Active Machine-Learning-Based Trading and Mutual Fund Performance

Xiaowen Hu (Southern Methodist University)
Maximilian Rohrer (NHH–Norwegian School of Economics)
Hanjiang Zhang (Washington State University)

University of Maryland / Singapore Management University / UBS Quant Investment Forum: AI and Finance

June 2026

Motivation

- The advancement of artificial intelligence (AI), particularly machine learning (ML), coupled with the rise of big data, has transformed the landscape of knowledge production (Abis and Veldkamp, 2024).
- It remains unclear whether professional investors, such as mutual funds, can effectively leverage this revolutionary technology and **utilize ML**-generated signals to enhance **investment strategy** and **performance**.

Motivation

- ML generates highly **profitable** trading signals (Gu et al., 2020) and asset managers have a strong **incentive** to utilize AI-driven strategy.
- Yet profitably adopting an ML-based trading strategy is **challenging**:
  - **Building** the strategy is complex, requiring human and physical capital.
  - **Implementing** the strategy is challenging: high turnover and trading costs, and constrained by market conditions and style (Avramov et al., 2023; Jensen et al., 2024).
  - Investment **mandates** can limit mutual fund managers' ability to deviate from their benchmarks.

Research Questions

- We comprehensively investigate the adoption of ML-based trading strategies by U.S. active equity mutual funds.
- We ask three main questions:
  1. Do mutual funds **utilize** ML-based trading strategies in their investment practices?
  2. Can ML-based trading help mutual funds achieve **superior portfolio performance**?
  3. What are the **mechanisms** that drive the performance of funds adopting the ML-based trading strategies?

Holding-Based Methodology and AMLT

- To address these questions, we propose an **Active Machine-Learning-Based Trading (AMLT)** measure.
- AMLT is a **signed active-share** measure based on mutual fund **holdings**.
  - It captures how a fund actively aligns its portfolio weight deviations from its benchmark with forward-looking ML-based trading signals.
- As a trade-alignment measure, AMLT provides a sharp identification of **AI adoption**, specifically in **enhancing investment** strategy and fund performance.

Preview of the Main Results

- We document a significant upward trend in the adoption of ML-based trading strategies in the mutual fund industry, with substantial variation across funds.
- High AMLT funds have high employee AI talent, while AMLT remains the only significant predictor of fund performance.
- Top-AMLT funds outperform bottom funds by 2.4% to 3.0% annually after risk adjustment. Outperformance comes from superior stock selection, lower expenses, and efficient trading-cost management.
- We identify two key drivers of superior performance:
  1. ML's capacity to process extensive information and capture complex interactions,
  2. Fund managers' ability to integrate AI with human investment expertise.

Related Literature

- **Mutual fund performance:**
  Daniel et al. (1997); Kacperczyk et al. (2007, 2008); Cremers and Petajisto (2009); Huang et al. (2011); Cremers et al. (2013); Petajisto (2013); Stambaugh (2014); Jiang and Zheng (2018); Abis and Lines (2024).
- **Information utilization and fund performance:**
  Kacperczyk and Seru (2007); Cohen et al. (2008); Fang et al. (2014); Abis and Veldkamp (2024).
- **Machine learning in finance:**
  Kelly et al. (2019); Freyberger et al. (2020); Gu et al. (2020); Kozak et al. (2020); Avramov et al. (2023); Chen et al. (2024); Choi et al. (2024); Feng et al. (2024); Jensen et al. (2024); Cong et al. (2025); Cong et al. (2026); Lu, Spiegel, and Zhang (2026); D'Acunto et al. (2019); Van Binsbergen et al. (2023); Liu (2022); Fuster et al. (2022); Zheng (2024); Cao et al. (2024); Li and Rossi (2020); Kaniel et al. (2023); DeMiguel et al. (2023); Ma et al. (2025).
- **AI and new data in asset management:**
  Bonelli and Foucault (2023); Zhang (2024); Cen et al. (2024); Sheng et al. (2024); Chen et al. (2025).

