based signals and provides sharp identification of ML utilization in trading practices, as opposed to other activities such as customer relations, marketing, retail financial planning, or general IT management. The methodology also allows us to examine the mechanisms underlying ML-related performance (analyzed in Section 5), as the AMLT metric can be readily adapted to measure trading strategies based on different ML models (e.g., linear versus nonlinear transformations of inputs) and different categories of information (e.g., quantitative versus textual inputs). Moreover, the methodology is applicable to the entire U.S. mutual fund universe at the fund level, enabling a comprehensive analysis of ML adoption, whereas other identification approaches (e.g., based on fund names, prospectuses, or labor-market information at the fund-advisor level) are more susceptible to data and coverage limitations.

We next provide external validation of the AMLT measure by examining self-designated AI and quantitative funds, as well as AI talent at fund advisory companies.

Self-designated AI Funds, and Quant Funds

We identify self-designated AI funds using CRSP fund names and online searches. Despite extensive searches of articles and fund websites, this approach identifies only 46 AI funds, highlighting its limitations. We also identify 46 quant funds based on fund names.

We examine the relation between the AMLT measures and self-designated AI and quant funds in Table 3 by regressing AMLT on AI fund and quant fund indicators, respectively. As shown in column (1), self-designated AI funds exhibit significantly higher AMLT than other funds within the same style-quarter, with an increase of 0.68, about one standard deviation of AMLT. Quant funds, as shown in column (4), also exhibit higher AMLT than the average fund in the same style, but the magnitude is much smaller compared to AI funds (i.e., about one-fifth that of AI funds).

To further understand AI versus quant funds, and their differential emphasis on strategies based on quantitative versus textual signals, we construct AMLT_quant and AML_textual measures by re-estimating the ML model using only quantitative or only textual inputs, respectively, and computing AMLT based on the resulting trading signals. Traditional quant funds rely on conventional statistical models and quantitative inputs (such as stock prices, earnings and market trends), whereas AI funds can be viewed as a more advanced subset that adopts cutting-edge AI technologies and incorporates a broader information set, including textual data. Consistent with this, we observe that, for AI funds, both AMLT_quant and

AMLT_text are significantly higher, with AMLT_quant being slightly stronger. For quant funds, however, only AMLT_quant is significant. Overall, despite the limited samples, these findings lend support to our methodology in capturing AI-adoption funds.

AI Talent of Fund Advisors

To provide further validation of our methodology, we next conduct a systematic examination of employee AI talent of mutual funds using employee profile data from Lightcast.[15] Implementing complex ML models in forming trading strategies requires special talent in AI technology. If AMLT captures AI adoption in mutual fund investments, we should expect funds with high AMLT to have high employee AI talent.

To investigate this, we start by matching mutual fund advisory companies reported in CRSP to companies covered in Lightcast profile data. Since most employees report employer names at fund advisory company level instead of fund level, analysis based on the AI talent measure using labor market information can only be conducted at fund advisory company level. Through name mapping, we are able to identify 1,183 mutual fund advisory companies, accounting for 62.7% of all advisory companies in our sample.

To identify employees of these mutual fund advisory companies, we classify employees' historical employer associations using Lightcast employee profile data, yielding over 2.15 million unique employee profiles linked to these firms. Next, we define an employee as having core AI skills if they have at least one of the following skills: Machine Learning, Artificial Intelligence, Natural Language Processing, and Computer Vision. We then construct an AI talent measure at the fund advisory company and quarter level, the Share_CoreAI, defined as the share of employees with core AI skills out of all employees with profiles associated with the company in quarter t.[16]

On average, based on existing employee profiles, mutual fund advisory companies have about 0.9% of employees with core AI skills, with a standard deviation of 2.6%. The time-series trend in AI adoption based on employee core AI skills in Figure 4 closely mirrors the holdings-based AMLT evidence in Figure 1, showing a clear upward trajectory in AI adoption in the mutual fund industry over the past two decades. The share of AI-skilled employees

[15] Lightcast provides both employee profile and job posting data. For our purposes, we focus on employee profile data. While job postings are useful for projecting companies' future demand for AI talent, employee profiles are better suited for measuring companies' current stock of AI talent. The latter is more relevant to our analysis, as we focus on mutual funds' existing capacity in adopting ML-based trading.

[16] In addition to measuring talent based on core AI skills, we also consider related AI skills and compute alternative AI talent measures following Babina et al. (2024). Specifically, we compute Share_related AI, which capture broader employee skills that are highly related to the core AI skills, using 10%, 15% and 20% cutoff thresholds. We repeat our analyses using these additional measures and find consistent results.

begins at around 1.0% in the early 2000s, accelerates after 2015, and reaches approximately 1.9% by 2022.[17]

To examine the relation between AMLT and Share_CoreAI, in Table 4, we regress AMLT on Share_CoreAI within the same quarter and style, controlling for other fund characteristics. We find a significant positive relation between the two measures. Column (1) indicates that a 10% increase in Share_CoreAI is associated with an increase in AMLT by 0.14, accounting for one fifth of a standard deviation of the measure.[18] Thus, the evidence shows that our trade-alignment-based AMLT measure is closely related to AI talent measure based on labor market information at fund companies, providing additional validation for AMLT as a proxy for AI adoption in mutual funds. It is important to note that, a key strength of AMLT lies in its ability to capture AI adoption specifically aimed at enhancing trading strategies and portfolio performance, whereas the labor-market-based AI talent measure may reflect AI adoption unrelated to investment activities. Further analysis in Section 6 confirms our conjecture.[19]

Given the strengths of the holdings-based methodology and the external validation of AMLT presented above, we use AMLT to examine AI-driven investment and performance in mutual funds in this study.

4 AMLT and Mutual Fund Performance

In this section, we examine the relation between AMLT strategy and subsequent mutual fund performance. We start by investigating the basic relation using both a portfolio approach and multivariate regressions. We then dive deeper into understanding the source and sustainability of the fund performance.

4.1 Portfolio-Based Analysis

We first employ a portfolio-based approach to examine the relation between AMLT and subsequent fund performance. Each quarter, we sort mutual funds into decile portfolios

[17] This overall trend in AI adoption that we document for the mutual fund industry is highly consistent with the recent evidence reported by Babina et al. (2024) on general companies and by Chen et al. (2025) on investment advisors over the recent decade. To facilitate comparisons, in the figure, we compute share of AI-skilled employees at advisor-employee position level. Unlike studies based on job postings, we focus on the existing stock of AI talent using employee CVs, providing unique and complementary evidence to early studies.

[18] As discussed earlier, many AI-related hirings are used in non-investment related tasks.

[19] The evaluation in Section 6 shows that the trade-alignment-based AMLT measure dominates the labor-market-based Share_CoreAI measure as the only significant predictor of future fund performance.

according to their AMLT measure of the quarter. Then, we compute equally weighted returns for each decile portfolio over the following three months and the portfolios are rebalanced quarterly. We use gross returns to capture investment skills. We estimate the risk-adjusted returns on the portfolios as intercept alphas from time-series regressions using the Carhart four-factor model.

The results are reported in Table 5. The first four rows are based on the AMLT measure that captures both the long and short sides of the ML trading signals, while the last four rows are based on the AMLT_long measure that emphasizes only the long side of the ML trading signals. We observe that AMLT strategy is positively associated with subsequent fund performance. In the quarter following the portfolio formation, funds in the top AMLT decile exhibit positive and significant abnormal performance ranging from 8.9 to 11.8 basis points based on various factor models, equivalent to 1.1% to 1.4% annually. Moreover, these top AMLT funds significantly outperform funds in the bottom AMLT decile by 19.6 to 23.8 basis, equivalent to 2.4% to 2.9% annually. Therefore, the evidence shows that mutual funds are able to gain superior performance through actively aligning with ML-based trading strategies.

Results based on AMLT_long in the next four rows that focus on long-only trading signals provide highly consistent and slightly stronger results. Funds in the top AMLT decile outperform funds in the bottom AMLT decile by 2.5% to 3.0% annually after risk adjustment. The evidence shows that, unlike individual anomalies whose profits primarily arise from short positions, ML-based signals generate substantial returns on the long side, which mutual funds, given their constraints on short selling, are better positioned to exploit.[20]

4.2 Multivariate Regression

We next use a multivariate regression approach to examine the relation between AMLT strategy and future mutual fund performance. This approach allows us to control for additional fund characteristics and fund styles that may affect performance. Specifically, we perform the following regression at a monthly frequency.

$$Perf_{f,t} = \beta_0 + \beta_1 AMLT_{f,t-1} + \boldsymbol{\beta}' \boldsymbol{X}_{f,t-1} + \gamma_t + \eta_s + \varepsilon_{f,t} \tag{2}$$

where the dependent variable $Perf_{f,t}$ is fund alpha, expressed in percent per month, computed as the difference between the fund's realized return in month t and its expected return

[20]To ensure robustness, we examine alternative transformations of the ML signals. First, we implement a more granular rank-based approach using 100 portfolios. Second, we consider a transformation that emphasizes the most informative signals (i.e., top- and bottom-three deciles). Appendix Table A.4 reports highly consistent results using these different transformations.

implied by the Carhart four-factor model. Factor loadings are estimated using fund returns over the prior 36 months (months $t-36$ to $t-1$), and expected returns are constructed using factor realizations in month t. The main independent variable is $AMLT_{f,t-1}$, computed at the most recent quarter end as of month $t-1$ and updated quarterly for each fund. We include a vector of various controls ($X_{f,t-1}$) in the regressions, including lagged fund performance ($Ret_{f,t-1}$) measured as the average fund return over the previous 12 months, expense ratio ($Exp\ ratio_{f,t-1}$), fund size ($log(TNA)_{f,t-1}$) measured as logarithm of TNA, turnover ratio ($Turnover_{f,t-1}$) and fund age ($Fund\ age_{f,t-1}$). All controls are lagged for a year. The regressions include time fixed effects (γ_t) and style fixed effects (η_s). Regressions are estimated at the monthly frequency, and standard errors are clustered at the fund level.

Table 6 reports the results. We report Carhart-alphas based on both net and gross fund returns. The first (last) two columns report results on AMLT (AMLT_long). Consistent with the portfolio-sorting evidence, we find a strong positive relation between AMLT strategy and subsequent fund performance. Using Carhart alpha reported in Column (2) as an example, the coefficient on AMLT is 0.05 and highly significant at 1% level. A decile portfolio spread in AMLT of 2.23 indicates that funds in the top decile outperform bottom funds by about 1.4% per year after controlling for fund characteristics and styles. In addition, we observe that low expense ratio, small fund size and low turnover are associated with high future fund performance, which is largely consistent with findings in the existing literature (Chen et al. (2004); Gil-Bazo and Ruiz-Verdu (2009); and Jiang and Zheng (2018)). Moreover, we find highly consistent results using net fund returns after expenses and using the AMLT_long strategy, which focuses on long-only ML signals, as shown in the last two columns. In addition, directly controlling for the active share measure has a negligible impact on our main effects, suggesting that our results are driven primarily by ML-powered active strategy.

Taken together, our evidence shows that mutual funds with effective adoption of AMLT in their investment practice are able to deliver superior portfolio performance.

4.3 Performance Decomposition

We next dive deeper to understand the source of the performance by performing performance decomposition. We follow Sialm and Zhang (2020) to decompose fund returns into stock selection, style timing, style selection, expenses, and return gap (i.e., interim trading benefits and trading costs).

$$RF_{i,t} = CS_{i,t} + CT_{i,t} + AS_{i,t} - EXP_{i,t} + RG_{i,t} \tag{3}$$

The decomposition procedure starts by decomposing the net fund return (RF) into the gross holdings return, fund expenses, and the return gap following Kacperczyk et al. (2008). The gross holdings return (RH) is defined as the value-weighted return of the previously disclosed fund holdings, where the weights correspond to the relative value of each fund position at the end of the previous month. The return gap (RG) captures the impact of unobserved actions on fund returns, including the trading costs and the interim trading benefits. Funds with lower trading costs and higher interim trading benefits will exhibit better return gaps. Second, following Daniel et al. (1997), we further divide the RH into a characteristic selectivity measure (CS), a characteristic timing measure (CT), and an average style measure (AS). CS is a measure of stock selection ability and uses as a benchmark the return of a portfolio of stocks that is matched to each of the fund's stock holdings each quarter along the dimensions of size, dividend yield, and momentum. CT is a measure of style timing ability, which examines whether fund managers can generate additional performance by exploiting time-varying expected returns of the benchmark portfolios. Finally, AS captures the returns due to a fund's tendency to hold stocks with certain style characteristics.

Table 7 reports results from our baseline regression (Equation (2)) using the various performance components as dependent variables. Panel A (Panel B) reports results based on univariate and multivariate regressions, respectively. We find that performance of AMLT strategy derives mainly from the ability to select securities and the average style of funds, while style timing ability is insignificant. Next, consistent with observations in Table 2, AMLT strategy is associated with low expense ratio, which may be consistent with the finding in previous literature that high performance funds that target sophisticated investors tend to charge low fees (Gil-Bazo and Ruiz-Verdu, 2009). Lastly, we observe a positive relation between AMLT and return gap. Despite being insignificant after including full controls, this finding illustrates that funds following AMLT strategy possess trading skills to efficiently manage trading costs and avoid eroding their performance while implementing ML strategy with high turnovers.

Overall, the superior performance of funds employing the AMLT strategy can be explained primarily by their superior stock selection ability, favorable style exposures, and the skill to efficiently manage trading costs.

5 Drivers of AMLT Performance

After establishing the positive relation between AMLT and fund performance, we next examine the underlying drivers of this performance. We investigate this along two dimensions: first, the role of ML methods and information inputs in helping mutual funds gain an information edge and achieve superior performance; and second, the role of managerial investment expertise in implementing AMLT strategies in practice, as reflected by sustaining performance despite implementation challenges.

5.1 AMLT and Information Advantages

5.1.1 The Volume of Information

First, we examine the role of information volume in driving AMLT performance. A key reason that ML technology revolutionizes information production is its strength in handling vast amounts of data at remarkable speed. Therefore, funds that adopt AMLT should tilt their portfolios toward stocks with more public information, where the use of ML technology can yield greater information advantages.

To test this, we measure funds' active portfolio tilt toward information-rich stocks. That is, we count the number of information inputs that each stock has out of the 235 individual predictors used in our AMLT model (N. Signals) and construct a fund-level active tilt measure (ATILT_N.Signals) toward N.Signals. Specifically, first, similarly to Daniel et al. (1997), we categorize the universe of U.S. common stocks into deciles based on N. Signals and assign each stock a score from -5 to 5 according to the rankings. We then compute the fund-level ATILT measure by multiplying the active portfolio weight deviations (from the benchmark) by the scores and aggregating the products over all positions. High ATILT measures capture mutual funds' active decision to invest in information-rich stocks relative to their style benchmarks. Similarly, we also construct three additional ATILT measures toward information-rich stocks based on quantitative signals (ATILT_N.Signals.quant), regulatory fillings (ATILT_N.Filings) and news articles (ATILT_N.Articles), respectively.

Table 8 Panel A presents regression results linking funds' active information-rich tilts (ATILTs) to AMLT. We find that funds following a high AMLT strategy allocate more capital to firms with a greater number of available information inputs compared to their style benchmarks. This is true whether we consider total information or its subcategories such as quantitative or textual information. Among the subcategories, AMLT funds tilt more strongly toward stocks with high levels of quantitative signals and news, rather than those with information primarily from regulatory filings. Using total information as an example, funds with a one standard deviation higher AMLT tilt toward stocks with high information inputs by 20.4%, suggesting that funds adopting AMLT make active portfolio decisions to leverage the strength of ML technology to gain an informational edge.

