Assessing the Performance of AI-Managed Portfolios

Mauricio Praxmarer and Iván Simon*†

This draft: April 2024

Abstract

We evaluate the performance of AI-managed funds. Using a hand-collected dataset of AI-managed U.S. mutual funds our analysis reveals that, while these funds do not outperform a market benchmark, they nonetheless generate slightly higher cumulative returns than their human-managed counterparts with similar investment objectives and fund characteristics. Notably, we observe a recent decline in the performance of AI-enhanced funds. Our findings indicate that, compared to their peers, these funds excel in market timing but struggle with stock selection. AI-managed funds tend to be less active. In addition, our findings indicate continued underperformance in both, AI and rival funds, with weak evidence of momentum in successful AI funds and reversal in rival funds.

Keywords: machine learning, artificial intelligence, natural language processing, portfolio management

(JEL G10, G11)

*Department of Banking and Finance, University of Innsbruck, Universitätstraße 15, A-6020 Innsbruck, Austria, corresponding emails: mauricio.praxmarer@uibk.ac.at and ivan.simon@student.uibk.ac.at

†We thank Moritz Dauber, Dennis Umlandt and University of Innsbruck for valuable comments. The computational results presented here have been achieved (in part) using the LEO HPC infrastructure of the University of Innsbruck.

1 Introduction

Previous literature has investigated the potential of including artificial intelligence (AI) in asset pricing contexts (e.g. Gu et al., 2020, Freyberger et al. (2020), Chen et al. (2024), Azevedo et al. (2023b), etc). However, Chen and Velikov (2023) show that most anomalies disappear once one accounts for external costs. Likewise, most AI-enhanced portfolio strategies report extremely high (paper) returns and subsequent high Sharpe Ratios (SR). Considering the limitations and overlooked costs associated with AI strategies (e.g. limits-to-arbitrage or post-publication decay), we aim to investigate the actual profitability of AI portfolio strategies in this study. Unlike previous research, which often neglects or proxies expenses, we adopt a hands-on approach. We screen all U.S. mutual funds for AI, machine learning (ML), or deep learning (DL) components in the portfolio creation process and investigate whether they have superior returns compared to conventional funds. This approach provides a more accurate, real world assessment of the profitability of AI-enhanced strategies.

Our findings indicate that the AI-managed funds investigated do not outperform a market portfolio. However, when compared to peer funds that are matched upon having similar fund characteristics, AI funds exhibit comparable performance.

Moreover, our analysis highlights the significance of how AI is utilized. Portfolios incorporating AI in the early stages of portfolio construction, such as for screening undervalued assets based on predefined characteristics, outperformed a subgroup of portfolios that utilized AI for optimizing allocations, such as estimating expected returns and variances. The former group of portfolios demonstrated substantially higher returns compared to the latter. Additionally, portfolios utilizing AI in the initial stages of portfolio construction exhibited significantly lower levels of activity.

In Section 2, supplemented by Appendix A, we outline our approach to distinguishing AI-managed funds from those that do not employ AI in their asset management processes. We screen the investment descriptions of funds through their 497K filings for AI-related keywords and manually verify matched funds. While this method ensures that the identified funds utilize some form of AI, machine learning, deep learning, etc., it has limitations. Ambiguous terms such as "quantitative methods," "mathematical models," or "statistical analysis" hint at AI utilization but do not definitively identify AI-managed funds. Consequently, such cases are excluded from our AI fund sample to maintain accuracy. This approach results in a selection bias, as we focus only on funds explicitly framing themselves as AI-enhanced, leading some to refer to our sample as AI-labeled rather than AI-optimized. We acknowledge this distinction and use both terms interchangeably, allowing readers to draw their own interpretations.

AI funds are compared to the market portfolio. Alternatively, we match similar funds to each identified AI fund. Conventional and AI funds are categorized by their attributes, such as size, momentum, or dividend yield. Close "neighbours" of AI-labelled funds are then bundled into a rival portfolio. This method allows for better comparisons, since it creates a fictional Non-AI portfolio with similar properties to the AI fund, leading to an appropriate counterfactual. Section 2 outlines how we match funds and the specifications in more detail. Additionally, Section 5 explores alternative specifications.

We test whether the advantages of AI over human fund managers translates into higher yields compared to conventional funds and the market benchmark. Our findings indicate that while AI-powered funds do not outperform the aggregate stock market, neither do comparable conventional rival funds demonstrate a sustainable advantage over AI funds. A portfolio comprising equal weights of AI funds yields a SR of 0.122 over the observation period. In comparison, a similar portfolio composed of conventional benchmark funds, which we match based on similar fund attributes, yields a SR of 0.153 during the same period, with the difference in SRs being statistically insignificant.

In a second step, we delve into the stock selection and timing abilities of AI funds. Leveraging AI technology allows funds and managers to process an unprecedented volume of data when making investment decisions. Pástor et al. (2020) decompose the excess return of a fund into two components: a fund-specific variable capturing the overall skill level of the fund, and a second component, namely a function that measures how actively a given fund exploits this specific set of skills. Building on this decomposition, we test two additional hypotheses. First, we investigate if AI funds exhibit any considerable level of skill. And consequently, we examine how actively an AI fund applies that skill. Our results in Section 3 reveal that, on average, AI funds display lower activity levels than their peers, especially those that apply AI at earlier stages of portfolio management.

Our results, further, reveal that AI funds exhibit timing abilities that surpass those of their rivals by 48 basis points (bps). However, this advantage is offset by an inferior stock-picking ability worth 58 bps.

Our study strongly relates to the findings of Chen and Ren (2022). We also find that AI funds per se do not outperform the market benchmark and one composed of similar rival funds. Additionally, we sub-categorize the results to explore the implications of employing AI in an early stage of portfolio management, including screening the investable universe and identifying potentially profitable assets, as well as using AI in a later stage to optimize allocation strategies. Furthermore, we include additional data in our analysis, which adds significant value to the study. The interest in "Artificial Intelligence" surged post-2020[1]. Consequently, we also observe a considerable increase in the number of funds compared to the 15 AI-managed funds examined in the comparative study.