Overview

- Introduction
- **Data and Sample Construction**
- AMLT Construction and Discussion
- AMLT and Mutual Fund Performance
- Mechanisms behind AMLT Performance
- Conclusion

Data and Sample Construction

- **Mutual Fund:**
  CRSP Survivorship Bias Free Mutual Fund Database, Thomson-Reuters Mutual Fund Holdings database. Actively managed U.S. equity mutual funds – 5,024 unique funds from 2000 to 2022.

- **Firm-Level:**
  NYSE, AMEX, and NASDAQ firms, covering 18,803 unique stocks, averaging 4,708 per month.

- **Validation:**
  Lightcast employee profile data – comprehensively linked to mutual fund advisory companies with over 2.15 million unique employee profiles.

Data and Sample Construction

- Firm-level information inputs – the ML trading signals
  - **Quantitative** information:
    - 94 stock-level characteristics and 74 industry dummies — Green et al. (2017), Gu et al. (2020).
  - **Textual** information:
    - **Corporate filings** (SEC Analytics) — 238,758 10-Ks · 666,633 10-Qs · 1.9 million 8-Ks – Loughran and McDonald (2011).
    - **Earnings call transcripts** (Wall Street Horizons) — 261,179 quarterly transcripts – Garcia et al. (2023).
    - **Firm-specific news** (RavenPack) — 697 million unique articles – Ke et al. (2019), Von Beschwitz et al. (2020), Jiang et al. (2021), Jeon et al. (2022), Lopez-Lira and Tang (2023).
  - Other inputs:
    - Macroeconomic variables and institutional ownership variables — Goyal and Welch (2008), Cao et al. (2024).

Overview

- Introduction
- Data and Sample Construction
- **AMLT Construction and Discussion**
  - Construction of AMLT
  - AMLT, Fund Characteristics, and Persistency
  - External Validation
- AMLT and Mutual Fund Performance
- Mechanisms behind AMLT Performance
- Conclusion

Construction of AMLT

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

AMLT at t: a fund aligns (1) its active positions at t with (2) ML signals for $t+1 \mid$ info t.

(1) Active weight deviation: $w_{i,t}^{f} - w_{i,t}^{B}$

- Capture a fund's **intentional** decisions, not passive benchmark tracking.
- Benchmark = the one of 21 S&P/Barra and Russell indices with the lowest active share (Cremers and Petajisto, 2009; Jiang and Zheng, 2018).

(2) ML trading signal$_{i,\,t\,+\,1|t}$

- Follow Gu et al. (2020) by implementing a DNN.
- Inputs: quantitative predictors + large-scale **textual data**.
- Rank-transform raw predictions to decile scores from −5 to +5 (Jensen et al., 2024) — corresponding to negative/positive returns (i.e., **sell/buy signals**).

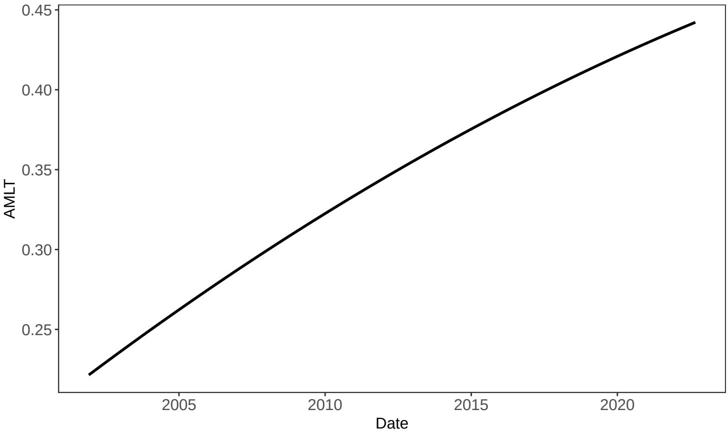
◀ ML Methodology Detailed ◀ Extraction of Textual Signals

Advantages of AMLT

- AMLT provides a sharp identification of ML adoption in the **investment process itself** — not in other activities (e.g., marketing, IT, customer service).
- Enables a detailed **mechanism** analysis.
  - Adapts to alternative ML models (linear/nonlinear) and information sets (quantitative/textual).
- Applies to the entire U.S. mutual fund universe, at the **fund level**.
- Alternative approaches — labor-market data, or fund self-disclosures — face limitations (Babina, 2026):
  - capture all-purpose adoption; manipulable; at adviser level
  - mainly serve as **validation** for AMLT.