Next, to confirm that funds adopting AMLT indeed gain by tilting more toward information-rich stocks, we conduct performance analysis in Panel B of Table 8. We compute a holdings-based N.Singals Measure at fund level and define High_Info as a dummy variable identifying funds ranked in the top-tercile in terms of holding information-rich stocks, where N.Singals are based on total information and its subcategories, respectively. We interact AMLT with High_Info and the results show that funds holding more information-rich stocks to leverage AI technology indeed perform better. Using total information in column (1) as an example, tilting toward stocks with a large number of information inputs enhances fund performance under the AMLT strategy by 54.8%. The results for the subcategories in the next three columns suggest that, consistent with funds tilting more toward quantitative information and news than regulatory filings, both types of information play particularly important roles in enhancing AMLT performance.

5.1.2 Information Diversity and Interactions

In addition to processing vast amounts of information, another key strength of ML technology lies in its flexibility to optimize and fit unspecified functional forms, allowing for highly complex interactions among diverse information inputs, a critical factor in generating valuable signals. Next, we explore how this strength enables ML-adopting funds to create value.

Non-Linear Interactions

We start by investigating the value added by ML technology in accommodating complex interactions among information sets. For this purpose, we consider two quasi-ML benchmarks that rely primarily on linear structures. The first is the Elastic Net (ENet) (Cao et al., 2024), which captures the essence of dimension reduction by ML models but relies solely on linear combinations of predictors. The second is Principal Components Regression (PCR), which extracts linear factors from the predictor set to reduce dimensionality before generating signals.

We estimate both ENet and PCR models to generate trading signals using the full set of information as in our base ML model. We then compute AMLT_ENet and AMLT_PCR by following Equation 1, replacing ML signals with ENet or PCR trading signals, respectively.

We examine the performance of the full AMLT strategy versus the AMLT_ENet and AMLT_PCR strategies in Panel A of Table 9. It is evident from Panel A that both the AMLT_ENet and AMLT_PCR strategies, which use the same amount of information but are based on technologies that do not accommodate nonlinear predictive relationships, generate performance that is only 46.3% of that of the full AMLT. This exercise clearly demonstrates that a key feature of the full-fledged ML model, its ability to capture complex interactions, is vital for delivering strong AMLT performance.

Quantitative versus Textual Information

Next, we examine the value of ML technology's ability to combine diverse information inputs. With the rise of big data, investors are increasingly utilizing not only traditional financial indicators but also new unstructured textual data. We investigate the value added from combining quantitative information with textual information in generating ML-based strategies. To do so, we classify the variables into two broad categories: quantitative signals and textual signals, then re-estimate our ML models using each type separately to generate trading signals and compute AMLT_quant and AMLT_textual measures accordingly.

Performance results are provided in Panel B of Table 9, where we regress Carhart alpha on the full AMLT, AMLT_quant, and AMLT_text separately. Results in columns (1) to (3) show that, in addition to the full AMLT strategy, both AMLT_quant and AMLT_text strategies are associated with significant future fund performance. However, in terms of magnitude, AMLT_quant and AMLT_text are able to generate performance that is only about one-half and one-third, respectively, of the magnitude of the full AMLT strategy. The findings illustrate the substantial value added by combining diverse information in ML to generate trading signals. They also highlight the power of adding unstructured textual information into the traditional ML model based on quantitative data in boosting the profits of investment strategy, which has not been explored before in the literature.

Overall, the results so far demonstrate that the superior performance of AMLT is driven by the capacity of ML models to effectively integrate diverse information and accommodate complex interactions among information inputs.

Multiple versus Individual Information Input

We next classify our information inputs into six anomaly and three textual-based categories to further understand the value of ML-based strategies that combine multiple information inputs versus those relying on a single specific input. We classify quantitative information into six anomaly categories following Hou et al. (2020): Momentum, Value versus Growth, Investments, Profitability, Intangibles, and Trading Frictions, and textual-based information into three categories: News, Earnings Calls transcripts, and SEC financial filings. We estimate nine ML models separately using each of the nine information sets.[21] We then feed each set of resulting trading signals into our AMLT calculations (to replace the ML signals) and derive nine AMLT measures based on the subset of information (i.e., AMLT_sub). We examine the performance of these strategies. To better understand both the long and short

[21] Given the smaller number of predictors per category, between 10 and 27, we employ a two-layer DNN with 32 and 16 neurons, following Gu et al. (2020). We demand at least 5 years of data availability for each category.

sides of these strategies, Figure 5 plots the Carhart alphas for the Long (top decile), Short (bottom decile), and Long-minus-Short portfolios based on these AMLT_sub strategies and the full AMLT strategy. Statistical significance is indicated by the color of the bars, with dark color representing significance and light color indicating insignificance. First, it is evident from both Panel B (Long) and Panel C (Long-minus-Short) that these AMLT_sub strategies fail to produce performance even close to that of the full AMLT, represented by the black bar on the right.

Second, examining these AMLT_sub strategies more closely, we find that none of the six anomaly categories exhibit significant performance.[22] This may reflect the fleeting nature of the signals, the fading of in-sample anomaly profits in recent years, and the concentration of the profits on the short side, making them difficult for mutual funds to capture.(McLean and Pontiff (2016); Avramov et al. (2023)). This is also consistent with the lack of evidence in the mutual fund literature on funds' ability to exploit individual anomalies.[23] Next, tuning to textual information, we observe that trading based on individual textual inputs, such as earnings calls, leads to positive but statistically insignificant performance.

Overall, our findings further illustrate the substantial value added by ML models in extracting information from a combination of multiple information categories, as opposed to relying on any individual information input. Moreover, these results also indicate that AMLT is a distinct strategy that some funds follow.

5.2 Sustained Performance and Managerial Skills

After understanding the role of ML technology and information in generating AMLT performance, we next turn to the role of managerial investment expertise.

Despite the considerable promise of ML-based trading signals, the literature has shown that implementing such strategies in the real world can be challenging. In this subsection, we investigate whether mutual funds can achieve sustained performance by adopting AMLT strategy, which helps to shed light on managerial skills in successfully implementing ML strategy in practice.

[22] For example, Momentum, a major anomaly category, shows insignificant and negative performance. In untabulated analyses, the estimated coefficient on Momentum is positive and insignificant when the models are estimated without controlling for the Momentum factor.

[23] In fact, the literature documents the "dumb money" effect of mutual fund flows toward major individual anomalies (Akbas et al., 2015).

Long-Term Performance

Since ML models tend to generate fleeting signals requiring high turnover and trading costs, sustainable investment profit may be hard to achieve. Therefore, an important question is whether mutual funds are capable of implementing the AMLT strategy to gain long-term performance.

To understand this, we investigate how AMLT is associated with future fund performance in the subsequent eight quarters using our baseline regression in Equation (2) with a full set of controls. In Table A.5, we report 3 snapshots: performance in the subsequent quarter (Quarter t), the first year (Quarter $t + 4$), and the 2nd year (Quarter $t + 8$). The results show that AMLT is significantly related to long-term fund performance.

AMLT Performance and Market Conditions

Another challenge of implementing ML-based strategy in practice is that the economic gains tend to be concentrated in volatile market conditions (Avramov et al., 2023). In Table 10, we investigate the performance outcomes associated with the AMLT strategy in different market conditions.

We define high and low market conditions based on the median cutoffs of three well-known market indicators, i.e., investor sentiment from Baker and Wurgler (2007) (Sentiment), monthly VIX index of implied volatilities of S&P 500 index options (VIX), and market illiquidity (Market_illiquidity), defined as the equally-weighted average of stock-level Amihud (2002) illiquidity for all NYSE/AMEX stocks in a month (Avramov et al., 2016). We augment our baseline regression in Equation (2) by High, an indicator variable denoting above median market condition, and an interaction term, AMLT $\times$ High. Our results, as shown in Table 10, demonstrate that funds with high AMLT outperform in both high and low market conditions, as indicated by the significantly positive coefficients associated with AMLT. In addition, consistent with previous literature that ML strategy generates greater profit in high market conditions, we find significant positive coefficients on all three interactions terms.

AMLT Performance and Fund Styles

ML-based trading strategies also dominate in small and illiquid stocks (Avramov et al. (2023); Jensen et al. (2024)), posing an additional challenge for practical implementation for mutual funds, especially since funds may need to follow style mandates and may not be able to invest in these stocks.

In Table 11, we investigate whether AMLT strategy can be implemented successfully by funds in different style groups. We divide our sample funds into four style groups based on the size and value scores of fund holdings following Daniel et al. (1997) and run our baseline regression for each subsample. Our results show that funds in all style groups are able to add value by utilizing AMLT strategy.

Overall, our evidence demonstrates that, despite challenges in ML-based trading, AMLT-adopting funds are able to generate persistent long-term performance and perform well in all market conditions and across all investment styles. In addition, these funds also manage trading costs well, as evidenced by their lower portfolio turnover and higher return gap. Our results therefore echo the insights of Cao et al. (2024) in that superior mutual fund performance from AMLT is the result of a powerful combination of human investment expertise and advanced ML technology.

6 Additional Analyses

In this section, we perform a series of additional analysis and robustness tests.

To evaluate mutual funds' ability to adopt ML technology as it advances, we conduct subsample analyses for earlier and later periods. In addition, we also examine an AMLT strategy based on a simpler ML model, the Random Forest, introduced at the beginning of our sample period in 2001 by Breiman (2001). In Appendix Table A.6, we show that AMLT constructed from Neural Network signals predicts fund performance in both the pre-2010 and post-2010 periods, with stronger predictive power in the later period. In addition, AMLT based on Random Forest model also yields significant performance, despite being weaker. Overall, these results suggest that mutual funds progressively improve their utilization of ML strategies as the technology advances.

To understand the strength of alternative AI adoption measures, in Appendix Table A.7, we re-estimate our baseline performance regression using both the holdings-based AMLT and the labor market information-based Share_CoreAI measures. We observe in column (1) that Share_CoreAI is positively related to future fund performance with a weak statistical significance. When putting both measures together in column (2), AMLT remains a significant and strong predictor of future fund performance while Share_CoreAI becomes insignificant. The finding illustrates the strength of the AMLT measure in capturing AI adoption specifically aimed at enhancing trading strategies and portfolio performance. In contrast, the AI talent measure is weak in predicting fund performance, which is consistent with that such labor market based measure can capture AI adoption for broader business operations unrelated to investment strategies, and may be coarse due to data coverage limitations.

Further, in our baseline analysis, we evaluate a fund's ML-based investment strategy relative to its benchmark portfolio by measuring deviations across its entire portfolio positions. As an alternative approach, we consider a purely trade-based measure, the turnover-adjusted ML trading measure (TOMLT), defined as a fund's dollar trading in a given quarter that follows ML signals (i.e., stocks in the top and bottom deciles of ML signals), divided by the fund's total dollar trading during the quarter. TOMLT focuses on mutual fund trading (i.e., weight change) toward strong ML signals, and adjusts for the fund's overall quarterly trading activity. We report consistent results in Appendix Table A.8: the alternative trading-based measure significantly and positively predicts future fund performance.

As robustness checks, we also consider alternative AMLT measures. First, in our baseline analysis, we use contemporaneous information available at quarter t to generate trading signals and compute AMLT at the same quarter, because the strength of ML model is its ability to rapidly incorporate the most up-to-date information. As a robustness check, we construct an alternative AMLT measure, denoted AMLT_begQInfo, which relies on beginning-of-quarter information. Specifically, AMLT in quarter t is computed using forward-looking ML-based trading signals generated from information available at the end of the previous quarter (i.e., quarter $t-1$). The results, reported in Appendix Table A.2, are highly consistent with our baseline findings.

Second, we examine alternative transformations of the ML signals. Specifically, we implement a more granular rank-based approach using 100 portfolios and an alternative transformation that places greater weight on extreme ML signals (i.e., assigning scores from -3 to +3 to stocks in the bottom and top deciles and zero to intermediate stocks). For both alternatives, we compute AMLT and AMLT_long measures. Appendix Table A.4 reports consistent results using these transformations.

7 Conclusion

Our study provides a comprehensive analysis of how U.S. equity mutual funds integrate cutting-edge ML technology into their investment strategies over the past two decades. We document a significant rising trend of ML adoption by mutual funds and substantial value added by such practice, with the superior performance attributable to better stock selection, lower expenses, and efficient trading-cost management. We identify two key mechanisms behind AMLT performance: the exceptional information-processing capacity of ML technology and fund managers' ability to integrate ML with human expertise to successfully implement AI-powered strategies in practice.

We construct a novel Active Machine-Learning-Based Trading (AMLT) measure, a signed active-share measure that captures how mutual funds actively align their portfolio weight deviations from its passive benchmark with forward-looking ML-based trading signals. Our methodology contributes to both mutual fund performance evaluation and the identification of AI adoptions in investments. The flexibility of the measure further enables analysis of the mechanisms behind performance and can be readily extended to other studies.

In addition to the holdings-based approach, we examine employee AI talent across all mutual fund companies to provide additional evidence on ML adoption based on labor market information. Together, these analyses offer a comprehensive and multi-dimensional investigation of ML adoption in mutual fund investments.

We contribute to the new and growing literature on the utilization of ML models in various areas in finance by providing the much needed evidence of real-world application of ML-based trading signals in mutual fund investment practices. Our results highlight how AI is reshaping traditional investment practices and how integrating ML-based signal with human investment skill can enhance decision making and generate long-term value. As the asset management industry continues to evolve with advancements in AI and alternative data sources, understanding the complementary role of AI and human expertise will remain a critical area for future research.

References

Abis, S. and L. Veldkamp (2024). The changing economics of knowledge production. *The Review of Financial Studies 37*(1), 89–118.

Akbas, F., W. J. Armstrong, S. Sorescu, and A. Subrahmanyam (2015). Smart money, dumb money, and capital market anomalies. *Journal of Financial Economics 118*(2), 355–382.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Avramov, D., S. Cheng, and A. Hameed (2016). Time-varying liquidity and momentum profits. *Journal of Financial and Quantitative Analysis 51*(6), 1897–1923.

Avramov, D., S. Cheng, and L. Metzker (2023). Machine learning vs. economic restrictions: Evidence from stock return predictability. *Management Science 69*(5), 2587–2619.

Babina, T., A. Fedyk, A. He, and J. Hodson (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics 151*, 103745.

Baker, M. and J. Wurgler (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives 21*(2), 129–151.

Bianchi, D., M. Büchner, and A. Tamoni (2021). Bond risk premiums with machine learning. *The Review of Financial Studies 34*(2), 1046–1089.

Bonelli, M. and T. Foucault (2023). Displaced by big data: Evidence from active fund managers. Working Paper.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

Cao, S., W. Jiang, J. Wang, and B. Yang (2024). From man vs. machine to man+ machine: The art and ai of stock analyses. *Journal of Financial Economics 160*, 103910.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance 52*(1), 57–82.

Cen, L., B. Han, Y. Han, and C. Jo (2024). Data scientists on wall street. Working Paper.

Chen, J., H. Hong, M. Huang, and J. D. Kubik (2004). Does fund size erode mutual fund performance? the role of liquidity and organization. *American Economic Review 94*(5), 1276–1302.

Chen, L., M. Pelger, and J. Zhu (2024). Deep learning in asset pricing. *Management Science 70*(2), 714–750.

Chen, S., C. Sialm, and D. X. Xu (2025). Artificial intelligence across asset management: Evidence from registered investment advisers. Working Paper.

Choi, D., W. Jiang, and C. Zhang (2024). Alpha go everywhere: Machine learning and international stock returns. Working Paper.

Cohen, L., A. Frazzini, and C. Malloy (2008). The small world of investing: Board connections and mutual fund returns. *Journal of Political Economy 116*(5), 951–979.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2022). Alphaportfolio: Direct construction through deep reinforcement learning and interpretable ai. *3554486*. Working Paper.

Cremers, K. M. and A. Petajisto (2009). How active is your fund manager? a new measure that predicts performance. *The Review of Financial Studies 22*(9), 3329–3365.