We further distinguish our research from other papers in several ways: While many prior investigations focus on AI-enhanced portfolio strategies, most are based on hypothetical returns. In contrast, our analysis centres on mutual fund data, allowing us to examine real-world strategies in action. This means we assess outcomes after factoring in actual trading fees, management fees, market impacts induced by trading, information costs, and other associated fees. Previous studies often rely on proxies for these costs, limiting their accuracy. Additionally, some theoretically profitable AI-enhanced strategies may not be feasible in practice due to factors like high turnover, limited liquidity for certain assets, post-publication decay, and other constraints to arbitrage (Avramov et al., 2023, Chen and Velikov, 2023).

The previous literature already investigated the potential upsides of AI over human investors. One notable advantage is AI's capability to analyze vast amounts of data — from news reports to financial statements to macroeconomic indicators. Noteworthy examples of AI's agility in decision-making include Deep Blue's victory over Garry Kasparov, the then-reigning chess champion, in 1997, and AlphaGo's triumph over Lee Sedol, one of the top human players, in four out of five games of "Go"[2] in 2016. Drawing from those landmark achievements, it's evident that AI excels

[1] According to Google's global interest index, interest in "Artificial Intelligence" was rated at 11 (out of 100) in December 2019, marking the end of the time-frame analyzed by Chen and Ren, 2022. However, as of February 2024, this index has reached its all-time high, defined at 100.

[2] "Go" is widely regarded as one of the most complex board games.

in complex decision-making tasks. AIEQ, the pioneering AI fund, claims to leverage historical and real-time economic data alongside analyzing one million news reports concurrently for security selection[3]. This capability enables AI to make decisions swiftly, surpassing the potential cognitive limitations and behavioural biases inherent in human fund managers. Studies such as those by Linnainmaa et al. (2021) have highlighted human managers' tendencies towards excessive trading, return chasing, and underdiversification, which have contributed to declining fund performance over time. In that regard, Barras et al. (2010) find a significant proportion of skilled (positive alpha) funds before 1996, but almost none by 2006. As a result, AI could pose an unbiased and data-driven alternative to improve security selection. Previous studies report higher SR for AI-enhanced strategies compared to linear alternatives. Some extreme cases, report SR beyond 2.0 and alphas beyond 13 % (Chen et al., 2024, Cong et al., 2021).

Those mentioned advantages led to research that included AI in trading strategies with exceptional risk-adjusted performances. Yet the applicability in real funds has not been answered: Kelly et al. (2024) theoretically demonstrate that "complex" models, with parameters exceeding the number of observations, should outperform simpler models in terms of return predictability. Azevedo et al. (2023a) test several sophisticated ML algorithms with constraints to reduce turnover rates and further cost mitigation techniques and accounting for trading costs through bid-ask estimates. Their ML-based strategies produce significant out-of-sample average monthly returns of up to 1.42% after *estimated* costs. On the other hand, Chen and Velikov (2023) show that anomalies disappear, when accounting for (1) transaction costs, (2) post-publication decay (c.f. McLean and Pontiff, 2016), and (3) higher liquidity and lower trading costs due to the new trading technologies that were implemented in the early 2000s (c.f. Chordia et al., 2014). This is even before accounting for other costs such as price impact, short-sale fees, or fund-related fees that arise when these anomalies are subsequently managed through a fund. Also, this does not account for further errors such as non-standard errors (see e.g. Walter et al., 2023). Furthermore, Martin and Nagel (2022) argue that an increase of predicting variables, i.e. information that may be processed, through machine learning still may lead to overfitting and that a Bayesian investor should still look for risk-based or behavioural explanations to make predictions and subsequent investment decisions. Chen and Ren (2022) take a real market approach and identify the first AI-managed funds available to investors and measure their performance. Similar to Chen and Velikov (2023) they do not find strong

[3]see `https://amplifyetfs.com/aieq/`

evidence for an outperformance of AI-managed strategies. However, their sample is rather short (due to the first AI-managed fund being launched in October 2017), which makes finding strong evidence difficult.

From an asset perspective, previous studies suggest that AI may also help to improve market efficiency. AI can process vast amounts of data, including data that would not have been incurred by human managers, and trade upon it. The initially neglected data may then find its way into market prices through AI-related trades. This may then enhance price informativeness, potentially leading to greater market efficiency (c.f. Zhu, 2019, Dugast and Foucault, 2018, Grennan and Michaely, 2021, etc.). However, Farboodi and Veldkamp (2020) argue that while increased price informativeness can reduce uncertainty regarding future payoffs, it also amplifies price sensitivity to shocks in expectations of cash flows or stock demand, introducing new sources of risk. Additionally, the authors highlight that as technology advances and price informativeness rises, trading against non-informed participants becomes a profitable strategy. This leads to funds that use AI to time the markets through statistical arbitrage, retail market making, countercyclical trading, or profiting off other investors' behaviour rather than buying undervalued stocks. AI, therefore, introduces the possibility of influencing picking and timing skills. Contrary to that, Begenau et al. (2018) suggest that data availability is mainly prevalent for large firms. As a result, price informativeness would only increase for large firms, leaving small firms out. Consequently, algorithms may overlook smaller firms, reducing their picking skill, or alternatively, leading to higher mispricing in these stocks, enabling AI funds to identify and exploit those, leading again to an increased picking activity.

This study proceeds as follows: We provide a concise overview of our datasets and the methodology for identifying AI-managed funds in Section 2, with a detailed explanation in Appendix A. Additionally, in Section 2, we outline how we match benchmark funds to each identified AI-labeled fund and compute the measures used to quantify the effects of AI in portfolio creation and optimization. The subsequent sections delve into our main analysis and present the results in Section 3 and Section 4. In Section 5 we revisit parts of the analysis and alter some of the specifications used in the two previous sections to test the robustness of our results. Finally, Section 6 summarizes our findings.