◂ ML Adoption in Asset Management

AMLT Adoption Rises Over Time



- A clear upward trend in AMLT adoption in mutual funds.
- The upward trend prevails in almost all style groups.

Figure Style

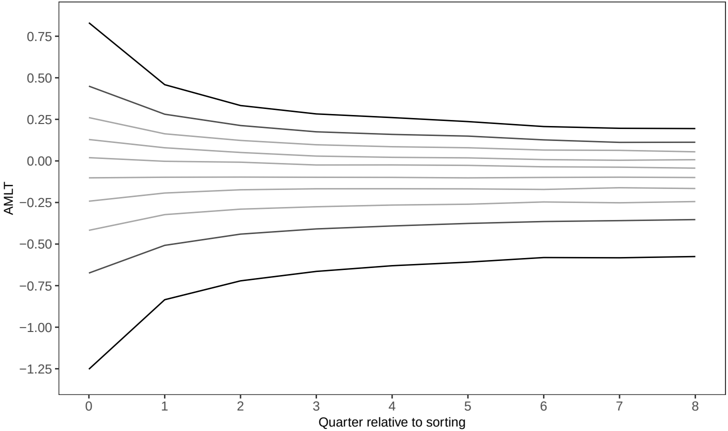
AMLT and Fund Characteristics

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)

Top AMLT funds:

- Actively deviate from their benchmark by 23.3% (AMLT_Tilt)
- Strong past performance, lower risk, lower fees.
- Lower turnover, smaller size.
- Distinct from active share.

Persistency of AMLT



- Relative ranks among the AMLT decile portfolios remain unchanged over two years.

Transition Matrix

Overview

- **External Validation of AMLT**
  - Self-Designated AI Funds and Quant Funds
  - Mutual Fund Employee AI Talent

AMLT and Self-Designated AI Funds and Quant Funds

	AI funds			Quant fund		
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***					
	(0.036)					
Quant $fund_t$				0.153***		
				(0.049)		
Controls	Yes			Yes		
Time fixed effects	Yes			Yes		
Style fixed effects	Yes			Yes		
Observations	140,577			140,577		
Adjusted R^2	0.095			0.089		

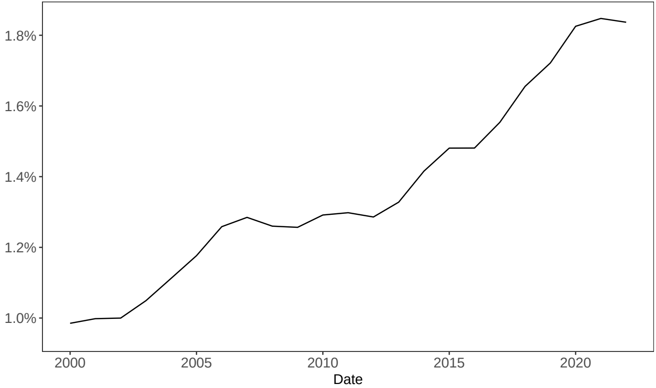
- Self-designated AI funds – substantially higher AMLT.
- Quant funds:
  - higher than average funds, about 1/5 of AI funds.
  -

AMLT and Self-Designated AI Funds and Quant Funds

	AI funds			Quant fund		
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)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
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

- Self-designated AI funds – substantially higher AMLT.
- Quant funds:
  - higher than average funds, about 1/5 of AI funds.
  - only correlated with AMLT_Quant, not AMLT_Text.

AMLT and Mutual Fund Employee AI Talent



- Mutual fund employee AI talent (Babina et al., 2024)
  - **Share_CoreAI**: the share of employees with core AI skills (i.e., AI, ML, NLP, and CV) – advisor level.
- A clear upward trend in fund AI talent, consistent with holdings-based AMLT evidence.