Cremers, M., M. A. Ferreira, P. Matos, and L. T. Starks (2013). The mutual fund industry worldwide: explicit and closet indexing, fees, and performance. Working Paper.

D'Acunto, F., N. Prabhala, and A. G. Rossi (2019). The promises and pitfalls of robo-advising. *The Review of Financial Studies 32*(5), 1983–2020.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance 52*(3), 1035–1058.

DeMiguel, V., J. Gil-Bazo, F. J. Nogales, and A. A. Santos (2023). Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics 150*(3), 103737.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pp. 1–15. Springer.

Evans, R. B. (2010). Mutual fund incubation. *The Journal of Finance 65*(4), 1581–1611.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fang, L. H., J. Peress, and L. Zheng (2014). Does media coverage of stocks affect mutual funds' trading and performance? *The Review of Financial Studies 27*(12), 3441–3466.

Feng, G., J. He, N. G. Polson, and J. Xu (2024). Deep learning in characteristics-sorted factor models. *Journal of Financial and Quantitative Analysis 59*(7), 3001–3036.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance 77*(1), 5–47.

Garcia, D., X. Hu, and M. Rohrer (2023). The colour of finance words. *Journal of Financial Economics 147*(3), 525–549.

Gil-Bazo, J. and P. Ruiz-Verdu (2009). The relation between price and performance in the mutual fund industry. *Journal of Finance 64*, 2153–2183.

Green, J., J. R. Hand, and X. F. Zhang (2017). The characteristics that provide independent information about average us monthly stock returns. *The Review of Financial Studies 30*(12), 4389–4436.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies 33*(5), 2223–2273.

Han, Y., A. He, D. E. Rapach, and G. Zhou (2024). Cross-sectional expected returns: new fama–macbeth regressions in the era of machine learning. *Review of Finance 28*(6), 1807–1831.

Hanley, K. W. and G. Hoberg (2019). Dynamic interpretation of emerging risks in the financial sector. *The Review of Financial Studies 32*(12), 4543–4603.

Hansen, L. K. and P. Salamon (1990). Neural network ensembles. *IEEE transactions on* *pattern analysis and machine intelligence 12*(10), 993–1001.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *The Review of Financial Studies 33*(5), 2019–2133.

Huang, J., C. Sialm, and H. Zhang (2011). Risk shifting and mutual fund performance. *The Review of Financial Studies 24*(8), 2575–2616.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance 48*(1), 65–91.

Jensen, T. I., B. T. Kelly, S. Malamud, and L. H. Pedersen (2024). Machine learning and the implementable efficient frontier. (22-63). Working Paper.

Jeon, Y., T. H. McCurdy, and X. Zhao (2022). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics 145*(2), 1–17.

Jiang, H., S. Z. Li, and H. Wang (2021). Pervasive underreaction: Evidence from high-frequency data. *Journal of Financial Economics 141*(2), 573–599.

Jiang, H. and L. Zheng (2018). Active fundamental performance. *The Review of Financial Studies 31*(12), 4688–4719.

Kacperczyk, M. and A. Seru (2007). Fund manager use of public information: New evidence on managerial skills. *The Journal of Finance 62*(2), 485–528.

Kacperczyk, M., C. Sialm, and L. Zheng (2008). Unobserved actions of mutual funds. *The Review of Financial Studies 21*(6), 2379–2416.

Kacperczyk, M., S. Van Nieuwerburgh, and L. Veldkamp (2012). Rational attention allocation over the business cycle. Technical report, National Bureau of Economic Research. Working Paper.

Kacperczyk, M. T., C. Sialm, and L. Zheng (2007). Industry concentration and mutual fund performance. *Journal of Investment Management 5*(1), 50–64.

Kaniel, R., Z. Lin, M. Pelger, and S. Van Nieuwerburgh (2023). Machine-learning the skill of mutual fund managers. *Journal of Financial Economics 150*(1), 94–138.

Ke, Z. T., B. T. Kelly, and D. Xiu (2019). Predicting returns with text data. NBER Working Paper.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the cross-section. *Journal of Financial Economics 135*(2), 271–292.

LeBaron, B. (1999). Technical trading rule profitability and foreign exchange intervention. *Journal of International Economics 49*(1), 125–143.

Lettau, M. and M. Pelger (2020). Factors that fit the time series and cross-section of stock returns. *The Review of Financial Studies 33*(5), 2274–2325.

Li, B. and A. G. Rossi (2020). Selecting mutual funds from the stocks they hold: A machine learning approach. Working Paper.

Liu, M. (2022). Assessing human information processing in lending decisions: A machine learning approach. *Journal of Accounting Research 60*(2), 607–651.

Lopez-Lira, A. and Y. Tang (2023). Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619*.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance 66*(1), 35–65.

McCulloch, W. S. and W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics 5*, 115–133.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *The Journal of Finance 71*(1), 5–32.

Mitchell, T. (1997). Machine learning. pp. 31.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Petajisto, A. (2013). Active share and mutual fund performance. *Financial Analysts Journal 69*(4), 73–93.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). Learning representations by back-propagating errors. *Nature 323*(6088), 533–536.

Russell, S. J. and P. Norvig (2016). *Artificial intelligence: a modern approach*. pearson.

Sheng, J., Z. Sun, B. Yang, and A. L. Zhang (2024). Generative ai and asset management. Working Paper.

Shumway, T. and V. A. Warther (1999). The delisting bias in crsp's nasdaq data and its implications for the size effect. *The Journal of Finance 54*(6), 2361–2379.

Sialm, C. and H. Zhang (2020). Tax-efficient asset management: evidence from equity mutual funds. *The Journal of Finance 75*(2), 735–777.

Stambaugh, R. F. (2014). Presidential address: Investment noise and trends. *The Journal of Finance 69*(4), 1415–1453.

Van Binsbergen, J. H., X. Han, and A. Lopez-Lira (2023). Man versus machine learning: The term structure of earnings expectations and conditional biases. *The Review of Financial Studies 36*(6), 2361–2396.

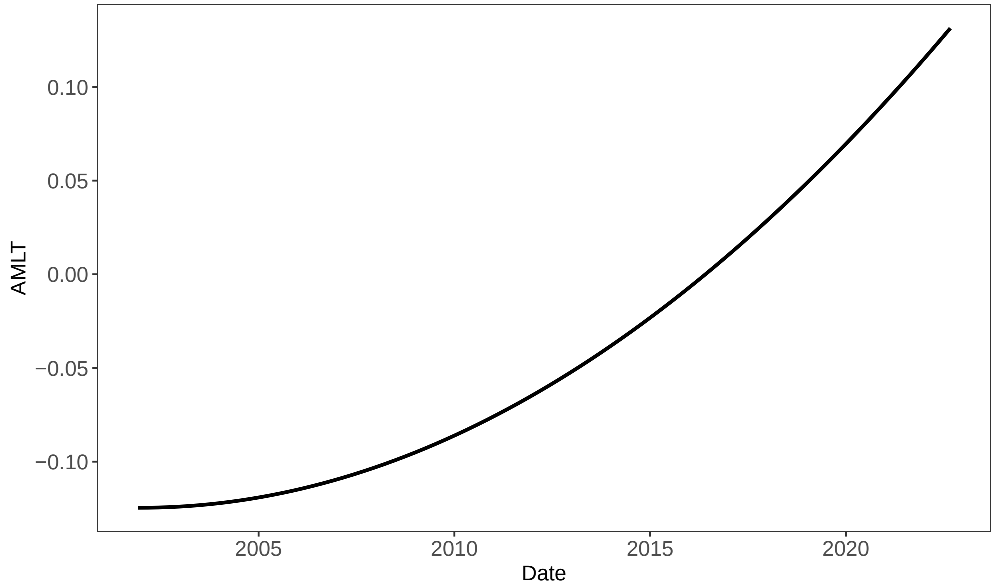
Von Beschwitz, B., D. B. Keim, and M. Massa (2020). First to âreadâ the news: News analytics and algorithmic trading. *The Review of Asset Pricing Studies 10*(1), 122–178.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies 21*(4), 1455–1508.

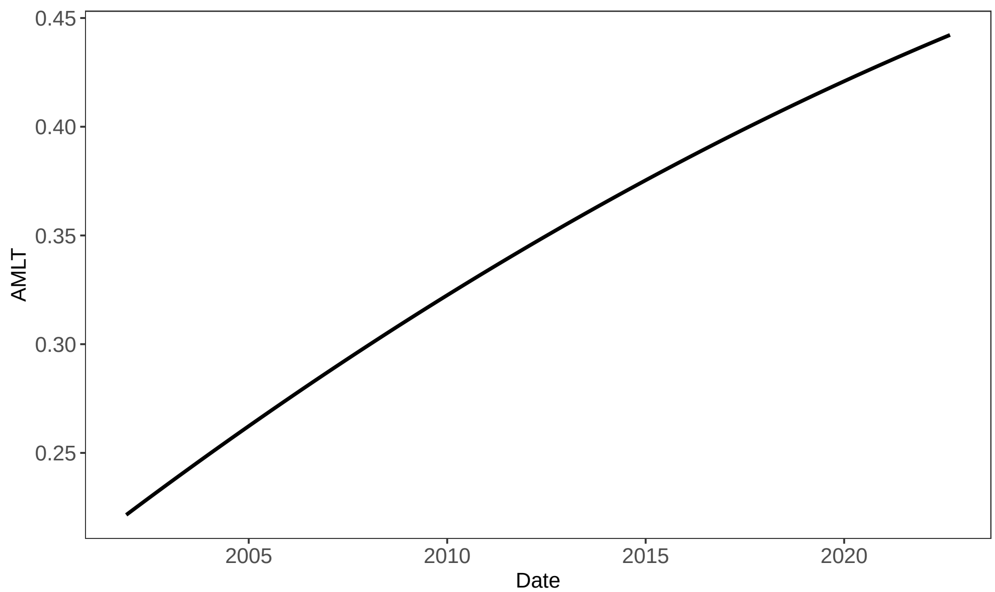
Wermers, R. (2000). Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance 55*(4), 1655–1695.

Zhang, Y. (2024). Do mutual funds benefit from the adoption of ai technology? Working Paper.

Zheng, X. (2024). How can innovation screening be improved? a machine learning analysis with economic consequences for firm performance. *Forthcoming: Journal of Financial and Quantitative Analysis*.



(a) 50th percentile



(b) 80th percentile

Figure 1: **Time-series Variation in AMLT**

The figure depicts second-order polynomial fits to the quarterly cross-sectional distribution of AMLT, a fund-level signed active share measure defined in Equation (1) using forward-looking ML trading signals. Panel A (B) reports the 50th (80th) percentile of AMLT.

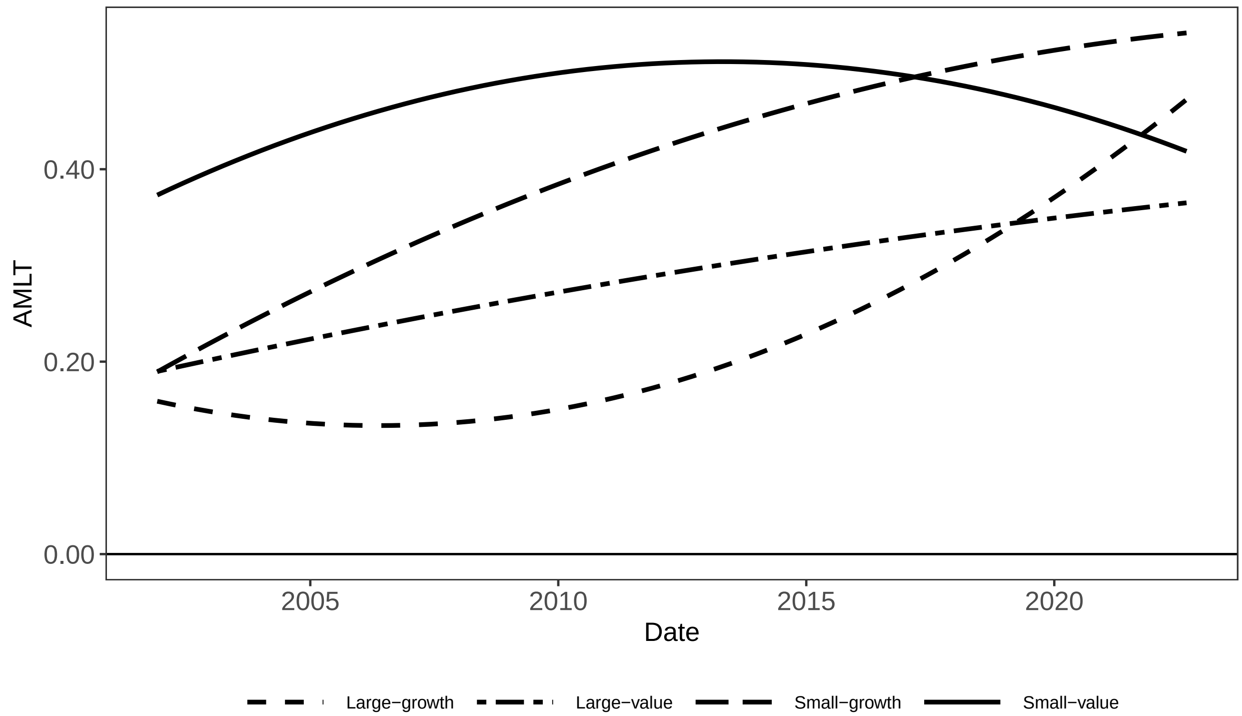


Figure 2: **Time-series Variation in AMLT per Style**

The figure depicts second-order polynomial fits to the quarterly 80th percentile of AMLT by fund style. Each quarter, funds are assigned to one of four styles. Funds are first split into two groups based on a size score and, within each size group, into two groups based on a value score. Size and value scores are constructed from the average quintile rankings of portfolio holdings using NYSE breakpoints for market capitalization and book-to-market ratios.

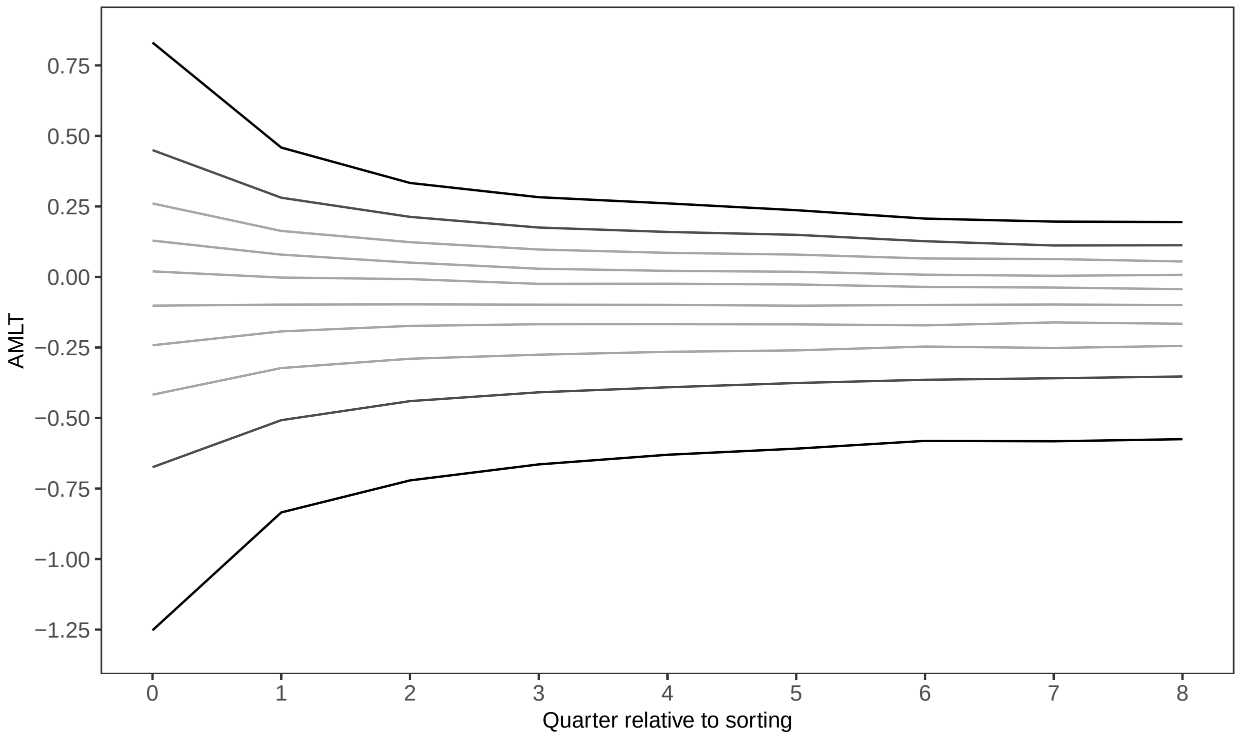


Figure 3: **Persistence of AMLT**

The figure depicts average AMLT levels for ten portfolios sorted on AMLT over the subsequent eight quarters.