2 Data & Methodology

2.1 Data

For US mutual funds data, we rely on the CRSP Survivor-Bias-Free U.S. Mutual Fund Database. Stock prices, returns, and dividends are from the monthly CRSP files. Balance sheet and income statement data at the firm level are from COMPUSTAT. We obtain factor returns and the risk-free rate from Kenneth French's website[4]. Inflation is the change in the Consumer Price Index (All Urban Consumers)[5] from the U.S. Bureau of Labor Statistics. Data on the US business cycle is from the National Bureau of Economic Research (NBER)[6].

2.2 Identifying AI-managed funds

We rely on SEC filings of the funds at the CIK level, which we subsequently merge with the CRSP mutual fund database. We especially focus on SEC filing 497K. The Summary Prospectus 497K contains general information about the respective fund itself, such as e.g. the fee structure, assigned manager, a fund description, or an outline of its applied methodology. The latter one is the one we are interested in. We string-search a predefined library of keywords in the Summary Prospectus. For all matches, we read through the description of the fund and its described methodology and decide based on the provided information, whether the fund tries to enhance its performance through the help of AI. We only label a fund as an AI fund if we are certain that AI is employed[7]. In ambiguous cases, we abstain from adding the fund to our sample of AI funds. A more detailed description of the filters we apply, the keywords we use and other additional information in that regard is in Appendix A.

We, further, sub-classify AI funds into two sub-groups based on the stage of portfolio management at which the fund resorts to AI. The first group applies AI to screen the market for an investable universe or mispriced constituents. This can be, e.g. to discover all stocks that operate in a given

[4] `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

[5] Series Id: CUUR0000SA0

[6] `https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions`

[7] For instance, we found a fund that claims to employ "LASSO" in its portfolio management process. One would assume that they apply a "least absolute shrinkage and selection operator" (LASSO) regression to identify suitable predictors of future premiums, or similar. However, upon further investigation, we found that this fund uses the acronym "LASSO" in the sense of "Long and Short Strategic Opportunities". Obviously, we eliminate such "false positives" from our sample of AI funds.

industry to create an enhanced industry ETF. Another example would be to use AI to uncover all stocks with a certain kind of mispricing that is not explicitly visible from balance sheet data. Zhu (2019), for instance, finds that funds that incorporate such alternative data, for instance, satellite data or consumer transactions, invest and divest more efficiently. Similarly, Grennan and Michaely (2021) and Dugast and Foucault (2018) also suggest that such alternative data collected by fintechs or hedge funds may enhance price informativeness. This group is labelled *AI_in_stock_sel* in our results. The second sub-group includes AI in a later stage of portfolio construction. An example would be to use AI to enhance predictions about future returns: Chinco et al. (2019), for instance, find that their LASSO regressions select predominantly stocks that contain news about fundamentals, leading to improved out-of-sample return predictions and subsequently to higher Sharpe Ratios. We refer to this kind of fund as *quant_strat* in our results.

However, our approach to identifying AI funds also faces a major flaw. We may only find those funds that actively mention that they employ AI to enhance their portfolio. This might leave us with a sample that inherits a selection bias, which may be coined "AI-enhanced-*labelled*". Funds that claim to apply "statistical methods" or "quantitative analysis" could potentially also utilize AI, ML or DL in their investment process. However, as long as they do not explicitly mention this within their SEC filings, our methodology is not capable of distinguishing between funds that apply "statistical methods" with a simple OLS and those that, for instance, compute multiple-layered neural networks to update any priors on expected returns. Since this would constitute an ambiguous case, we drop this fund from the list of identified AI funds. As a result, we identify those funds that explicitly claim that AI is part of their portfolio management process. Our results might, thus, alternatively be interpreted as how the label "AI" influences returns or fund flows in addition to or rather than whether AI itself affects portfolio returns. We leave that decision to the reader and use the terms AI-managed and AI-labelled interchangeably.

2.3 Identifying benchmark funds

Besides the best possible identification of AI-optimized funds, the second crucial aspect of investigating realized AI-related returns is the selection of an appropriate counterfactual. In that regard, we adapt the methodology of Hoberg et al. (2018). The underlying idea is to define a set of fund characteristics and subsequently find "rival" funds that exhibit similar characteristics. The fund characteristics are derived from the individual characteristics of their respective holdings. Based on (orthogonalized) ranks and z-scores they identify rival funds as funds with the smallest Euclidean distance in an n-dimensional space, where n is the number of fund characteristics used. We alter this methodology into a two-step approach. First, we filter all funds that operate within the same fund style as provided by CRSP. Like this, we ensure we do not compare funds that follow different objectives. We do not consider fund style in the matching vector since the fund style is measured by a categorical variable, meaning that one cannot interpret distances between the categories. However, the methodology relies on those *metric* distances. In the second step, we search for similar funds based on fund properties.

For the main analysis, we compute z-scores based on the value-weighted stock characteristics in each fund. We winsorize all matching variables at the 1% and the 99% levels. We then orthogonalize the respective z-scores. This gives us the matching vector $C_{f,t} = \begin{pmatrix} z_{lme} & z_{lmom} & z_{ldy} \end{pmatrix}$, where $lme \equiv log(me)$, $lmom \equiv log(1 + r_{t-2,t-12})$ and ldy is the log dividend yield. We then adapt the cutoff distance $d^* = 8.858\%$ from Hoberg et al. (2018). However, according to them, the results are robust to the choice of d^*. We, furthermore, impose a minimum of ten rival funds to achieve a broader sample of comparable fund performance. This process identifies several benchmark funds for each of the previously identified AI-managed funds. We then proceed to aggregate these rival funds, in an attempt to create a portfolio of rival funds that better mimic the characteristics of a specific AI-enhanced fund. We employ two aggregations that differ only in the weights they assign to the rival funds, equal weights and distance weights. The latter uses the inverse Euclidean distance to the respective AI fund as weight. In such a distance-weighted portfolio rival funds which are more alike the AI fund receive a higher weight and vice versa.