AMLT and Mutual Fund Employee AI Talent

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

- High AMLT is associated with high employee AI talent.

Overview

AMLT and Future Fund Performance: Portfolio Approach

	Bottom 3				**Top 3**			
Portfolio	1	2	3	⋯	8	9	10	H-L
				AMLT				
CAPM-α	-0.117	-0.038	-0.016	⋯	0.020	0.058	0.118*	0.235***
	(0.089)	(0.061)	(0.050)		(0.041)	(0.050)	(0.067)	(0.080)
FF-α	-0.131**	-0.047	-0.022	⋯	0.015	0.051	0.108***	0.238***
	(0.065)	(0.046)	(0.039)		(0.028)	(0.031)	(0.040)	(0.079)
Carhart-α	-0.106	-0.045	-0.027	⋯	0.007	0.042	0.090**	0.196**
	(0.064)	(0.046)	(0.039)		(0.028)	(0.031)	(0.040)	(0.076)
PS-α	-0.111*	-0.049	-0.029	⋯	0.006	0.041	0.089**	0.199***
	(0.063)	(0.044)	(0.038)		(0.028)	(0.031)	(0.039)	(0.075)
				AMLT_Long				
CAPM-α	-0.120	-0.034	-0.031	⋯	0.030	0.063	0.122**	0.243***
	(0.082)	(0.060)	(0.049)		(0.043)	(0.050)	(0.062)	(0.079)
FF-α	-0.132**	-0.042	-0.037	⋯	0.024	0.057*	0.114***	0.246***
	(0.063)	(0.046)	(0.036)		(0.028)	(0.031)	(0.041)	(0.078)
Carhart-α	-0.107*	-0.041	-0.043	⋯	0.014	0.047	0.098**	0.205***
	(0.062)	(0.046)	(0.037)		(0.028)	(0.031)	(0.041)	(0.075)
PS-α	-0.112*	-0.044	-0.045	⋯	0.013	0.046	0.097**	0.208***
	(0.060)	(0.045)	(0.036)		(0.028)	(0.030)	(0.041)	(0.075)

- Top-decile AMLT funds earn significant alphas of 1.1% to 1.4% annually; outperform bottom-decile funds by 2.4% to 3.0%
- **AMLT_Long** yield highly consistent and slightly stronger results.

AMLT and Future Fund Performance: Multivariate Regressions

Dependent Variable:	Carhart-α_t			
Model:	Net return (1)	Gross return (2)	Net return (3)	Gross return (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

- Mutual funds following an active ML-based strategy earn significantly better risk-adjusted performance in the subsequent quarter, after additional controls.

Performance Decomposition

- We follow Daniel, Grinblatt, Titman, and Wermers (1997) and Sialm and Zhang (2020) to decompose net fund returns:

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

  - CS (Characteristic Selectivity) – stock selection ability
  - CT (Characteristic Timing) – style timing
  - AS (Average Style) – returns from average style
  - EXP – Expense Ratio
  - RG (Return Gap) – interim trading benefits and trading costs

Performance Decomposition

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

- Superior performance of AMLT funds can be explained primarily by:
  - Superior stock selection ability (CS)
  - Lower expenses (Exp)
  - Skill to efficiently manage trading costs (RG)

Overview

Mechanisms behind AMLT Performance

1. **ML Technology and Information Advantage**
   - **Volume of information** – ML thrives on data: do high-AMLT funds leverage this to tilt toward information-rich stocks, and profit from it?
   - **Linear vs. nonlinear models** – does the edge come from ML's ability to capture complex interactions that linear models miss?
   - **Quantitative vs. textual inputs** – is the value added from combining traditional data with new unstructured text?
   - **Individual information categories** – is it from the combination, not any single anomaly category, or individual text data source?