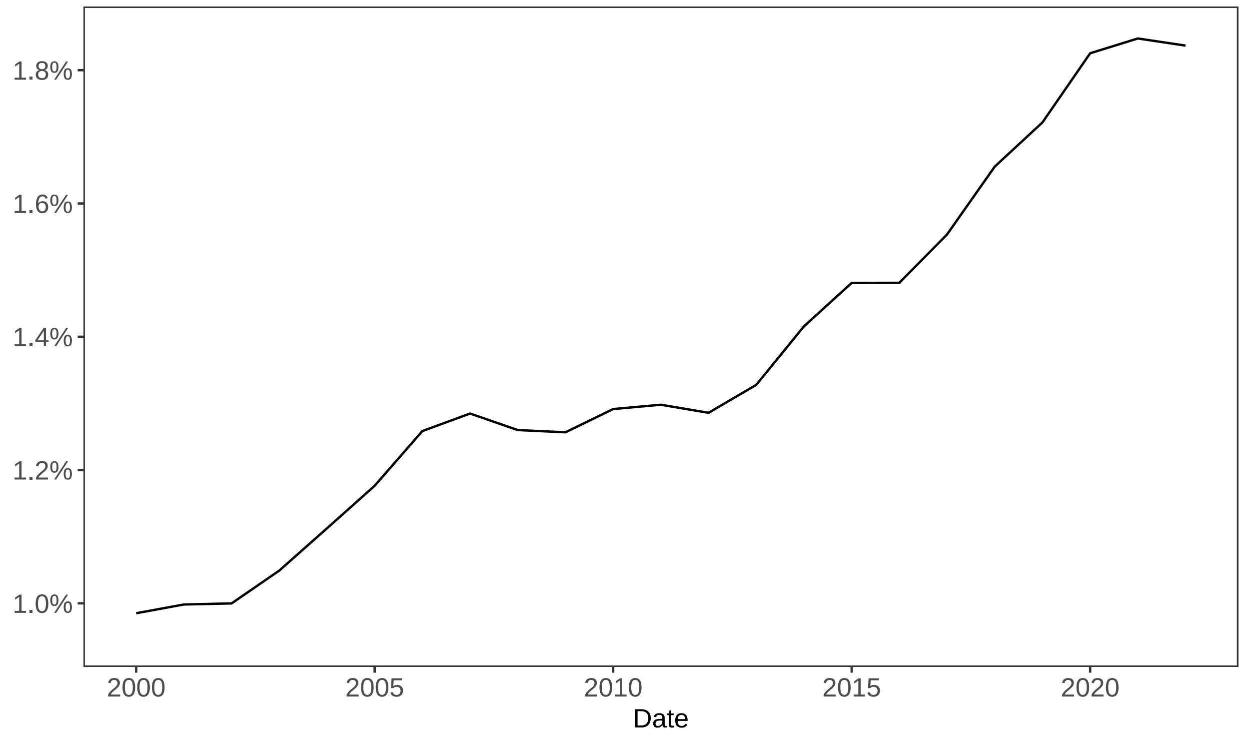
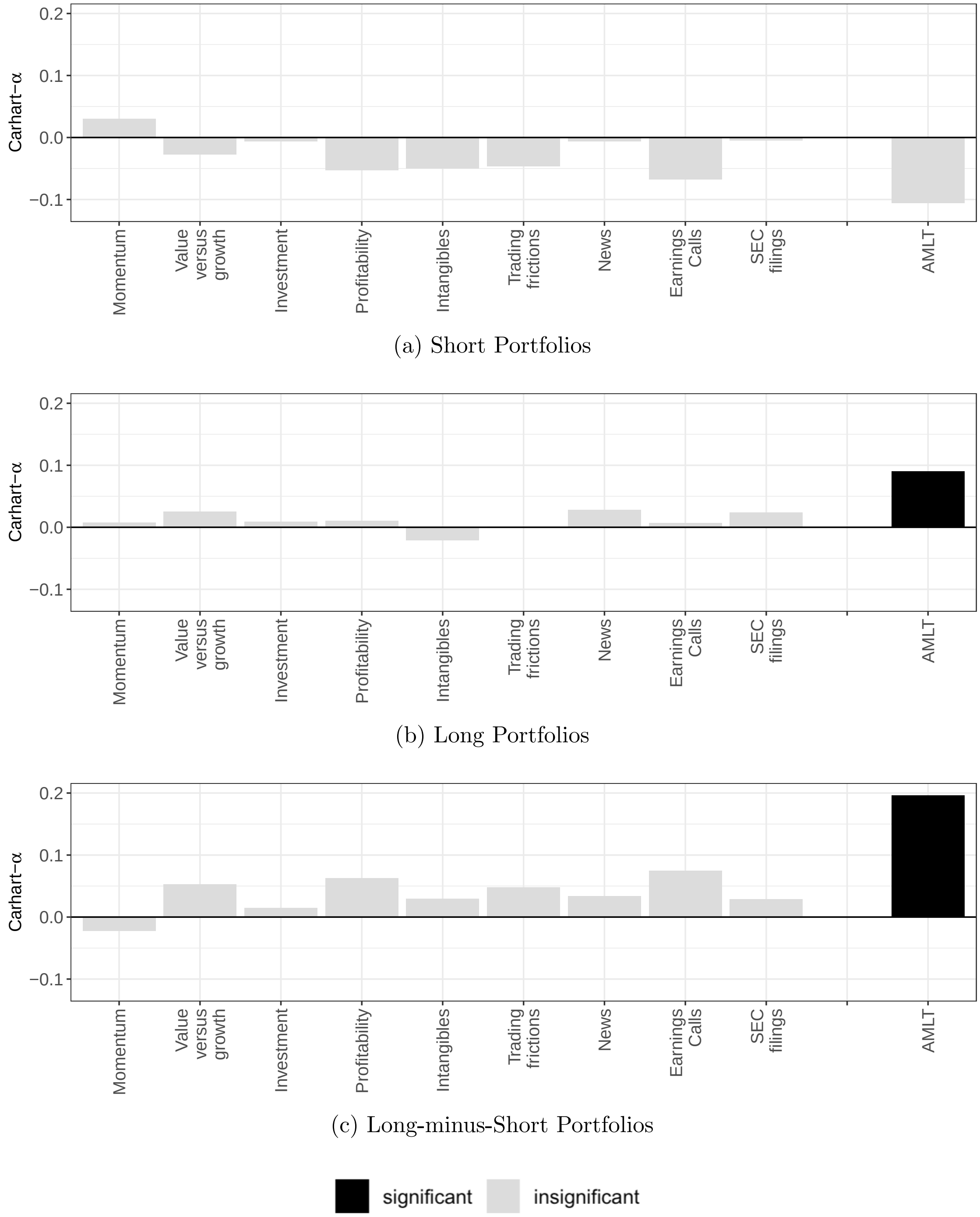


Figure 4: **Time-series Variation in Percentage of Employees with AI Talent**

The figure depicts the percentage of employees with AI talent at mutual fund advisory companies. An employee is classified as having AI talent if their profile lists at least one of the following core skills: artificial intelligence, machine learning, natural language processing, or computer vision.



(a) Short Portfolios

(b) Long Portfolios

(c) Long-minus-Short Portfolios

Figure 5: **Machine-Learning Models with Individual Information Set**

The figure depicts average monthly Carhart-α returns of mutual fund portfolios sorted by the most recent AMLT measure under alternative restrictions on the information set inputs of the machine-learning model. We consider nine information sets. Panel A (B) reports returns for the short (long) portfolio, and Panel C reports returns for the long-minus-short portfolio. Intercepts are obtained from Carhart four-factor regressions.

Table 1: **Descriptive Statistics**

This table reports descriptive statistics for the sample of mutual funds at the monthly frequency. AMLT is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. *AMLT* is constructed using ML signals for all positions, while *AMLT_long* uses ML signals for long-side trading opportunities only. *AMLT_tilt* is a simplified AMLT measure that assigns +1 or -1 to each active deviation depending on whether it aligns with the ML signal. The table reports summary statistics for fund characteristics, fund returns, risk-adjusted returns, performance decomposition measures, and AI talent. Risk-adjusted returns are measured by Carhart-α. Characteristic selectivity (CS), characteristic timing (CT), and average style (AS) follow Daniel et al. (1997), and return gap (RG) follows Kacperczyk et al. (2008). *Share_CoreAI* is the share of employees of a registered investment adviser with AI talent. An employee is classified as having AI talent if the employee's profile lists at least one of the following core skills: artificial intelligence, machine learning, natural language processing, or computer vision. The table reports the mean, standard deviation, 25th percentile, median, and 75th percentile of each variable.

Variable	Mean	Std. Dev.	q25	Median	q75
AMLT	-0.109	0.649	-0.407	-0.007	0.269
AMLT_long	-0.071	0.432	-0.279	-0.007	0.183
AMLT_tilt	-0.027	0.182	-0.110	-0.001	0.077
Active share (in %)	70.544	23.187	61.670	75.781	87.527
Turnover (in % per year)	61.509	48.835	26.000	50.000	84.000
Fund age (in years)	15.370	13.516	6.000	12.000	20.000
TNA (in $M)	2,401.968	12,113.415	94.700	345.300	1,334.900
Past perf (in % per month)	0.720	1.600	-0.027	0.894	1.638
Total risk (in % per month)	4.842	1.725	3.546	4.661	5.869
β_{MKT}	1.033	0.221	0.927	1.015	1.139
Idio vol (in % per month)	1.158	0.833	0.736	1.009	1.386
Size (range [1 to 5])	4.148	0.959	3.507	4.622	4.895
Value (range [1 to 5])	2.862	0.412	2.573	2.880	3.145
Momentum score (range [1 to 5])	3.095	0.429	2.802	3.088	3.367
Share_CoreAI (in %)	0.829	2.392	0.000	0.000	0.752
Net return (net, in % per month)	0.676	5.074	-1.854	1.157	3.626
Gross return (gross, in % per month)	0.763	5.069	-1.766	1.245	3.708
Carhart-α (net, in % per month)	-0.131	1.615	-0.864	-0.101	0.624
Carhart-α (gross, in % per month)	-0.045	1.615	-0.778	-0.025	0.709
CS (in % per month)	0.023	1.400	-0.670	0.023	0.726
CT (in % per month)	-0.004	0.703	-0.331	0.001	0.322
AS (in % per month)	0.809	4.880	-1.653	1.357	3.592
Exp ratio (in % per month)	0.085	0.038	0.063	0.085	0.106
RG (in % per month)	-0.077	0.987	-0.362	-0.059	0.221

Table 2: **Fund Characteristics**

This table reports fund characteristics by AMLT and examines the determinants of AMLT. AMLT is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. Panel A reports characteristics of mutual fund portfolios sorted on the most recent *AMLT* at portfolio formation. The reported characteristics include *AMLT_tilt*, a simplified AMLT measure that assigns a value of +1 or -1 to each active deviation depending on whether it aligns with the ML signal, expense ratio, turnover, fund age, total net assets, family-level total net assets, fund past performance, total risk, market beta, idiosyncratic volatility, and active share. In each quarter, we compute the cross-sectional average of each characteristic within each portfolio and then report time-series averages across quarters. The high-minus-low differences are computed quarterly using Newey-West standard errors with four lags. Panel B reports coefficient estimates and standard errors from quarterly regressions analyzing the relation between AMLT and fund characteristics at portfolio formation. Fund characteristics include expense ratio, fund size, family size, turnover, fund age, fund performance, *Ret*, measured as the average fund return over the previous 12 months, total risk, style scores, recession, and market volatility. Style scores are based on the average quintile ranks of portfolio holdings' market capitalization and book-to-market ratio using NYSE breakpoints. All regressions include time fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A – Fund Characteristics by AMLT

Portfolio	AMLT	AMLT_tilt	Exp ratio	Turnover	Fund age	TNA	TNA family	Past perf	Total risk	β_{MKT}	Idio vol	Active share
1	-1.354	-0.365	1.217	0.707	15.262	1,720	57,822	0.653	5.578	1.152	1.683	0.850
2	-0.691	-0.179	1.120	0.659	16.107	2,060	70,429	0.718	5.093	1.080	1.331	0.780
3	-0.427	-0.108	1.069	0.638	16.377	2,440	82,883	0.706	4.905	1.048	1.204	0.739
4	-0.244	-0.058	1.009	0.606	16.451	3,460	99,455	0.720	4.779	1.025	1.097	0.676
5	-0.102	-0.023	0.933	0.562	16.151	3,785	113,950	0.724	4.698	1.011	1.027	0.609
6	0.016	0.007	0.846	0.501	15.535	3,618	99,057	0.729	4.643	1.005	0.942	0.543
7	0.122	0.033	0.891	0.526	15.561	2,414	74,419	0.715	4.619	0.996	0.952	0.595
8	0.259	0.069	1.003	0.591	15.797	1,919	67,164	0.719	4.648	0.992	1.044	0.700
9	0.451	0.119	1.041	0.589	15.460	1,758	59,322	0.734	4.716	1.000	1.107	0.748
10	0.872	0.233	1.084	0.575	14.627	1,486	47,827	0.774	4.813	1.009	1.228	0.816
H-L	2.227***	0.598***	-0.133***	-0.133***	-0.636***	-235*	-9,995**	0.121	-0.765***	-0.143***	-0.455***	-0.034***
	(0.098)	(0.023)	(0.011)	(0.023)	(0.229)	(121.614)	(4,914.243)	(0.083)	(0.144)	(0.031)	(0.035)	(0.005)

Panel B – Determinants of AMLT

Dependent Variable:	AMLT$_t$		
Model:	(1)	(2)	(3)
Exp ratio$_t$	-0.148***	-0.168***	-0.236***
	(0.018)	(0.018)	(0.018)
Log(TNA)$_t$	-0.023***	-0.021***	-0.025***
	(0.005)	(0.005)	(0.005)
Log(TNA family)$_t$	0.003	0.003	0.002
	(0.003)	(0.003)	(0.003)
Turnover$_t$	-0.095***	-0.094***	-0.142***
	(0.015)	(0.015)	(0.016)
Fund age$_t$	0.000	0.000	0.001***
	(0.001)	(0.001)	(0.001)
Ret$_t$	0.043***	0.035***	0.008***
	(0.006)	(0.006)	(0.002)
Total risk$_t$	-0.106***	-0.122***	-0.064***
	(0.008)	(0.009)	(0.004)
Size score$_t$		-0.061***	-0.034***
		(0.010)	(0.009)
Value score$_t$		0.096***	0.111***
		(0.018)	(0.018)
Momentum score$_t$		0.005	0.052***
		(0.014)	(0.013)
US recession$_t$			-0.105***
			(0.013)
High VIX$_t$			0.128***
			(0.008)
Time fixed effects	Yes	Yes	
Observations	140,576	140,576	140,576
Adjusted R^2	0.113	0.123	0.075