2.4 Defining Measures of Skill and Activeness

In the model motivated by Pástor et al. (2020) the expected adjusted gross return is defined as the product of a fund-specific constant representing skill and a function that quantifies how actively that skill is measured. To measure the respective skill of an AI-enhanced fund or a rival fund we adapt the timing and picking skill defined by Kacperczyk et al. (2014) as

$$Timing_t^j = \sum_{i=1}^{N^j} \left(w_{i,t}^j - w_{i_t}^m \right) \left(\beta_{i,t} R_{t+1}^m \right) \tag{1}$$

and

$$Picking_t^j = \sum_{i=1}^{N^j} \left(w_{i,t}^j - w_{i_t}^m \right) \left(R_{t+1}^i - \beta_{i,t} R_{t+1}^m \right), \tag{2}$$

where N is the number of stocks that are in a fund's portfolio, j. $\left(\beta_{i,t} R_{t+1}^m\right)$ in the *Timing*-equation is the systematic part of the future returns. $\left(R_{t+1}^i - \beta_{i,t} R_{t+1}^m\right)$ in the *Picking*-equation on the other hand measures the idiosyncratic part of the future return of asset i within the portfolio. $\left(w_{i,t}^j - w_{i_t}^m\right)$ captures the over- or under-weighting that the portfolio has in a given stock compared with the market benchmark, meaning how much the portfolio exposes itself to the systematic or idiosyncratic part of the returns. This means that a fund with high ”timing” ability over- (under-)weights high (low) beta assets before market upswings and, vice versa, over- (under)weights low (high) beta assets before market downturns. Similarly, a fund with high ”Picking” ability, would build up positions in stocks with high idiosyncratic returns in advance of market increases and reduce them in sight of market downswings.

For the activeness measure we stick to the measure proposed by Pástor et al. (2020)[8]

$$Activeness = TL^{-1/2}, \tag{3}$$

where T is the turnover scaled by the NAV. $L = (\Sigma_{i=1}^N \frac{w_i^2}{m_i})^{-1}$ is portfolio liquidity where N is the number of stocks in a given portfolio, w_i and m_i are the respective weights of constituent i in the fund's portfolio and the market portfolio. Activeness is supposed to be positively related to turnover – to generate excess returns an active fund has to move away from the (market) benchmark – and negatively with L, as mispriced stocks tend to be less liquid and smaller.

[8]Equation 37

3 Performance

As we mention in the previous section, we identify 70 funds that apply artificial intelligence either for screening the investable universe to identify suitable fund constituents or in the subsequent stage for integrating selected stocks, bonds, etc., into a portfolio.

Figure 1 gives a first overview of the identified AI-managed funds. The left plot illustrates the distribution of the size (measured by Assets under Management (AuM)) of these funds. At the end of 2017, there was only one AI-managed fund, AIEQ. Subsequently, AIEQ experienced growth, coinciding with the emergence of numerous new funds over time, as depicted in the right plot. The solid grey line represents the total number of unique funds utilizing AI, while the dashed black line adjusts for share classes and funds sharing the same portfolio. Those new funds started with modest AuM[9]. Relative to these smaller funds, AIEQ's magnitude became significant, resulting in a notable spread between the 5th and 95th percentiles, particularly during the years 2020 to 2022. The issue with the smaller-sized funds is that overhead cost can be a substantial factor in the performance after fees, which we want to investigate. Additionally, we posit that funds may not accord adequate attention to portfolios below a certain threshold, potentially hindering the ability to generate outperformance. Consequently, for subsequent analysis, we opt to exclude all AI-labeled funds below a minimum size of five million, as below this threshold, net fund returns may not be deemed credible, with performance potentially not entirely attributable to fund skill. However, as a robustness check, we revisit the subsequent analysis, including smaller funds, in Section 5.

Our primary inquiry centres on whether AI usage furnishes fund managers with an advantage over counterparts eschewing artificial assistance. As an initial step, we examine the cumulative returns of such funds and compare them with their respective peers. Figure 2 displays those returns. We aggregate AI-managed funds into portfolios and compute the respective returns. The left plot delineates the aggregated returns over time (solid line) for an equally weighted portfolio (after deducting management fees and 12b fees), while the right plot does so in an AuM-weighted manner. With the dot-dashed line, we compare the returns of the AI portfolios with their matched non-AI peers. Similarly, the dotted line represents the market benchmark, defined as the equal- and value-weighed CRSP stock market portfolio. Examination of the size-weighted outcomes reveals that between

[9]The smallest AuM reported by CRSP is 0.1, indicating a fund size up to 100,000 USD.