2. Managerial Skill and Sustained Performance

Volume of Information – Active Portfolio Tilt

Dependent Variables:	ATILT_N.Signals$_t$	ATILT_N.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

- AMLT funds tilt significantly more toward information-rich stocks — where ML yields a greater information advantage.

Volume of Information – Performance

Dependent Variables:	Carhart-α_t			
Information proxy:	N.Signals	N.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)
Controls	Yes	Yes	Yes	Yes
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

- High_Info = funds that hold the most information-rich stocks (top tercile).
- AMLT funds, focusing on information-rich stocks, achieves even stronger performance.

Linear vs. Nonlinear Models

Dependent Variables:	Carhart-α_t		
ML model:	DNN	ENet	PCR
Model:	nonlinear, with interactions (1)	linear, shrinkage + selection (2)	linear, factor reduction (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

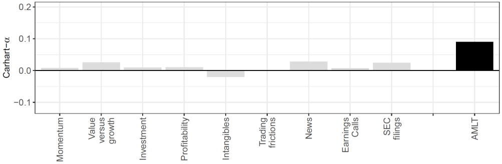
- AMLT based on the nonlinear DNN more than **doubles** the performance gain of the linear ENet and PCR.
- Substantial value added by ML's ability to capture **complex interactions**.

Quantitative vs. Textual Inputs

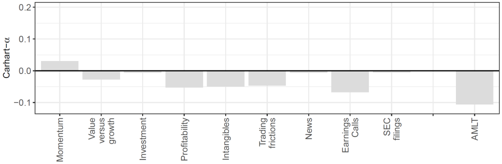
Dependent Variables:	Carhart-α_t		
Information input:	All	Quantitative	Textual
Model:	(1)	(2)	(3)
AMLT_{t-1}	0.054***	0.028***	0.017***
	(0.006)	(0.005)	(0.006)
Controls	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes
Style fixed effects	Yes	Yes	Yes
Observations	406,897	406,897	406,897
Adjusted R^2	0.070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

- AMLT's performance gain from the full information set is **2–3 times** that from quantitative or textual data alone.
- Substantial value added by combining **diverse information**.

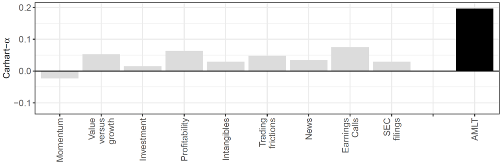
Multiple vs. Individual Information Categories



(a) Top Portfolios



(b) Bottom Portfolios



(c) Top-minus-Bottom Portfolios

- Classify information into:
  - 6 anomaly categories (Hou et al. (2020))
  - 3 textual categories
- AMLT is a distinct strategy — no single-category strategy comes close.

Mechanisms behind AMLT Performance

1. ML Technology and Information Advantages

2. **Managerial Skill and Sustained Performance**
    - Long-Term Performance
    - Market Conditions
    - Fund Styles

Long-Term Performance

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

- High AMLT funds achieve significant long-term performance.

Market Conditions

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)
$AMLT_{t-1} \times High_{t-1}$	0.033***	0.069***	0.045***
	(0.011)	(0.012)	(0.011)
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

- High AMLT funds achieve persistent performance across all market conditions, with greater performance in high market conditions.

Fund Styles

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)
Controls	Yes	Yes	Yes	Yes
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

- Funds in all style groups are able to add value by utilizing AMLT strategy.

Additional Analyses and Robustness Tests

- Subperiods and Early ML Model
- AMLT vs. Employee AI Talent
- Controlling for Management/Advisory Company Fixed Effects
- Controlling for Active Share Measure
- Controlling for Active Anomaly Trading (Stambaugh et al. (2012))
- Alternative Measures: TO-Adjusted ML Trading, Beginning-of-Qtr Info, and Signal Transformations

Conclusion

- We introduce AMLT, a **holdings-based active-share measure** that directly captures ML adoption in mutual fund investments, validated with comprehensive employee AI talent data.
- We document a significant rising trend of **ML adoption by mutual funds** over the past two decades, as well as significant value added by such practice.
- **Superior performance** is attributable to better stock selection, lower expenses, and efficient trading-cost management.
- We identify two key drivers of superior performance: ML technology and fund managers' ability to integrate AI.