Table 3: **AMLT and Self-designated AI Funds and Quant Funds**

This table reports coefficient estimates and standard errors from quarterly regressions analyzing the relation between AMLT and self-designated fund strategies. The dependent variables are *AMLT*, *AMLT_quant*, and *AMLT_text*. AMLT is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of active deviations of fund portfolio weights from benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. In Columns (1) and (4), *AMLT* is constructed using machine-learning signals based on the full set of information inputs. In Columns (2) and (5), *AMLT_quant* is constructed using quantitative information inputs only. In Columns (3) and (6), *AMLT_text* is constructed using textual information inputs only. The main independent variables, *AI fund* and *Quant fund*, are indicators equal to one if a fund is identified as an AI fund or a quant fund, and zero otherwise. We identify five funds with "AI" in their fund names and an additional 41 through online searches, and 46 funds with "quant" or "quantitative" in their fund names. Control variables include fund performance, *Ret*; expense ratio, *Exp ratio*; fund size, *Log(TNA)*; turnover ratio, *Turnover*; and fund age, *Fund age*. All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variables:	$AMLT_t$	$AMLT\_quant_t$	$AMLT\_text_t$	$AMLT_t$	$AMLT\_quant_t$	$AMLT\_text_t$
Model:	(1)	(2)	(3)	(4)	(5)	(6)
AI $fund_t$	0.680***	0.694***	0.496***			
	(0.036)	(0.041)	(0.042)			
Quant $fund_t$				0.153***	0.188***	0.048
				(0.049)	(0.051)	(0.037)
Exp $ratio_t$	-0.171***	-0.173***	-0.133***	-0.173***	-0.174***	-0.135***
	(0.019)	(0.020)	(0.016)	(0.019)	(0.020)	(0.016)
$Log(TNA)_t$	-0.020***	-0.016***	-0.022***	-0.020***	-0.016***	-0.022***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)
$Turnover_t$	-0.119***	-0.126***	-0.076***	-0.121***	-0.128***	-0.076***
	(0.016)	(0.017)	(0.014)	(0.016)	(0.017)	(0.014)
Fund age_t	0.000	0.000	0.001	0.000	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ret_t	0.039***	0.045***	0.014***	0.039***	0.045***	0.014***
	(0.005)	(0.006)	(0.004)	(0.005)	(0.006)	(0.004)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	140,577	140,577	140,577	140,577	140,577	140,577
Adjusted R^2	0.095	0.086	0.059	0.089	0.080	0.055

Table 4: **AMLT and Employee AI Talent**

This table reports coefficient estimates and standard errors from quarterly regressions analyzing the relation between AMLT and AI talent. The dependent variable is *AMLT*. AMLT is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. The main independent variable is the share of employees of a registered investment adviser with AI talent, *Share_CoreAI*. An employee is classified as having AI talent if the employee's profile lists at least one of the following core skills: artificial intelligence, machine learning, natural language processing, or computer vision. Control variables include fund performance (*Ret*) measured as the average fund return over the previous 12 months, expense ratio (*Exp ratio*), fund size (*Log(TNA)*) measured as logarithm of TNA, turnover ratio (*Turnover*) and fund age (*Fund age*) the natural logarithm of one plus fund age. All control variables are measured over the prior year. All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable:	$AMLT_t$	
Model:	(1)	(2)
$Share\_CoreAI_t$	1.424***	1.247**
	(0.453)	(0.499)
Exp $ratio_t$		-0.156***
		(0.033)
$Log(TNA)_t$		-0.011
		(0.008)
$Turnover_t$		-0.111***
		(0.028)
Fund age_t		0.000
		(0.001)
Ret_t		0.034***
		(0.009)
Time fixed effects	Yes	Yes
Style fixed effects	Yes	Yes
Observations	42,076	40,876
Adjusted R^2	0.070	0.084

Table 5: **AMLT and Future Fund Performance: Portfolio Approach**

This table reports abnormal monthly returns of decile mutual fund portfolios sorted by the most recent AMLT measure. AMLT is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. *AMLT* is constructed using ML signals for all positions, while *AMLT_long* uses ML signals for long-side trading opportunities only. Decile portfolios are formed and rebalanced at the end of each quarter, and portfolio performance is computed over the subsequent three months for each decile and the high-minus-low (H-L) decile portfolio. The top (bottom) panel reports risk-adjusted returns based on the CAPM, the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, and the Pástor and Stambaugh (2003) five-factor model for portfolios sorted on *AMLT* (*AMLT_long*). Returns are expressed in percent per month. Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Portfolio	1	2	3	4	5	6	7	8	9	10	H-L
						AMLT					
CAPM-α	-0.117	-0.038	-0.016	-0.020	-0.002	-0.008	0.011	0.020	0.058	0.118*	0.235***
	(0.089)	(0.061)	(0.050)	(0.041)	(0.036)	(0.034)	(0.035)	(0.041)	(0.050)	(0.067)	(0.080)
FF-α	-0.131**	-0.047	-0.022	-0.025	-0.006	-0.012	0.008	0.015	0.051	0.108***	0.238***
	(0.065)	(0.046)	(0.039)	(0.031)	(0.028)	(0.026)	(0.025)	(0.028)	(0.031)	(0.040)	(0.079)
Carhart-α	-0.106	-0.045	-0.027	-0.031	-0.011	-0.019	0.002	0.007	0.042	0.090**	0.196**
	(0.064)	(0.046)	(0.039)	(0.031)	(0.028)	(0.026)	(0.025)	(0.028)	(0.031)	(0.040)	(0.076)
PS-α	-0.111*	-0.049	-0.029	-0.033	-0.013	-0.020	0.001	0.006	0.041	0.089**	0.199***
	(0.063)	(0.044)	(0.038)	(0.031)	(0.028)	(0.025)	(0.025)	(0.028)	(0.031)	(0.039)	(0.075)
						AMLT_long					
CAPM-α	-0.120	-0.034	-0.031	-0.013	-0.003	-0.010	0.000	0.030	0.063	0.122**	0.243***
	(0.082)	(0.060)	(0.049)	(0.042)	(0.038)	(0.037)	(0.038)	(0.043)	(0.050)	(0.062)	(0.079)
FF-α	-0.132**	-0.042	-0.037	-0.018	-0.008	-0.014	-0.004	0.024	0.057*	0.114***	0.246***
	(0.063)	(0.046)	(0.036)	(0.031)	(0.028)	(0.027)	(0.026)	(0.028)	(0.031)	(0.041)	(0.078)
Carhart-α	-0.107*	-0.041	-0.043	-0.024	-0.012	-0.022	-0.009	0.014	0.047	0.098**	0.205***
	(0.062)	(0.046)	(0.037)	(0.031)	(0.028)	(0.027)	(0.026)	(0.028)	(0.031)	(0.041)	(0.075)
PS-α	-0.112*	-0.044	-0.045	-0.026	-0.014	-0.023	-0.011	0.013	0.046	0.097**	0.208***
	(0.060)	(0.045)	(0.036)	(0.031)	(0.028)	(0.027)	(0.026)	(0.028)	(0.030)	(0.041)	(0.075)

Table 6: **AMLT and Future Fund Performance: Multivariate Regressions**

This table reports coefficient estimates and standard errors from monthly regressions of Equation (2) analyzing the relation between AMLT and monthly fund performance over the subsequent quarter. The dependent variable is the monthly abnormal fund return, Carhart-α, calculated as the difference between the fund return and the expected return based on factor loadings estimated using the Carhart (1997) four-factor model over the past 36 months in a rolling window. Carhart-α is estimated for both net and gross fund returns. The main independent variable, *AMLT*, is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. *AMLT* is constructed using ML signals for all positions, while *AMLT_long* uses ML signals for long-side trading opportunities only. Control variables include lagged fund performance (*Ret*) measured as the average fund return over the previous 12 months, expense ratio (*Exp ratio*), fund size (*Log(TNA)*) measured as logarithm of TNA, turnover ratio (*Turnover*) and fund age (*Fund age*) the natural logarithm of one plus fund age. All control variables are measured over the prior year. All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable:	Carhart-α_t			
	Net return	Gross return	Net return	Gross return
Model:	(1)	(2)	(3)	(4)
AMLT$_{t-1}$	0.054***	0.053***		
	(0.006)	(0.006)		
AMLT_long$_{t-1}$			0.086***	0.085***
			(0.008)	(0.008)
Ret$_{t-1}$	0.016**	0.015**	0.016**	0.016**
	(0.007)	(0.007)	(0.007)	(0.007)
Exp ratio$_{t-1}$	-1.080***	-0.125	-1.084***	-0.129
	(0.105)	(0.105)	(0.105)	(0.104)
Log(TNA)$_{t-1}$	-0.005***	-0.005***	-0.005***	-0.005***
	(0.002)	(0.002)	(0.002)	(0.002)
Turnover$_{t-1}$	-0.056***	-0.056***	-0.056***	-0.056***
	(0.007)	(0.007)	(0.007)	(0.007)
Fund age$_{t-1}$	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Time fixed effects	Yes	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes	Yes
Observations	406,897	406,462	406,897	406,462
Adjusted R^2	0.070	0.069	0.070	0.069

Table 7: **AMLT and Future Fund Performance: Performance Decomposition**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between AMLT and fund performance decomposition measures over the subsequent quarter. The dependent variables are characteristic selectivity (CS), characteristic timing (CT), average style (AS), expense ratio (*Exp Ratio*), and return gap (RG). The CS, CT, and AS measures follow Daniel et al. (1997), with benchmarks formed using market capitalization, dividend yield, and prior 12-month returns. The RG measure follows Kacperczyk et al. (2008). The main independent variable, *AMLT*, is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. Panel A reports univariate regressions without control variables. Panel B includes the same control variables as in Table 6. Because the expense ratio is highly persistent, lagged *Exp Ratio* is omitted as a control in column (4). All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A – Univariate Regression

Dependent Variables:	CS_t	CT_t	AS_t	Exp $Ratio_t$	RG_t
Model:	(1)	(2)	(3)	(4)	(5)
$AMLT_{t-1}$	0.050***	0.001	0.009***	-0.006***	0.008**
	(0.006)	(0.002)	(0.003)	(0.001)	(0.004)
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	430,322	418,605	418,605	429,791	415,982
Adjusted R^2	0.063	0.085	0.952	0.219	0.031

Panel B – Multivariate Regression

Dependent Variables:	CS_t	CT_t	AS_t	Exp $Ratio_t$	RG_t
Model:	(1)	(2)	(3)	(4)	(5)
$AMLT_{t-1}$	0.043***	0.002	0.010***	-0.005***	0.003
	(0.006)	(0.002)	(0.003)	(0.001)	(0.004)
Controls	Yes	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	414,576	409,157	409,157	414,447	403,875
Adjusted R^2	0.066	0.085	0.952	0.331	0.032

Table 8: **AMLT and Volume of Information**

Panel A reports coefficient estimates and standard errors from quarterly regressions analyzing the relation between AMLT and active portfolio tilt. The dependent variables are active portfolio tilt toward stocks with a high number of signals (*ATILT_N.Signals*), quantitative (*ATILT_N.Signals.quant*), regulatory filings (*ATILT_N.Filings*), and newspaper articles (*ATILT_N.Articles*). The main independent variable, *AMLT*, is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. Panel B reports coefficient estimates and standard errors from monthly regressions analyzing the relation between the interactions of AMLT with holdings-based information proxies and monthly fund performance over the subsequent quarter. The dependent variable is monthly abnormal fund return, Carhart-α. The main independent variables are the interactions between *AMLT* and holdings-based information proxies. *High info* is an indicator equal to one if the holding-size-weighted average number of firm-level signals is in the highest cross-sectional tercile. All regressions include the same control variables as in Table 6, as well as time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A – AMLT and Active Portfolio Tilt

Dependent Variables:	ATILT_N.Signals$_t$	ATILT_N.Signals.quant$_t$	ATILT_N.Filings$_t$	ATILT_N.Articles$_t$
Model:	(1)	(2)	(3)	(4)
AMLT$_t$	0.204***	0.217***	0.065***	0.222***
	(0.009)	(0.012)	(0.009)	(0.011)
Exp ratio$_t$	-0.099***	-0.050**	-0.149***	-0.117***
	(0.015)	(0.021)	(0.015)	(0.018)
Log(TNA)$_t$	-0.003	0.000	-0.003	0.020***
	(0.004)	(0.005)	(0.004)	(0.004)
Turnover$_t$	0.046***	-0.029*	0.143***	0.139***
	(0.013)	(0.017)	(0.013)	(0.016)
Fund age$_t$	0.000	-0.001	0.000	-0.001
	(0.000)	(0.001)	(0.000)	(0.000)
Ret$_t$	-0.005	0.000	0.003	-0.007
	(0.004)	(0.005)	(0.004)	(0.005)
Style fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	140,577	140,577	140,577	140,577
Adjusted R^2	0.117	0.097	0.087	0.122

Panel B – Performance Regression

Dependent Variables:	Carhart-α_t			
Information proxy:	N.Signals	N.Signals.quant	N.Filings	N.Articles
Model:	(1)	(2)	(3)	(4)
AMLT_{t-1}	0.042***	0.040***	0.046***	0.040***
	(0.007)	(0.007)	(0.007)	(0.007)
High info_{t-1}	0.012**	0.005	-0.019***	0.049***
	(0.006)	(0.006)	(0.007)	(0.007)
AMLT_{t-1} × High info_{t-1}	0.023**	0.028**	0.014	0.047***
	(0.011)	(0.012)	(0.014)	(0.013)
Ret_{t-1}	-0.004	-0.004	-0.005	-0.004
	(0.008)	(0.008)	(0.008)	(0.008)
Exp ratio_{t-1}	-1.061***	-1.072***	-1.093***	-1.047***
	(0.110)	(0.109)	(0.111)	(0.109)
Log(TNA)_{t-1}	-0.003*	-0.003*	-0.004**	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)
Turnover_{t-1}	-0.058***	-0.058***	-0.058***	-0.057***
	(0.007)	(0.007)	(0.007)	(0.007)
Fund age_{t-1}	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Style fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	371,727	371,727	371,727	371,727
Adjusted R^2	0.066	0.066	0.066	0.066

Table 9: **AMLT and Future Fund Performance: Linear ML Models and Quantitative vs. Textual Information Inputs**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between AMLT and monthly fund performance over the subsequent quarter. The dependent variable is monthly abnormal fund return, Carhart-α. Panel A considers AMLT measures constructed from machine-learning signals generated by alternative models. Columns (1), (2), and (3) report results using ML signals derived from a three-layer DNN, an elastic net, and a principal components regression model based on the full set of predictors, respectively. Panel B considers AMLT measures constructed from machine-learning signals based on alternative information sets. Columns (1), (2), and (3) report results using AMLT constructed from machine-learning signals generated by DNN models using all predictors, quantitative data only, and textual data only, respectively. All regressions include the same control variables as in Table 6, as well as time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A – Alternative ML Models

Dependent Variables:	Carhart-α_t		
ML model:	DNN	ENet	PCR
Model:	(1)	(2)	(3)
AMLT$_{t-1}$	0.054***	0.025***	0.025***
	(0.006)	(0.005)	(0.006)
Controls	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes
Observations	406,897	406,897	406,897
Adjusted R^2	0.070	0.070	0.070

Panel B – Quantitative vs. Textual Information Inputs

Dependent Variables:	Carhart-α_t		
Information input:	All	Quantitative	Textual
Model:	(1)	(2)	(3)
AMLT$_{t-1}$	0.054***	0.028***	0.017***
	(0.006)	(0.005)	(0.006)
Controls	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes
Observations	406,897	406,897	406,897
Adjusted R^2	0.070	0.070	0.069