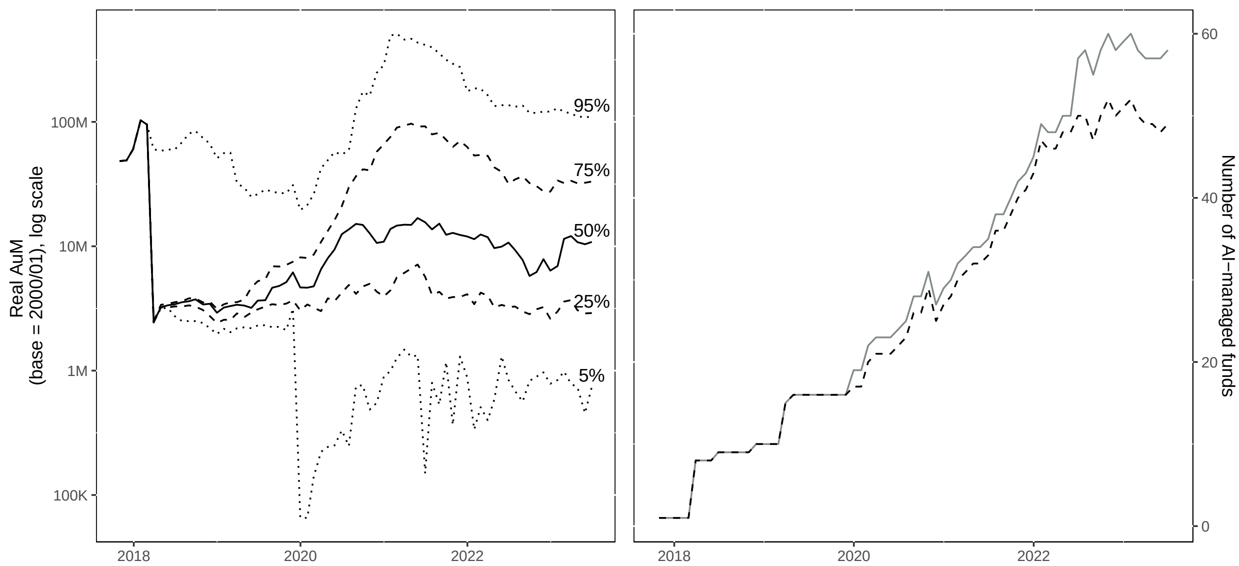

Figure 1: **Distribution AuM & Number of AI-managed funds/portfolios:** The left plot depicts the distribution of the Assets under Management (AuM) across all sub-classes of the AI-managed funds. The right plot shows the number of AI-managed funds (grey solid line) and unique portfolios (i.e. funds adjusted for different share classes of the same portfolio; black dashed line) over time. We further consider portfolios with different portfolio numbers identical if their time-series correlation of daily returns is above 99.9 % and if they are offered by the same company. For more extensive information on that regard we refer to Appendix A.

2020 and 2022, AI-managed funds potentially outperformed their fund and market benchmarks. However, over time, cumulative returns diminish, with performance falling below that of the stock market portfolio. Yet, the value-weighted portfolio of the AI-managed funds still yields a higher accumulated performance than the non-AI equivalents. Likewise, the equal-weighted portfolios exhibit a high relevance since the AIEQ makes up to 30% of the AuM of AI-managed portfolios at certain points in time. Again, the AI-managed funds surpass their conventional analogues. Additionally, the small AI-managed funds seem to outperform the equal-weighted stock market.

Table 1 delves deeper into these differences in (risk-adjusted) returns. Panel A and B report the average excess returns for equal and value-weighted AI and rival funds. Since we obtain several potential rivals for each AI fund, we aggregate them with equal weights (*ew*) and based on their relevance with the inverse of the Euclidean distance (*dw*) to the respective AI fund. One could interpret these portfolios as fictitious, AI-fund-style mimicking, non-AI counterfactual benchmark fund. We then aggregate these rival fund portfolios, representing one fund for each AI fund, into equal-weighted portfolios (Panel A) and value-weighted portfolios (Panel B). The average monthly

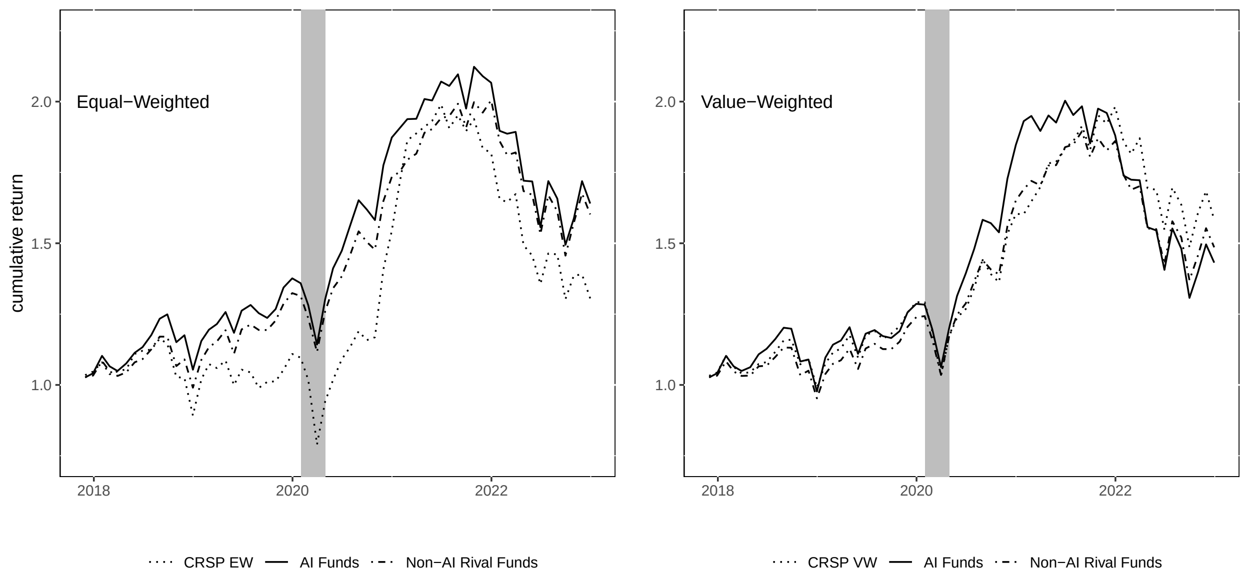

Figure 2: **Return over time:** We aggregate the AI-managed and the non-AI-managed funds into equal- and value-weighted portfolios and compute the respective cumulative returns. The left panel represents the equal-weighted aggregation, and the right is value-weighted. The rival funds were identified using orthogonal z-scores based on Hoberg et al. (2018). The dotted line represents the CRSP Equal-/Value-Weighted Stock Market. We require AI-labelled funds to have a minimum fund size of five million. NBER recessions are shaded in grey.