ML- Methodology

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

- 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.
- Three-layered neural network architecture with 32, 16, 8 neurons. Across all layers, we employ the rectified linear unit (ReLU) activation function.
- The dependent variable is rank-transformed and scaled to range from 0 to 1. Predictors are normalized in the cross-section.
- We employ an ensemble approach following Hansen and Salamon (1990) and Dietterich (2000) – we average forecasts across ten trained networks with random seeds.
- We tune hyperparameters through grid search utilizing a 10-year training window and a 4-year validation window. We re-tune hyperparameters annually.
- We re-estimate the model using the selected hyper-parameters for every monthly prediction.
- **ML 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.**

Return

Textual Signal Extraction

1. Corporate filings (SEC Analytics)
   - 238,758 annual reports (10-Ks), 666,633 quarterly reports (10-Qs), and 1.9 million current reports (8-Ks).
   - 27 predictors from SEC Analytics (9 per corpora) including file size, word count, as well as sentiment (and other language markers) as proposed by Loughran and McDonald (2011).
2. Earnings calls (Wall Street Horizons)
   - 261,179 quarterly earnings call transcripts
   - 19 predictors including file size, timing, as well as sentiment (and other language markers) constructed following García et al. (2023) and Loughran and McDonald (2011), separately for introduction and Q&A.
3. Firm-specific news articles (RavenPack News Analytics)
   - 697 million unique news articles
   - Ten predictors from RavenPack aggregated at firm-month level. Predictors cover news frequency and concentrations, as well as several variants of sentiment, separately accounting for level and momentum.

Return

ML Adoption in Asset Management

- Deep neural networks (DNNs) were conceptually developed as early as the 1940s and became practically implementable in the 1980s.
- Anecdotal evidence that hedge funds were already experimenting with machine learning models in the 1990s.
  - E.g., Renaissance Technologies' Medallion Fund and Goldman Sachs' Global Alpha Funds used computer-based models.
- 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.
  - LeBaron (1999) demonstrates that neural networks can forecast asset returns.
- Increased adoption after 2010 due to explosion of big data and advancements in deep learning.
  - Many adopters such as BlackRock, Two Sigma, AQR, Vanguard Group, State Street Global Advisors, and Cerebel- lum Capital, etc.
- We examine an earlier machine learning model—Random Forest, introduced around 2001—and find that it generates significant performance, despite being weaker compared to more advanced models.

Return

Subsample Analysis for Earlier and Later Periods

Dependent var.	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 FE	Yes	Yes	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	406,897	175,621	231,276	406,897	175,621	231,276
Adj. R^2	0.070	0.049	0.096	0.070	0.049	0.096

- AMLT constructed from DNN predicts performance in both pre- and post-2010 periods, with stronger predictive power in the later period. AMLT based on Random Forest (Breiman (2001)) also yields significant performance, despite being weaker.

Return

Employee AI Talent and Future Fund Performance

Dependent var.	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 FE	Yes	Yes
Style FE	Yes	Yes
Obs.	118,885	118,885
Adj. R^2	0.069	0.070

- The strength of AMLT: Capture AI adoption specifically aimed at enhancing trading strategies and portfolio performance.

Return

Controlling for Active Share Measure

Dependent var.	Carhart-α_t			
	AMLT		AMLT_Long	
	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)
Active $Share_{t-1}$	0.000	0.003	0.000	0.003
	(0.014)	(0.014)	(0.014)	(0.014)
Ret_{t-1}	0.016**	0.015**	0.016**	0.016**
	(0.007)	(0.007)	(0.007)	(0.007)
Exp $ratio_{t-1}$	-1.081***	-0.134	-1.085***	-0.138
	(0.117)	(0.117)	(0.116)	(0.116)
$Log(TNA)_{t-1}$	-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 FE	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	Yes	Yes
Obs.	406,897	406,462	406,897	406,462
Adj. R^2	0.070	0.069	0.070	0.069

- Negligible impact, results primarily driven by ML-powered active strategy.