Table 10: **AMLT and Future Fund Performance: Market Conditions**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between the interactions of AMLT with market conditions and monthly fund performance over the subsequent quarter. The dependent variable is monthly abnormal fund return, Carhart-α, calculated as the difference between the fund return and the expected return based on factor loadings estimated using the Carhart four-factor model over the past 36 months in a rolling window. The main independent variables are the interactions between *AMLT* and market condition proxies. *AMLT* is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. *High* is an indicator equal to one if the market proxy in a given month exceeds its full-sample median. Column (1) conditions on investor sentiment from Baker and Wurgler (2007) (*Sentiment*). Columns (2) and (3) condition on monthly *VIX*, the S&P 500 option-implied volatility index, and *Market Illiquidity*, defined as the equally weighted average of stock-level Amihud (2002) illiquidity across NYSE/AMEX stocks in a given month (Avramov et al., 2016), respectively. All regressions include the same control variables as in Table 6, as well as time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variables:	Carhart-α_t		
Market states:	Sentiment	VIX	Market Illiquidity
Model:	(1)	(2)	(3)
AMLT_{t-1}	0.037***	0.017***	0.031***
	(0.008)	(0.007)	(0.007)
$\text{AMLT}_{t-1} \times \text{High}_{t-1}$	0.033***	0.069***	0.045***
	(0.011)	(0.012)	(0.011)
Ret_{t-1}	0.016**	0.015**	0.015**
	(0.007)	(0.007)	(0.007)
Exp ratio_{t-1}	-1.082***	-1.085***	-1.080***
	(0.105)	(0.105)	(0.105)
Log(TNA)_{t-1}	-0.005***	-0.004***	-0.004***
	(0.002)	(0.002)	(0.002)
Turnover_{t-1}	-0.056***	-0.055***	-0.056***
	(0.007)	(0.007)	(0.007)
Fund age_{t-1}	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Time fixed effects	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes
Observations	406,897	406,897	406,897
Adjusted R^2	0.070	0.070	0.070

Table 11: **AMLT and Future Fund Performance: Fund Styles**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between AMLT and monthly fund performance over the subsequent quarter, estimated separately for funds with different styles. The dependent variable is monthly abnormal fund return, Carhart-α, calculated as the difference between the fund return and the expected return based on factor loadings estimated using the Carhart four-factor model over the past 36 months in a rolling window. The main independent variable, *AMLT*, is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter.. At the beginning of each quarter, funds are assigned to one of four styles. Funds are first sorted into two groups based on a size score and, within each size group, into two groups based on a value score. The size (value) score corresponds to the average quintile rank of portfolio holdings' market capitalization (book-to-market ratio) using NYSE breakpoints. All regressions include the same control variables as in Table 6, as well as time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variables:	Carhart-α_t			
Sample:	Small Value	Small Growth	Large Value	Large Growth
Model:	(1)	(2)	(3)	(4)
AMLT$_{t-1}$	0.038***	0.069***	0.046***	0.055***
	(0.010)	(0.010)	(0.010)	(0.012)
Ret$_{t-1}$	0.019	0.034**	0.006	-0.015
	(0.018)	(0.014)	(0.015)	(0.014)
Exp ratio$_{t-1}$	-0.886***	-0.639***	-1.429***	-1.416***
	(0.238)	(0.212)	(0.168)	(0.191)
Log(TNA)$_{t-1}$	-0.008**	-0.008*	-0.005**	-0.002
	(0.004)	(0.004)	(0.002)	(0.003)
Turnover$_{t-1}$	-0.034**	-0.085***	-0.037***	-0.032***
	(0.015)	(0.012)	(0.010)	(0.012)
Fund age$_{t-1}$	0.001	0.000	0.000	0.000
	(0.001)	(0.001)	(0.000)	(0.000)
Time fixed effects	Yes	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes	Yes
Observations	100,508	102,282	101,937	102,170
Adjusted R^2	0.129	0.201	0.114	0.084

Appendix
Active Machine-Learning-Based Trading and Mutual Fund Performance

A.1 The Evolution of AI and ML Technologies and their Adoption in Asset Management

Artificial intelligence (AI) broadly refers to computational systems designed to replicate human cognitive functions such as pattern recognition, decision-making, and problem-solving (Russell and Norvig, 2016). Within AI, machine learning (ML) is a subset that focuses on algorithms that learn from data and improve predictive performance without explicit programming (Mitchell, 1997).

ML methods can be categorized based on their complexity and ability to model relationships in data. Regularized linear models, such as ridge regression, LASSO regression, and elastic net, provide robust predictive modeling by penalizing large coefficients and selecting relevant features. Beyond linear methods, decision trees, and ensemble models like random forests allow for nonlinearities. Further advancements, such as gradient boosting techniques refine predictive accuracy by iteratively improving weak models. At a more advanced level, neural networks and deep learning methods enable the capture of highly complex, nonlinear relationships.

Early adoption of ML in asset management can be traced back to the 1990s, as computational advancements enabled firms to develop more sophisticated trading and risk management strategies. One of the earliest and most successful adopters was Renaissance Technologies, which established its flagship Medallion Fund in 1988. The firm uses computer-based models, such as pattern recognition, analyzing as much data as can be gathered to identify market patterns. In the mid-1990s, Goldman Sachs launched the Global Alpha Fund, a quantitative hedge fund that employed computer-driven models to evaluate risks and identify investment opportunities across various asset classes. These early implementations of ML and AI laid the groundwork for the widespread adoption of advanced computational techniques in asset management, significantly influencing modern investment strategies.

A major breakthrough occurred in the early 2000s with the increasing adoption of neural networks in financial modeling. Originally conceptualized by McCulloch and Pitts (1943), neural networks gained practical relevance following the development of the backpropagation algorithm (Rumelhart et al., 1986), which enabled efficient training of multi-layer networks. LeBaron (1999) demonstrated the viability of neural networks in forecasting asset returns, further solidifying their importance in quantitative finance. With the increasing availability of financial data and improvements in computational power, neural networks emerged as a feasible and scalable solution for asset managers in the early 2000s.

The 2010s marked a paradigm shift in AI adoption within asset management, driven by the explosion of big data and advancements in deep learning. Asset management firms increasingly turned to natural language processing for sentiment analysis, reinforcement learning for dynamic portfolio optimization, and deep neural networks for anomaly detection. Leading asset management firms including both hedge funds and mutual funds, such as BlackRock, Two Sigma, AQR, Vanguard Group, State Street Global Advisors, and Cerebellum Capital, leveraged these AI-driven models to extract predictive signals from diverse data sources, including macroeconomic reports, alternative datasets, and textual sentiment analysis.[24] The integration of deep learning in asset management has enabled more sophisticated investment strategies, reinforcing the industry's reliance on AI.

Despite indications of AI adoption in asset management from news, online articles, and fund websites, the extent to which the industry integrates ML into trading strategies remains largely unknown, as asset managers may not have incentives to disclose their underlying trading strategies to public or competitors. This is especially true for mutual funds, who already operate under greater scrutiny and face stricter disclosure requirements.

A.2 Machine Learning Model

One of the key components of the proposed AMLT measure, defined in Equation 1, is the ML-based trading signal. In this Appendix section, we establish our ML return prediction

[24] https://hedgefundalpha.com/ai-hedge-fund-returns
https://www.wired.com/2016/01/the-rise-of-the-artificially-intelligent-hedge-fund
https://www.statestreet.com/alpha/insights/artificial-intelligence
https://ir.vcm.com/news/news-details/2016/Victory-Capital-Announces-Investment-in-Cerebellum-Capital
https://www.bloomberg.com/news/articles/2024-02-06/vanguard-quietly-embraces-ai-in-13-billion-of-quant-stock-funds
https://loanworks.ai/artificial-intelligence-in-financial-services-a-historical-review

procedure and the profitability of the resulting ML trading signals.

Machine Learning Based Trading Signals

It is paramount that the ML-based trading signal is out-of-sample, not suffering from any forward looking bias, so that a hypothetical mutual fund could have traded on it in real time. We follow the ML procedures outlined in Gu et al. (2020), describing a stock's expected return with a flexible prediction function:

$$E_t(r_{i,t+1}) = \hat{g}(z_{i,t}; \theta) \tag{A.1}$$

where $E_t(r_{i,t+1})$ represents the expected excess return $r_{i,t+1}$ for stock i during the subsequent quarter $t+1$, vector $z_{i,t}$ is the full set of predictors, and θ denotes the parameter set of the prediction model.

We select a traditional feed-forward neural network architecture as our prediction model $g(\cdot)$. This structure consists of an input layer comprising raw predictors, one or more hidden layers that capture interactions and apply nonlinear transformations, and an output layer that synthesizes information from the hidden layers to generate predictions.

We select a three-layer architecture (NN3) consisting of 32, 16, and 8 neurons in successive hidden layers. Across all layers, we employ the rectified linear unit (ReLU) activation function, a widely used choice that promotes sparsity and facilitates efficient gradient-based optimization. The dependent variable is rank-transformed and scaled to range from 0 to 1. Following Green et al. (2017), predictors are normalized in the cross-section.

We also employ an ensemble approach to training our neural networks, following the methodologies of Hansen and Salamon (1990) and Dietterich (2000). Specifically, we initialize the neural network estimation with multiple random seeds and aggregate predictions by averaging forecasts across ten trained networks. This technique mitigates prediction variance by reducing sensitivity to the stochastic nature of the optimization process, which can lead to variations in forecasts depending on the initialization.

To optimize model performance, we tune hyperparameters through grid search, utilizing a 10-year training window and a 4-year validation window. Our out-of-sample predictions commence in 2002. Given the substantial computational cost of machine learning models, we do not re-estimate the network each month. Instead, we re-tune hyperparameters annually, rolling the training and validation windows forward by one year while maintaining a consistent sample length. However, we re-estimate the model using the selected hyper-parameters for every monthly prediction. This approach balances computational efficiency with adaptability, ensuring that the model remains responsive to evolving market conditions.

Taken together, at the end of each month t and using the estimated neural network $\hat{g}(\cdot)$, we generate the expected return for each stock i during the subsequent quarter.

Machine Learning Performance

Next, we validate the ML-based trading signal by documenting that a portfolio strategy based on it creates alpha. At the end of each quarter, we sort stocks into deciles based on the ML trading signal and compute the monthly portfolio returns over the subsequent three months. The portfolios are rebalanced quarterly, and stocks within each decile are assigned equal weights. Finally, we then construct a zero-net-investment portfolio that takes a long position in the stocks within the highest decile (decile 10) and a short position in those within the lowest decile (decile 1).

Table A.1 presents the out-of-sample performance results for the ML portfolios. On average, each portfolio consists of 463 stocks. The bottom decile portfolios exhibit either negative monthly returns or returns that are not significantly different from zero, whereas the top decile portfolios generate significantly positive average monthly returns. The top long portfolio (decile 10) achieves an average monthly raw return of 1.2% (equivalent to 15.0% annualized) with an average monthly volatility of 4.5% (annualized at 15.6%). Meanwhile, the long-short portfolio, which goes long in the top decile and short in the bottom decile, earns an average monthly raw return of 2.1% (25.3% annualized) with an average monthly volatility of 7.7% (26.7% annualized). The magnitudes are highly consistent with Gu et al. (2020)[25].

We estimate the risk-adjusted returns on the portfolios as intercept alphas from time-series regressions, using the capital asset pricing model (CAPM) with the market factor, Fama and French's (1993) three-factor model and Carhart's (1997) four-factor model, which augments the Fama-French factors with Jegadeesh and Titman's (1993) momentum factor and the five-factor model, which adds Pástor and Stambaugh's (2003) liquidity risk factor to the four-factor model. For instance, the Carhart four-factor alpha is the intercept from

[25]See Gu et al. (2020) Table 7, performance of the machine learning portfolios for NN3.

the following time-series regression:

$$R_{p,t} - R_{f,t} = \alpha_p + \beta_{p,m}(R_{m,t} - R_{f,t}) + \beta_{p,SMB} SMB_t + \beta_{p,HML} HML_t + \beta_{p,UMD} UMD_t + \varepsilon_{p,t} \quad \text{(A.2)}$$

where $R_{p,t}$ is the return in month t for fund portfolio p, $R_{f,t}$ is the 1-month Treasury-bill rate in month t, $R_{m,t}$ is the value-weighted stock market return in month t, SMB_t is the difference in returns between small and large capitalization stocks in month t, HML_t is the return difference between high and low book-to-market stocks in month t, and UMD_t is the return difference between stocks with high and low past returns in month t.

Risk-adjusted performance measures further confirm the economic significance of the ML portfolios. For all models, risk-adjusted monthly returns are all positive and highly significant for both the top long portfolios and the long-short portfolio. For instance, the long-short portfolio delivers a monthly CAPM alpha of 2.6%, translating to an annualized alpha of 31.3%, underscoring the ability of ML-based signals to generate abnormal returns beyond traditional risk factors.

Table A.1: **Performance of the Machine Learning Portfolios**

This table reports the performance of stock portfolios sorted on the ML signal. Stocks are sorted into deciles at the end of each quarter based on the ML signal for the subsequent quarter. We report equally weighted monthly portfolio returns for each decile and the high-minus-low portfolio, both raw and risk-adjusted using the CAPM (CAPM-α), the Fama and French (1993) three-factor model (FF-α), the Carhart (1997) four-factor model (Carhart-α), and the Pástor and Stambaugh (2003) five-factor model (PS-α). The table also reports the standard deviation of monthly returns (*SD*), Sharpe ratios (*SR*), the most negative monthly return (*Max 1M Loss*), maximum drawdowns (*Max DD*), the average number of firms per portfolio (*Nr firms*), and the average quarterly turnover (*Turnover*). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Portfolio	Raw	CAPM-α	FF-α	Carhart-α	PS-α	SD	SR	Max 1M loss	Max DD	Nr firms	Turnover
1	-0.859	-1.946***	-1.994***	-1.712***	-1.721***	10.052	-0.296	-26.958	96.849	463	87.988
	(0.779)	(0.479)	(0.403)	(0.349)	(0.351)						
2	0.476	-0.528*	-0.564**	-0.363*	-0.372*	8.123	0.203	-28.210	73.932	463	115.808
	(0.621)	(0.305)	(0.241)	(0.209)	(0.208)						
3	0.635	-0.267	-0.297*	-0.137	-0.144	6.959	0.316	-25.247	65.797	463	124.194
	(0.519)	(0.227)	(0.168)	(0.150)	(0.147)						
4	0.720	-0.121	-0.145	-0.037	-0.045	6.279	0.397	-25.629	64.189	463	127.097
	(0.452)	(0.188)	(0.124)	(0.112)	(0.106)						
5	0.812**	0.035	0.014	0.085	0.081	5.723	0.491	-24.817	61.615	463	128.834
	(0.405)	(0.156)	(0.108)	(0.104)	(0.101)						
6	0.923**	0.192	0.173*	0.209**	0.205**	5.347	0.598	-22.510	58.543	463	129.691
	(0.376)	(0.153)	(0.094)	(0.095)	(0.093)						
7	0.923**	0.198	0.181*	0.200**	0.196**	5.279	0.606	-23.310	59.707	463	129.459
	(0.371)	(0.155)	(0.093)	(0.093)	(0.091)						
8	0.962***	0.274*	0.259***	0.258***	0.256***	4.997	0.667	-22.045	53.154	463	127.592
	(0.340)	(0.147)	(0.082)	(0.080)	(0.079)						
9	0.970***	0.311**	0.296***	0.283***	0.280***	4.787	0.702	-23.627	52.011	463	121.561
	(0.321)	(0.142)	(0.092)	(0.089)	(0.088)						
10	1.249***	0.661***	0.648***	0.601***	0.599***	4.503	0.961	-23.838	42.636	464	98.312
	(0.301)	(0.182)	(0.146)	(0.139)	(0.140)						
H-L	2.108***	2.608***	2.643***	2.313***	2.320***	7.726	0.945	-38.027	67.362		
	(0.607)	(0.499)	(0.453)	(0.382)	(0.385)						