excess returns over the risk-free rate of the equal-weighted AI-labelled funds in Panel A amount to 71 basis points (bps). On average, AI funds even slightly underperform their peers. Equal-weighted rival funds yield an average monthly excess return 8.5 bps higher, while distance-weighted rival fund portfolios generate an average excess performance over AI funds of 9.4 bps. A similar pattern is observed in Panel B for the value-weighted aggregation of AI and rival funds, with a 32 bps average monthly excess return for AI funds and 69 bps and 68 bps for equal-/distance-weighted alternatives. However, a t-test indicates no significance for the average difference. Panel C further investigates the Sharpe Ratios (SR). Adjusted for risk, AI-labelled funds exhibit smaller returns compared to conventional funds. The SR of equally-weighted AI funds stands at 0.122 since 10/2017, while non-AI rivals exhibit SRs of 0.153 for equal-weighted rivals and 0.155 for distance-weighted ones. The inferior performance of the value-weighted AI fund portfolios in Panel B also results in smaller SRs. The risk-adjusted return is 0.051, half of what one would obtain with equal weights. Additionally, the SR difference is larger, with equal and distance-weighted peer funds yielding SRs more than twice as high as AI funds. A significance test of the SR differences following Ledoit and Wolf (2008) using the Parzen kernel fails to reject the hypotheses that the SRs differ for all combinations of equal- and value-weighted AI funds and equal- and distance-weighted connected peer funds. Our results align with the previous literature: Zhang et al. (2023) find similar results for the Chinese Mutual Fund market. Their studied "Big Data" funds have average Sharpe Ratios 0.123, far below the 0.208 of the conventional funds benchmark. The average monthly return among those big data funds is 0.70 %, whereas traditional funds average 1.28 % over the same period.

In Table 2 we explore potential alphas and the loadings of AI funds on factors, interpreted as fund styles, via several factor regressions based on monthly returns:

$$r_{j,t}^{AI} - r_t^f = \alpha_j + \sum_{k=1}^{K} \beta_{j,k} f_{k,t} + \varepsilon_{j,t}, \tag{4}$$

where $r_{j,t}^{AI}$ is the return of AI-labelled portfolio j in month t, r_t^f is the risk-free rate and r_t^m is the value-weighted market return of CRSP firms. $f_{k,t}$ is the return of zero-investment factor-mimicking portfolios constructed based on size, book-to-market, operating profitability, investment and past returns, respectively. We vary K from 1 to 6 and denote these variations as CAPM, FFC4 and FFC6, when $f_{k,t}$ is $\left\{\left(r_t^m - r_t^f\right)\right\}$, $\left\{\left(r_t^m - r_t^f\right) \quad SMB_t \quad HML_t \quad MOM_t\right\}$, and $\left\{\left(r_t^m - r_t^f\right) \quad SMB_t \quad HML_t \quad RMW_t \quad CMA_t \quad MOM_t\right\}$ respectively. Columns (3)-(5) and (8)-(10) give the results of the regressions for equal and value-weighted portfolios of AI-labelled fund returns. Most importantly, none of the alphas are significant across all columns, neither for equal-weighted nor value-weighted AI portfolios. This result blends in with the results of Zhang et al. (2023), who study big data funds in the Chinese mutual fund market. Their identified big data funds do not produce any improvements in alphas compared to traditional funds. The reported average alpha is significantly smaller than those of conventional funds. Columns (4), (9) and (10) reveal that AI-enhanced funds tilt themselves towards small-sized firms. Furthermore, columns (4) & (9) and (5) & (10) show that these funds tilt towards growth stocks and weak stocks with low operating profitability.

Additionally, we regress excess fund returns on excess rival returns, *ew* and *dw*,

$$r_{j,t}^{AI} - r_t^f = \alpha + \beta_j^{MKT} \left(r_{j,t}^{Rv} - r_t^f\right) + \varepsilon_t, \tag{5}$$

where $r_{j,t}^{Rv}$ is the aggregated return, *ew* or *dw*, of the rival funds of AI fund j at month t. The results are reported in Table 2, columns (1) & (2) and (6) & (7), respectively. The idea is to see whether AI funds can generate some alpha beyond conventional funds, in the sense that Jensen's alpha investigates an outperformance over the market return. Once again, the constants show no significance, indicating that AI funds fail to generate alpha beyond their matched rival funds. This finding mirrors those from Table 1.

Table 1: **Performance metrics:** We evaluate the performance of AI-enhanced funds compared to matched rival funds. Suitable rival funds are matched based on the methodology used in Hoberg et al. (2018). For each AI fund, we construct a portfolio comprising similar rival funds. Subsequently, these rival funds are aggregated into two distinct portfolios: one employing equal-weighted (*ew*) methodology, and the other utilizing distance-weighted (*dw*) methodology. In the *dw* approach, weights are determined based on the inverse of Euclidean distance, assigning higher weights to closer funds and lower weights to less similar ones within the rival portfolio. Panel A and B present the average *excess* returns and return differences for both AI and rival portfolios. We calculate these metrics using both equal-weighted and value-weighted portfolios of AI funds and their respective rival portfolios. We multiply the returns by 100. Returns, and consequently, the return differences in columns (4)-(6) are reported in per cent. Panel C displays the Sharpe Ratios (SR) and their differences. Columns (1)-(4) present SR for portfolios with equal weights, while columns (5)-(8) consider weights based on the total net assets (TNA) of the constituting funds. Standard Errors for SR differences are computed following Ledoit and Wolf (2008), based on the prewhitened Parzen kernel.