Return

Controlling for Active Trading toward Composite Anomaly (Stambaugh et al., 2012)

Dependent var.	Carhart-α_t	
Model	(1)	(2)
AMLT_{t-1}	0.054***	0.051***
	(0.006)	(0.006)
$\text{AT\_Composite}_{t-1}$		0.014
		(0.011)
Ret_{t-1}	0.016**	0.015**
	(0.007)	(0.007)
Exp ratio_{t-1}	-1.080***	-1.075***
	(0.105)	(0.105)
Log(TNA)_{t-1}	-0.005***	-0.005***
	(0.002)	(0.002)
Turnover_{t-1}	-0.056***	-0.057***
	(0.007)	(0.007)
Fund age_{t-1}	0.000	0.000
	(0.000)	(0.000)
Time FE	Yes	Yes
Style FE	Yes	Yes
Obs.	406,897	406,730
Adj. R^2	0.070	0.070

- AMLT is essentially unchanged and remains highly significant. Active anomaly trading (AT) is insignificant.

Return

Alternative Turnover-Adjusted ML Trading Measure

Dependent var.	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 FE	Yes	Yes
Style FE	Yes	Yes
Obs.	377,330	376,937
Adj. R^2	0.072	0.071

- TOMLT focuses on mutual fund trading toward strong ML signals, and adjusts for the fund's overall quarterly trading activity. Consistent with AMLT, TOMLT also significantly and positively predicts future fund performance.

Return

Alternative AMLT Using Beginning-of-Quarter Information

Dependent var.	Carhart-α_t	
Model	Net return (1)	Gross return (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 FE	Yes	Yes
Style FE	Yes	Yes
Obs.	406,897	406,462
Adj. R^2	0.070	0.069

- AMLT based on ML trading signals generated using information available at the end of the previous quarter yields results highly consistent with the baseline findings.

Return

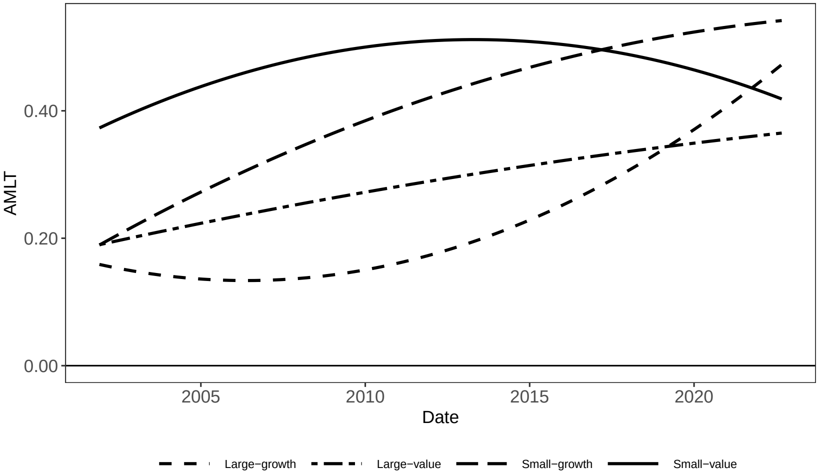
Alternative ML Signal Transformations

	AMLT_Alt1				AMLT_Long_Alt1			
	CAPM	FF3	Carhart	PS	CAPM	FF3	Carhart	PS
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**			
	CAPM	FF3	Carhart	PS	CAPM	FF3	Carhart	PS
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)

- Results are highly consistent with the baseline findings.

Return

AMLT Adoption Across Style Groups



Transition Matrix for Funds Sorted into AMLT Quintiles

Panel A: One Quarter Ahead

	Portfolio$_{t+1}$							
Portfolio$_t$	1	2	3	4	5	Attrition	Δ Off-diag.	S.E.
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 Quarters Ahead

	Portfolio$_{t+2}$							
Portfolio$_t$	1	2	3	4	5	Attrition	Δ Off-diag.	S.E.
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.	S.E.
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)		

Return