Table A.2: **AMLT and Future Fund Performance: Multivariate Regressions – Beginning-Quarter Information**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between *AMLT_begQInfo* and monthly fund performance over the subsequent quarter. The dependent variable is monthly abnormal fund return, Carhart-α, calculated as the difference between the fund return and the expected return based on factor loadings estimated using the Carhart four-factor model over the past 36 months in a rolling window. The main independent variable, *AMLT_begQInfo*, is a fund-level signed active share measure. Similar to the AMLT measure, *AMLT_begQInfo* in quarter t is computed using an ML-based trading signal generated from information available at quarter $t-1$, that is, the beginning of quarter t. Carhart-α is estimated for both net and gross fund returns. Control variables include lagged fund performance, *Ret*, measured as the average fund return over the previous 12 months; expense ratio, *Exp ratio*; fund size, *Log(TNA)*, measured as the logarithm of TNA; turnover ratio, *Turnover*; and fund age, *Fund age*, measured as the natural logarithm of one plus fund age. All control variables are measured over the prior year. All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable:	Carhart-α_t	
	Net return	Gross return
Model:	(1)	(2)
AMLT_begQInfo$_{t-1}$	0.044***	0.043***
	(0.005)	(0.005)
Ret$_{t-1}$	0.016**	0.016**
	(0.007)	(0.007)
Exp ratio$_{t-1}$	-1.086***	-0.131
	(0.104)	(0.104)
Log(TNA)$_{t-1}$	-0.005***	-0.005***
	(0.002)	(0.002)
Turnover$_{t-1}$	-0.056***	-0.056***
	(0.007)	(0.007)
Fund age$_{t-1}$	0.000	0.000
	(0.000)	(0.000)
Time fixed effects	Yes	Yes
Style fixed effects	Yes	Yes
Observations	406,897	406,462
Adjusted R^2	0.070	0.069

Table A.3: **Transition Matrix**

This table reports transition probabilities for funds sorted into AMLT quintiles over one-, two-, and four-quarter horizons in Panels A, B, and C, respectively. Mutual funds are sorted into AMLT quintiles at quarter t, and the table reports the probability that a fund is in each quintile at $t+1$, $t+2$, and $t+4$. Columns (2) through (6) report transition probabilities, Column (7) reports attrition rates, and Column (8) reports the difference between diagonal and average off-diagonal transition probabilities. The last row reports the difference in attrition between quintiles five and one. All probabilities are expressed in percent. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A – One Quarter Ahead

	Portfolio$_{t+1}$							
Portfolio$_t$	1	2	3	4	5	Attrition	Δ Off-Diag	
1	66.2	20.2	6.5	4.0	3.1	1.9	56.7***	(0.021)
2	20.3	41.7	21.4	11.1	5.5	2.0	26.7***	(0.020)
3	6.6	21.9	41.0	22.3	8.2	2.0	25.9***	(0.019)
4	4.1	10.8	23.0	41.1	21.1	1.7	26.0***	(0.020)
5	3.3	5.4	8.1	21.4	62.0	1.7	51.7***	(0.021)
						-0.3***		
						(0.001)		

Panel B – Two Quarter Ahead

	Portfolio$_{t+2}$							
Portfolio$_t$	1	2	3	4	5	Attrition	Δ Off-Diag	
1	57.9	21.5	8.7	6.4	5.5	3.4	45.7***	(0.025)
2	21.7	34.7	21.4	13.6	8.5	3.3	17.8***	(0.020)
3	9.0	22.0	35.9	22.5	10.7	3.0	19.3***	(0.019)
4	6.3	13.5	23.0	34.8	22.3	3.1	17.9***	(0.019)
5	5.5	8.4	10.8	22.6	52.8	2.9	39.9***	(0.024)
						-0.7***		
						(0.002)		

Panel C – One Year Ahead

	Portfolio$_{t+4}$							
Portfolio$_t$	1	2	3	4	5	Attrition	Δ Off-Diag	
1	50.6	22.1	11.0	8.7	7.7	6.8	35.6***	(0.021)
2	22.8	30.9	20.1	15.3	11.0	6.1	12.7***	(0.014)
3	11.2	20.6	32.6	22.8	12.7	5.8	14.9***	(0.019)
4	8.3	15.2	23.3	30.8	22.5	5.5	12.8***	(0.013)
5	7.8	11.4	12.9	22.1	45.8	5.2	30.5***	(0.023)
						-1.6***		
						(0.004)		

Table A.4: **AMLT and Future Fund Performance: Portfolio Approach – Alternative AMLT**

This table reports abnormal monthly returns of decile mutual fund portfolios sorted by the most recent AMLT measure under two alternative AMLT constructions. Alternative 1 applies a granular rank-based transformation of the raw ML predictions into 100 portfolios. Alternative 2 applies a rank-based transformation that places greater weight on extreme ML signals: stocks are assigned scores from -3 to +3, with bottom (top) decile stocks receiving -3 (+3) and middle-decile stocks assigned zero. *AMLT_long* denotes the long-only version of each alternative AMLT construction, assigning positive scores to stocks in the corresponding top portfolios and zero to all other stocks. Decile portfolios are formed and rebalanced at the end of each quarter, and portfolio performance is computed over the subsequent three months for each decile and the high-minus-low portfolio. The table reports risk-adjusted returns based on the CAPM, the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, and the Pástor and Stambaugh (2003) five-factor model for portfolios sorted on *AMLT* and *AMLT_long*. Returns are expressed in percent per month. Standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

	AMLT_Alt1				**AMLT_long_Alt1**			
	CAPM-α	FF-α	Carhart-α	PS-α	CAPM-α	FF-α	Carhart-α	PS-α
1	-0.121	-0.134**	-0.109*	-0.114*	-0.114	-0.125**	-0.103*	-0.108*
	(0.089)	(0.066)	(0.065)	(0.063)	(0.080)	(0.062)	(0.061)	(0.060)
10	0.117*	0.106***	0.089**	0.088**	0.120**	0.112***	0.097**	0.095**
	(0.067)	(0.041)	(0.040)	(0.040)	(0.060)	(0.041)	(0.041)	(0.041)
H-L	0.237***	0.240***	0.198**	0.202***	0.235***	0.237***	0.200***	0.203***
	(0.081)	(0.080)	(0.077)	(0.076)	(0.078)	(0.077)	(0.075)	(0.074)
	AMLT_Alt2				**AMLT_long_Alt2**			
1	-0.126	-0.138**	-0.116*	-0.120*	-0.113	-0.122**	-0.104*	-0.108*
	(0.086)	(0.066)	(0.065)	(0.064)	(0.076)	(0.062)	(0.061)	(0.060)
10	0.114*	0.104**	0.085**	0.084**	0.127**	0.119***	0.105**	0.103**
	(0.065)	(0.043)	(0.042)	(0.041)	(0.057)	(0.042)	(0.042)	(0.042)
H-L	0.240***	0.242***	0.201***	0.204***	0.239***	0.242***	0.208***	0.212***
	(0.081)	(0.080)	(0.077)	(0.077)	(0.077)	(0.076)	(0.075)	(0.074)

Table A.5: **AMLT and Future Fund Performance: Long-Horizon Performance**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between AMLT and fund performance over longer horizons. The dependent variables are monthly abnormal fund returns, Carhart-α_t, Carhart-α_{t+4}, and Carhart-α_{t+8}, corresponding to fund performance over the subsequent quarter, subsequent year, and subsequent two years, respectively. Carhart-α is calculated as the difference between the fund return and the expected return based on factor loadings estimated using the Carhart four-factor model over the past 36 months in a rolling window. The main independent variable, *AMLT*, is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. Control variables include lagged fund performance, *Ret*, measured as the average fund return over the previous 12 months; expense ratio, *Exp ratio*; fund size, *Log(TNA)*, measured as the logarithm of TNA; turnover ratio, *Turnover*; and fund age, *Fund age*, measured as the natural logarithm of one plus fund age. All control variables are measured over the prior year. All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variables:	Carhart-α_t	Carhart-α_{t+4}	Carhart-α_{t+8}
Model:	(1)	(2)	(3)
AMLT$_{t-1}$	0.054***	0.041***	0.033***
	(0.006)	(0.006)	(0.006)
Ret$_{t-1}$	0.016**	0.037***	-0.006
	(0.007)	(0.008)	(0.010)
Exp ratio$_{t-1}$	-1.080***	-1.041***	-1.092***
	(0.105)	(0.092)	(0.111)
Log(TNA)$_{t-1}$	-0.005***	-0.004**	-0.002
	(0.002)	(0.002)	(0.002)
Turnover$_{t-1}$	-0.056***	-0.034***	-0.038***
	(0.007)	(0.007)	(0.007)
Fund age$_{t-1}$	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Time fixed effects	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes
Observations	406,897	376,795	340,076
Adjusted R^2	0.070	0.071	0.073

Table A.6: **AMLT and Future Fund Performance: Random Forest and Early Periods**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between AMLT and monthly fund performance over the subsequent quarter. The dependent variable is monthly abnormal fund return, Carhart-α, calculated as the difference between the fund return and the expected return based on factor loadings estimated using the Carhart four-factor model over the past 36 months in a rolling window. The main independent variables are *AMLT* and *AMLT_RandomF*. AMLT is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. *AMLT_RandomF* is constructed analogously using ML signals generated by a Random Forest model. Columns (1) and (4), (2) and (5), and (3) and (6) report results for the full sample, the subsample ending in 2010, and the subsample starting in 2011, respectively. Control variables include lagged fund performance, *Ret*, measured as the average fund return over the previous 12 months; expense ratio, *Exp ratio*; fund size, *Log(TNA)*, measured as the logarithm of TNA; turnover ratio, *Turnover*; and fund age, *Fund age*, measured as the natural logarithm of one plus fund age. All control variables are measured over the prior year. All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variables:	Carhart-α_t					
Sample period:	Full	$\leq$ 2010	> 2010	Full	$\leq$ 2010	> 2010
Model:	(1)	(2)	(3)	(4)	(5)	(6)
AMLT_{t-1}	0.054***	0.045***	0.067***			
	(0.006)	(0.008)	(0.008)			
$\text{AMLT\_RandomF}_{t-1}$				0.032***	0.026***	0.031***
				(0.005)	(0.009)	(0.006)
Ret_{t-1}	0.016**	-0.021*	0.040***	0.014**	-0.021*	0.037***
	(0.007)	(0.012)	(0.008)	(0.007)	(0.012)	(0.008)
Exp ratio_{t-1}	-1.080***	-0.949***	-1.291***	-1.152***	-1.036***	-1.345***
	(0.105)	(0.169)	(0.116)	(0.104)	(0.167)	(0.116)
Log(TNA)_{t-1}	-0.005***	-0.006**	-0.006***	-0.005***	-0.007**	-0.007***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
Turnover_{t-1}	-0.056***	-0.044***	-0.076***	-0.062***	-0.050***	-0.080***
	(0.007)	(0.010)	(0.009)	(0.007)	(0.010)	(0.009)
Fund age_{t-1}	0.000	0.001*	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Time fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	406,897	175,621	231,276	406,897	175,621	231,276
Adjusted R^2	0.070	0.049	0.096	0.070	0.049	0.096

Table A.7: **Employee AI Talent and Future Fund Performance – Multivariate Regressions**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between AMLT, employee AI talent, and monthly fund performance over the subsequent quarter. The dependent variable is monthly abnormal fund return, Carhart-α, calculated as the difference between the fund return and the expected return based on factor loadings estimated using the Carhart four-factor model over the past 36 months in a rolling window. The main independent variables are *AMLT* and *Share_CoreAI*. AMLT is a fund-level signed active share measure defined in Equation (1), computed quarterly as the sum of the products of the active deviations of fund portfolio weights from their benchmark weights during the quarter and the forward-looking ML trading signal indicators for the subsequent quarter. *Share_CoreAI* is the share of employees of a registered investment adviser with AI talent. Control variables include lagged fund performance, *Ret*, measured as the average fund return over the previous 12 months; expense ratio, *Exp ratio*; fund size, *Log(TNA)*, measured as the logarithm of TNA; turnover ratio, *Turnover*; and fund age, *Fund age*, measured as the natural logarithm of one plus fund age. All control variables are measured over the prior year. All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable:	Carhart-α_t	
Model:	(1)	(2)
AMLT_{t-1}		0.068***
		(0.010)
$\text{Share\_CoreAI}_{t-1}$	0.632*	0.547
	(0.353)	(0.353)
Ret_{t-1}	0.017	0.014
	(0.011)	(0.011)
Exp ratio_{t-1}	-1.789***	-1.662***
	(0.239)	(0.242)
Log(TNA)_{t-1}	-0.008**	-0.007**
	(0.003)	(0.003)
Turnover_{t-1}	-0.047***	-0.041***
	(0.012)	(0.012)
Fund age_{t-1}	0.000	0.000
	(0.000)	(0.000)
Time fixed effects	Yes	Yes
Style fixed effects	Yes	Yes
Observations	118,885	118,885
Adjusted R^2	0.069	0.070

Table A.8: **Turnover-Adjusted ML Trading Measure and Future Fund Performance: Multivariate Regressions**

This table reports coefficient estimates and standard errors from monthly regressions analyzing the relation between *TOMLT* and monthly fund performance over the subsequent quarter. The dependent variable is monthly abnormal fund return, Carhart-α, calculated as the difference between the fund return and the expected return based on factor loadings estimated using the Carhart four-factor model over the past 36 months in a rolling window, and estimated for both net and gross fund returns. The main independent variable, *TOMLT*, is defined as a fund's dollar trading during the quarter that follows ML signals, scaled by the fund's total dollar trading during the quarter. Formally,

$$TOMLT_{f,t} = \frac{MLBuy_{f,t} - MLSell_{f,t}}{TotTrade_{f,t}} \tag{A.3}$$

where $MLBuy_{f,t}$ $(MLSell_{f,t}) = \sum_{i=1}^{n} \left(N_{i,t}^{f} - N_{i,t-1}^{f} \right) \times Prc_{i,t-1}$ for stocks i in the top (bottom) decile of the ML signal, with changes in shares adjusted for stock splits, and $TotTrade_{f,t} = \sum_{i=1}^{n} | \left(N_{i,t}^{f} - N_{i,t-1}^{f} \right) \times Prc_{i,t-1}|$. Control variables include lagged fund performance, *Ret*, measured as the average fund return over the previous 12 months; expense ratio, *Exp ratio*; fund size, *Log(TNA)*, measured as the logarithm of TNA; turnover ratio, *Turnover*; and fund age, *Fund age*, measured as the natural logarithm of one plus fund age, all measured over the prior year. All regressions include time and style fixed effects. Standard errors are clustered by fund and reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Dependent Variable:	Carhart-α_t	
	Net return	Gross return
Model:	(1)	(2)
TOMLT_{t-1}	0.063***	0.061***
	(0.021)	(0.021)
Ret_{t-1}	0.014*	0.013*
	(0.007)	(0.007)
Exp ratio_{t-1}	-1.191***	-0.236**
	(0.109)	(0.109)
Log(TNA)_{t-1}	-0.005***	-0.005***
	(0.002)	(0.002)
Turnover_{t-1}	-0.060***	-0.060***
	(0.007)	(0.007)
Fund age_{t-1}	0.000	0.000
	(0.000)	(0.000)
Time fixed effects	Yes	Yes
Style fixed effects	Yes	Yes
Observations	377,330	376,937
Adjusted R^2	0.072	0.071

Internet Appendix
Active Machine-Learning-Based Trading and Mutual Fund Performance

Table IA.1: **Variable Definition and Sources**

This table lists all explanatory variables used in the machine-learning model, grouped by category, along with their definitions and data start dates.