Panel A: Equal Weighted Returns

				$r^x_{AI} - r^x_{Rv}$			
	$\overline{r^x}$	se	p-val	Δr^x	se	p-val	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AI funds	0.714	0.745	0.338				62
Rival funds (ew)	0.799	0.662	0.227	−0.085	0.996	0.932	62
Rival funds (dw)	0.808	0.662	0.222	−0.094	0.996	0.925	62

Panel B: Value Weighted Returns

				$r^x_{AI} - r^x_{Rv}$			
	$\overline{r^x}$	se	p-val	Δr^x	se	p-val	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AI funds	0.320	0.794	0.686				62
Rival funds (ew)	0.691	0.667	0.300	−0.371	1.037	0.721	62
Rival funds (dw)	0.680	0.669	0.310	−0.360	1.038	0.730	62

Panel C: Sharpe Ratios

	Equal Weighted Portfolios				Value Weighted Portfolios			
	SR	ΔSR	se	pval	SR	ΔSR	se	pval
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AI	0.122				0.051			
Rivals (ew)	0.153	−0.032	0.031	0.300	0.132	−0.080	0.050	0.108
Rivals (dw)	0.155	−0.033	0.031	0.276	0.129	−0.078	0.048	0.103

Table 2: **Fama/French/Carhart Factor Regressions:** Columns (1)&(2) and (5)&(6) report the coefficients obtained from regressing the excess returns of AI funds onto the excess returns of equal-weighted and value-weighted portfolios of Rival funds. Rival funds are identified using the methodology outlined in Hoberg et al., 2018. For each AI fund, we select a minimum of ten similar funds and aggregate them into a portfolio designed to replicate a counterpart fund with comparable characteristics to the AI funds. This aggregation employs both equal weights (*ew*) and distance weights (*dw*), where we compute weights for *dw* with the inverse of the Euclidean distance of each rival fund. In columns (3)-(5) and (8)-(10), we follow the conventional approach of factor regressions, reporting the coefficients obtained from regressing the equal-weighted and value-weighted AI returns onto the factor returns. Columns (1)-(5) employ equal weights on AI fund returns, while columns (6)-(10) use fund size (TNA) as weights.

	Dependent variable:									
	AI fund returns (ew)					AI fund returns (vw)				
	Rivals (ew)	Rivals (dw)	CAPM	FFC4	FFC6	Rivals (ew)	Rivals (dw)	CAPM	FFC4	FFC6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
rvex	1.089***					1.116***				
	(0.062)					(0.068)				
rvvx		1.089***					1.117***			
		(0.060)					(0.064)			
mkt			1.026***	0.989***	1.026***			1.053***	0.996***	1.053***
			(0.065)	(0.076)	(0.077)			(0.073)	(0.078)	(0.049)
smb				0.181***	0.056				0.397***	0.194*
				(0.069)	(0.087)				(0.117)	(0.102)
hml				−0.219***	−0.187***				−0.311***	−0.239***
				(0.043)	(0.069)				(0.055)	(0.066)
rmw					−0.267***					−0.422***
					(0.075)					(0.101)
cma					0.041					0.020
					(0.091)					(0.071)
mom				−0.021	−0.046				0.026	−0.007
				(0.053)	(0.052)				(0.070)	(0.071)
Constant	−0.157	−0.166	−0.017	0.008	0.100	−0.452	−0.439	−0.423	−0.379	−0.213
	(0.173)	(0.181)	(0.272)	(0.229)	(0.176)	(0.337)	(0.320)	(0.390)	(0.284)	(0.216)
Observations	62	62	64	64	64	62	62	64	64	64
R^2	0.936	0.936	0.905	0.931	0.939	0.879	0.887	0.842	0.901	0.917
Adjusted R^2	0.935	0.935	0.903	0.927	0.932	0.877	0.885	0.840	0.894	0.909

Note: Newey and West (1994) Standard Errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01

Given the absence of substantial return differences between AI-enhanced funds and the stock market or other funds, as well as the lack of significant alphas, we delve deeper into whether AI may be more beneficial at the first or second stage of portfolio management. Figure 3, thus, disentangles the returns reported in Figure 2 into funds that try to exert the power of AI in identifying the relevant constituents to invest in, and those that employ AI to arrive at a weighting for the previously selected stocks, bonds, etc. We tag the first group "*Selection*" while we refer to the latter as "*Allocation*". Again, the two solid lines represent the performance of the respective AI-optimized portfolios in the *Selection* and the *Allocation* group. The dot-dashed lines represent the matched rival fund returns. The dotted lines again compare the funds to the aggregated stock market. The left plot illustrates returns when all funds have identical weights in their respective portfolios, whereas the right plot weights funds based on their AuM. Only the *Selection* sub-portfolio with equal weights manages to outpace the market benchmark. The *Allocation* AI funds perform even worse than the sub-portfolio of their rivals. In the value-weighted case, both AI subgroups of funds do not manage to surpass the stock market. Regardless of that underperformance compared to the aggregate stock market, these results hint that AI seems to help its fund managers find better constituents for their portfolios: The AI-*Selection* portfolios yield higher cumulated returns than the conventional benchmark funds, especially in the equal-weighted case that gives less weight to the AIEQ. In addition, the *Selection* funds accomplish considerably higher returns than the funds that employ AI in the portfolio optimization stage. These *Allocation* funds also do not exhibit considerable excess returns compared to their peers.

Figure 3 suggests that AI might play a beneficial role in pinpointing potentially profitable assets for portfolio construction. To further explore this we study whether those AI funds demonstrate skill and effectively utilise it. Pástor et al. (2020) posit that a fund's excess return, $a = \mu g\,(T, L)$, comprises a fund-specific constant (μ) representing its skill level, and a measure (g) indicating how actively the fund applies its skill. Activeness is gauged by turnover (T) and a liquidity/diversification metric (L). Activeness increases in turnover T. The rationale is that an active fund has to trade away from the benchmark to generate returns over its benchmark. Activeness is negatively related to liquidity L, because mispricing is greater among less-liquid and smaller stocks. We recompute the fund activeness measure for the funds and then regress fund activeness on the Dummy variable for AI funds, including time-fixed effects. The results are in columns (1) and (2) of Table 3. AI-labelled funds tend to be less active than their peers, a trend also observed for funds employing AI