Acronym	Definition	Start Date
Financial variables, following Green et al. (2017)		
absacc	Absolute value of *acc*	1980-01
acc	Annual income before extraordinary items (*ib*) minus operating cash flows (*oancf*) divided by average total assets (*at*); if *oancf* is missing then set to change in *act* - change in *che* - change in *lct* + change in *dlc* + change in *txp* - *dp*	1980-01
aeavol	Average daily trading volume (*vol*) for 3 days around earnings announcement minus average daily volume for 1-month ending 2 weeks before earnings announcement divided by 1-month average daily volume. Earnings announcement day from Compustat quarterly (*rdq*)	1980-01
age	Number of years since first Compustat coverage	1980-01
agr	Annual percent change in total assets (*at*)	1980-01
baspread	Monthly average of daily bid-ask spread divided by average of daily spread	1980-01
beta	Estimated market beta from weekly returns and equal weighted market returns for 3 years ending month t-1 with at least 52 weeks of returns	1980-01
bm	Book value of equity (*ceq*) divided by end of fiscal year-end market capitalization	1980-01
bm_ia	Industry adjusted book-to-market ratio	1980-01
cash	Cash and cash equivalents divided by average total assets	1980-01
cashdebt	Earnings before depreciation and extraordinary items (*ib* + *dp*) divided by avg. total liabilities (*lt*)	1980-01
cashpr	Fiscal year-end market capitalization plus long-term debt (*dltt*) minus total assets (at) divided by cash and equivalents (*che*)	1980-01
cfp	Operating cash flows divided by fiscal-year-end market capitalization	1980-01
cfp_ia	Industry adjusted *cfp*	1980-01
chatoia	2-digit SIC fiscal year mean adjusted change in sales (*sale*) divided by average total assets (*at*)	1980-06
chcsho	Annual percent change in shares outstanding (*csho*)	1980-01
chempia	Industry-adjusted change in number of employees	1980-01
chfeps	Mean analyst forecast in month prior to fiscal period end date from I/B/E/S summary file minus same mean forecast for prior fiscal period using annual earnings forecasts	1989-01
chinv	Change in inventory (*inv*) scaled by average total assets (*at*)	1980-01
chmom	Cumulative returns from months t-6 to t-1 minus months t-12 to t-7	1980-01
chnanalyst	Change in analyst from month t-3 to month t	1989-04
chpmia	2-digit SIC fiscal year mean adjusted change in income before extraordinary items (*ib*) divided by sales (sale)	1980-01

Table IA.1 continued

Acronym	Definition	Start Date
chtx	Percent change in total taxes (*txtq*) from quarter t-4 to t	1980-01
cinvest	Change over one quarter in net PP&E	1980-01
convind	An indicator equal to 1 if company has convertible debt obligations	1980-01
currat	Current assets / current liabilities	1980-01
depr	Depreciation divided by PP&E	1980-01
disp	Standard deviation of analyst forecasts in month prior to fiscal period end date divided by the absolute value of the mean forecast; if $meanest = 0$, then scalar set to 1. Forecast data from I/B/E/S summary files	1989-01
divi	An indicator variable equal to 1 if company pays dividends but did not in prior year	1980-01
divo	An indicator variable equal to 1 if company does not pay dividend but did in prior year	1980-01
dy	Total dividends (*dvt*) divided by market capitalization at fiscal year-end	1980-01
ear	Sum of daily returns in three days around earnings announcement. Earnings announcement from Compustat quarterly file (*rdq*)	1980-01
egr	Annual percent change in book value of equity (*ceq*)	1980-01
ep	Annual income before extraordinary items (*ib*) divided by end of fiscal year market cap	1980-01
fgr5yr	Most recently available analyst forecasted 5-year growth	1989-01
gma	Revenues (*revt*) minus cost of goods sold (*cogs*) divided by lagged total assets (*at*)	1980-01
grCAPX	Percent change in capital expenditures from year t-2 to year t	1980-06
grltnoa	Growth in long-term net operating assets	1980-01
herf	2-digit SIC fiscal year sales concentration (sum of squared percent of sales in industry for each company)	1980-01
hire	Percent change in number of employees (*emp*)	1980-01
idiovol	Standard deviation of residuals of weekly returns on weekly equal weighted market returns for 3 years prior to month end	1980-01
ill	Average of daily (absolute return / dollar volume)	1980-01
indmom	Equal weighted average industry 12-month returns	1980-01
invest	Annual change in gross property, plant, and equipment (*ppegt*) + annual change in inventories (*invt*) all scaled by lagged total assets (*at*)	1980-01
IPO	An indicator variable equal to 1 if first year available on CRSP monthly stock file	1980-01
lev	Total liabilities (*lt*) divided by fiscal year-end market capitalization	1980-01
mom12m	11-month cumulative returns ending one month before month end	1980-01
mom1m	1-month cumulative return	1980-01
mom36m	Cumulative returns from months t-36 to t-13	1980-01
ms	Sum of 8 indicator variables for fundamental performance	1980-01
mve	Natural log of market capitalization at end of month t-1	1980-01
mve_ia	2-digit SIC industry adjusted fiscal year-end market capitalization	1980-01
nanalyst	Number of analyst forecasts from most recently available I/B/E/S summary files in month prior to month of portfolio formation. nanalyst set to zero if not covered in I/B/E/S summary file	1989-01
nincr	Number of consecutive quarters (up to eight quarters) with an increase in earnings (*ibq*) over same quarter in the prior year	1980-02
operprof	Revenue minus cost of goods sold - SG&A expense - interest expense divided by lagged common shareholders' equity	1980-01
orgcap	Capitalized SG&A expenses	1980-01
pchcapx_ia	2-digit SIC fiscal year mean adjusted percent change in capital expenditures (*capx*)	1980-01

Table IA.1 continued

Acronym	Definition	Start Date
pchcurrat	Percent change in *currat*	1980-01
pchdepr	Percent change in *depr*	1980-01
pchgm_pchsale	Percent change in gross margin (*sale-cogs*) minus percent change in sales (*sale*)	1980-01
pchsale_pchinvt	Annual percent change in sales (*sale*) minus annual percent change in inventory (*invt*)	1980-01
pchsale_pchrect	Annual percent change in sales (*sale*) minus annual percent change in receivables (*rect*)	1980-01
pchsale_pchxsga	Annual percent change in sales (*sale*) minus annual percent change in SG&A (*xsga*)	1980-01
pchsaleinv	Percent change in *saleinv*	1980-01
pctacc	Same as *acc* except that the numerator is divided by the absolute value of *ib*; if $ib = 0$ then ib set to 0.01 for denominator	1980-01
pricedelay	The proportion of variation in weekly returns for 36 months ending in month t explained by 4 lags of weekly market returns incremental to contemporaneous market return	1980-01
ps	Sum of 9 indicator variables to form fundamental health score	1980-01
rd	An indicator variable equal to 1 if R&D expense as a percentage of total assets has an increase greater than 5%	1980-07
rd_mve	R&D expense divided by end-of-fiscal-year market capitalization	1980-01
rd_sale	R&D expense divided by sales (*xrd/sale*)	1980-01
realestate	Buildings and capitalized leases divided by gross PP&E	1984-12
retvol	Standard deviation of daily returns from month t-1	1980-01
roaq	Income before extraordinary items (*ibq*) divided by one quarter lagged total assets (*atq*)	1980-01
roavol	Standard deviation for 16 quarters of income before extraordinary items (*ibq*) divided by average total assets (*atq*)	1981-11
roeq	Earnings before extraordinary items divided by lagged common shareholders' equity	1980-01
roic	Annual earnings before interest and taxes (*ebit*) minus nonoperating income (*nopi*) divided by non-cash enterprise value (*ceq*+*lt*-*che*)	1980-01
rsup	Sales from quarter t minus sales from quarter t-4 (*saleq*) divided by fiscal-quarter-end market capitalization ($cshoq \times precq$)	1980-01
salecash	Annual sales divided by cash and cash equivalents	1980-01
saleinv	Annual sales divided by total inventory	1980-01
salerec	Annual sales divided by accounts receivable	1980-01
secured	Total liability scaled secured debt	1980-12
securedind	An indicator equal to 1 if company has secured debt obligations	1980-12
sfe	Analysts mean annual earnings forecast for nearest upcoming fiscal year from most recent month available prior to month of portfolio formation from I/B/E/S summary files scaled by price per share at fiscal quarter end	1989-01
sgr	Annual percent change in sales (*sale*)	1980-01
sin	An indicator variable equal to 1 if a company's primary industry classification is in smoke or tobacco, beer or alcohol, or gaming	1980-01
SP	Annual revenue (*sale*) divided by fiscal year-end market capitalization	1980-01
std_dolvol	Monthly standard deviation of daily dollar trading volume	1980-01
std_turn	Monthly standard deviation of daily share turnover	1980-01
stdcf	Standard deviation for 16 quarters of cash flows divided by sales (*saleq*); if $saleq = 0$, then scale by 0.01. Cash flows defined as *ibq* minus quarterly accruals	1981-11

Table IA.1 continued

Acronym	Definition	Start Date
sue	Unexpected quarterly earnings divided by fiscal-quarter-end market cap. Unexpected earnings is I/B/E/S actual earnings minus median forecasted earnings if available, else it is the seasonally differenced quarterly earnings before extraordinary items from Compustat quarterly file	1980-01
tang	Cash holdings + 0.715×receivables + 0.547×inventory + 0.535×PPE / total assets	1980-01
tb	Tax income, calculated from current tax expense divided by maximum federal tax rate, divided by income before extraordinary items	1980-01
turn	Average monthly trading volume for most recent 3 months scaled by number of shares outstanding in current month	1980-01
zerotrade	Turnover weighted number of zero trading days for most recent 1 month	1980-01
News, sourced from Ravenpack		
sentiment_av	Monthly average of daily sentiment (*css*)	2000-01
sentiment_vw	Monthly average sentiment (*css* weighted by squared relevance (*relevance*)	2000-01
articles_pct_neg	Share of articles with a sentiment score below 25	2000-01
articles_pct_pos	Share of articles with a sentiment score above 75	2000-01
newsConcentration	HHI inspired measure of news concentration during a month. Sum of squared daily share of articles during a month	2000-01
sentiment_av_mom1	One month change in *sentiment_av*	2000-02
sentiment_vw_mom1	One month change in *sentiment_vw*	2000-02
sentiment_av_mom12	Average monthly change in *sentiment_av* over the past 12 months	2001-01
sentiment_vw_mom12	Average monthly change in *sentiment_vw* over the past 12 months	2001-01
abn_articles_std	Average number of articles per day standardized by 12 month trailing average and standard deviation	2001-01
Earnings calls, sourced from Streetviews and Seeking Alpha		
earningsCallMonth	An indicator variable equal to 1 if there was an earnings call during the month	2009-07
WC_intro	Total number of words from the introduction section	2009-07
ml_neg_intro	Percentage of negative words from the introduction section using GHR (Garcia et al. (2023)) machine learning dictionary	2009-07
ml_pos_intro	Percentage of positive words from the introduction section using GHR machine learning dictionary	2009-07
lm_neg_intro	Percentage of negative words from the introduction section using LM (Loughran and McDonald (2011)) dictionary	2009-07
lm_pos_intro	Percentage of positive words from the introduction section using LM dictionary	2009-07
lm_modal_weak_intro	Percentage of weak words from the introduction section using LM dictionary	2009-07
lm_modal_strong_intro	Percentage of strong words from the introduction section using LM dictionary	2009-07
lm_litigious_intro	Percentage of litigious words from the introduction section using LM dictionary	2009-07
lm_uncertainty_intro	Percentage of uncertainty words from the introduction section using LM dictionary	2009-07
WC_qa	Total number of words from the Q&A section	2009-07
ml_neg_qa	Percentage of negative words from the Q&A section using GHR machine learning dictionary	2009-07
ml_pos_qa	Percentage of positive words from the Q&A section using GHR machine learning dictionary	2009-07

Table IA.1 continued

Acronym	Definition	Start Date
lm_neg_qa	Percentage of negative words from the Q&A section using LM dictionary	2009-07
lm_pos_qa	Percentage of positive words from the Q&A section using LM dictionary	2009-07
lm_modal_weak_qa	Percentage of weak words from the Q&A section using LM dictionary	2009-07
lm_modal_strong_qa	Percentage of strong words from the Q&A section using LM dictionary	2009-07
lm_litigious_qa	Percentage of litigious words from the Q&A section using LM dictionary	2009-07
lm_uncertainty_qa	Percentage of uncertainty words from the Q&A section using LM dictionary	2009-07
EDGAR, sourced from SEC Analytic Suite		
k10 filingMonth	An indicator variable equal to 1 if the firm filed a 10-K during the month	1997-03
k10 fsize	File size of the 10-K	1997-03
k10 word count	Total number of words from 10K	1997-03
k10 lm negative	Percentage of negative words from 10K using LM dictionary	1997-03
k10 lm positive	Percentage of positive words from 10K using LM dictionary	1997-03
k10 lm modal weak	Percentage of weak words from 10K using LM dictionary	1997-03
k10 lm modal strong	Percentage of strong words from 10K using LM dictionary	1997-03
k10 lm litigious	Percentage of litigious words from 10K using LM dictionary	1997-03
k10 lm uncertainty	Percentage of uncertainty words from 10K using LM dictionary	1997-03
q10 filingMonth	An indicator variable equal to 1 if the firm filed a 10-Q during the month	1997-03
q10 fsize	File size of the 10-Q	1997-03
q10 word count	Total number of words from 10Q	1997-03
q10 lm negative	Percentage of negative words from 10Q using LM dictionary	1997-03
q10 lm positive	Percentage of positive words from 10Q using LM dictionary	1997-03
q10 lm modal weak	Percentage of weak words from 10Q using LM dictionary	1997-03
q10 lm modal strong	Percentage of strong words from 10Q using LM dictionary	1997-03
q10 lm litigious	Percentage of litigious words from 10Q using LM dictionary	1997-03
q10 lm uncertainty	Percentage of uncertainty words from 10Q using LM dictionary	1997-03
k8 nr individual filings	Number of 8Ks the firm filed during the month	1997-03
k8 fsize	Average file size of the 8K	1997-03
k8 word count	Total number of words from 8K	1997-03
k8 lm negative	Percentage of negative words from 8K using LM dictionary	1997-03
k8 lm positive	Percentage of positive words from 8K using LM dictionary	1997-03
k8 lm modal weak	Percentage of weak words from 8K using LM dictionary	1997-03
k8 lm modal strong	Percentage of strong words from 8K using LM dictionary	1997-03
k8 lm litigious	Percentage of litigious words from 8K using LM dictionary	1997-03
k8 lm uncertainty	Percentage of uncertainty words from 8K using LM dictionary	1997-03
Macro-economic variables, following Welch and Goyal (2008)		
dp	Dividend Price Ratio is the difference between the log of dividends and the log of prices. Dividends are 12-month moving sums of dividends paid on the S&P 500 index	1980-01
ep	Earnings Price Ratio is the difference between the log of earnings and the log of prices. Earnings are 12-month moving sums of earnings on the S&P 500 index	1980-01
bm	Book-to-Market Ratio is the ratio of book value to market value for the Dow Jones Industrial Average	1980-01
ntis	Net Equity Expansion is the ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks	1980-01

Table IA.1 continued

Acronym	Definition	Start Date
tbl	3-Month Treasury Bill rates, secondary Market Rate from the economic research data base at the Federal Reserve Bank at St. Louis	1980-01
tms	Term Spread is the difference between the long term yield on government bonds and the Treasury-bill	1980-01
dfy	Default Yield Spread is the difference between BAA and AAA-rated corporate bond yields	1980-01
svar	Stock Variance is computed as sum of squared daily returns on the S&P 500	1980-01
Ownership, sourced from 13F-filings		
own num	The number of 13F institutional investors holding the stock	1980-03
own hfidx	The ownership concentration as measured by the Herfindahl-Hirschman Index	1980-03
own percent	The percentage of shares outstanding owned by 13F institutional investors	1980-03