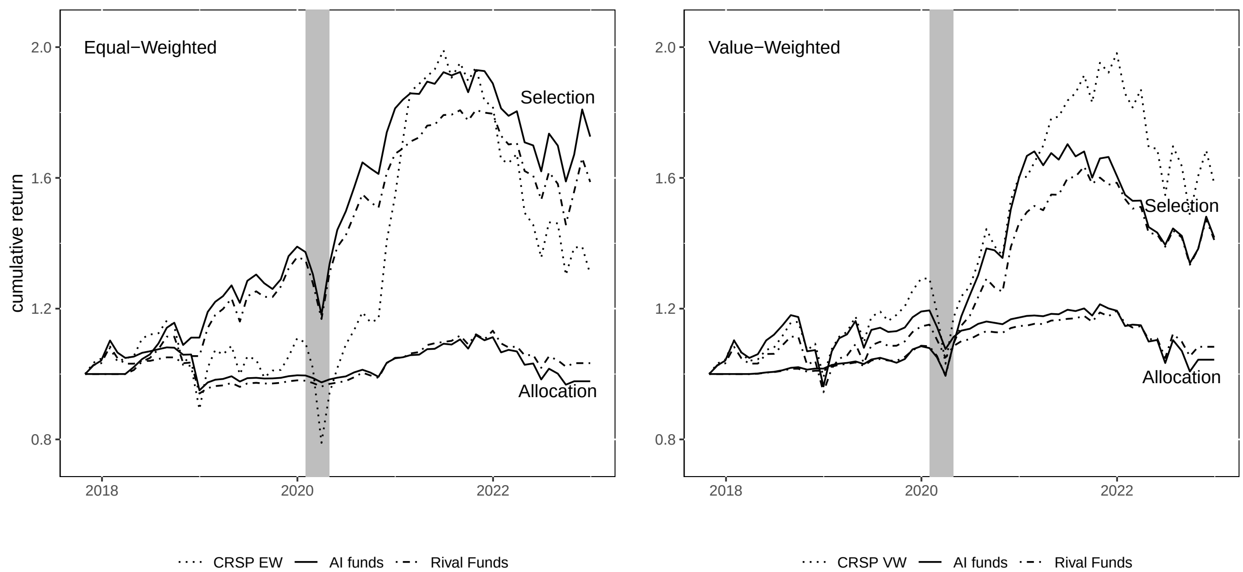

Figure 3: **AI to screen the universe and allocate portfolios:** We sort AI-labelled funds into two categories: The first uses AI to screen the market and establish an investment universe of potentially profitable constituents. We label this group *Selection*. The second group of funds employs AI to arrive at desired portfolio weights. This group has the tag *Allocation*. The solid lines are the cumulative returns of an equal- and value-weighted portfolio of those groups of AI funds. The dashed lines are the cumulative returns of the portfolio of benchmark funds which we again match using orthogonal z-scores based on Hoberg et al. (2018). The left plot depicts the returns, if all funds have the same weights in their respective portfolios. The right plot weights the funds based on their AuM. Further information on how we select funds and how we find appropriate benchmark funds is given in Appendix A. US recessions as dated by the NBER are highlighted in grey.

in the selection process. In contrast, funds utilizing variations of AI in asset weighting tend to be more active than conventional funds.

Furthermore, we assess the general skill level of the AI and Non-AI funds using the two measures motivated in Kacperczyk et al. (2014), *Timing* and *Picking*. We explain the measures in Section 2. Timing evaluates how well a portfolio times its holdings (with over-/underweighting) to the systematic component of returns. Picking captures the idiosyncratic part of the portfolio's holdings. We compute the betas of the holdings using rolling windows of the previous 12 months. We adopted a short model where the only explanatory variable is whether or not a fund uses (or advertises to use) AI and additionally, we added some further covariates. Contrary to other parts of the previous analysis, we do not from *ew* or *dw* portfolios of the set of rival portfolios, but consider them as they are. We print the results of these regressions in columns (4)-(14) in Table 3. Across all specifications, AI funds exhibit significantly better timing abilities than their conventional rival funds. Their over/underweighting creates additional returns between 45 bps in the short regression and 28 bps in the long model without the matching variables, compared to their non-AI peers. Conversely, AI funds significantly underperform their rivals in terms of Picking. Their inferior picking ability creates negative returns of 58 bps for overall AI-labelled funds in the short model to 34 bps in the model that includes further explanatory variables without the matching variables. Separating AI-labelled funds into subgroups based on the presumed stage of AI application yields similar results. In the short model, the AI portfolios have relatively higher timing ability worth 48 and 38 bps in extra return units. This timing return is, nevertheless, offset by the comparably worse picking ability of approximately the same size, producing returns of -59 bps and -55 bps on average. These results continue to hold if we again add further explanatory variables, though the effect size is smaller. Timing is positive for the subgroups of AI portfolios with 25 and 41 bps of return while picking cancels these returns out with -36 and -43 bps, relative to the matched rivals. These results are opposite to those reported in Chen and Ren (2022), which find relative picking ability and no significant difference in timing ability. However, the significance of the picking ability is low, indicating that extending the time series and cross-section with new AI-managed funds may improve the power of tests, providing a clearer picture of AI-labelled fund skill levels.

Table 3: **Activeness and Skill:** We run Fixed Effects Regressions with fixed time effects to examine the impact of AI funds on activeness and skill. We employ a dummy variable indicating the use of AI at any stage of portfolio management as the explanatory variable. Additionally, we disaggregate this AI dummy into two distinct stages: AI employed in the stock selection process and AI utilized in determining the weighting scheme. The *Activeness* measure, as proposed in Pástor et al. (2020), is defined as $Activeness = TL^{-1/2}$, where T represents fund turnover and L denotes a measure of portfolio liquidity. Skill is assessed through *Timing* and *Picking* metrics as outlined in Kacperczyk et al. (2014): $Timing_t^j = \sum_{i=1}^{N^j} \left(w_{i,t}^j - w_{i_t}^m\right)\left(\beta_{i,t} R_{t+1}^m\right)$ and $Picking_t^j = \sum_{i=1}^{N^j} \left(w_{i,t}^j - w_{i_t}^m\right)\left(R_{t+1}^i - \beta_{i,t} R_{t+1}^m\right)$. For the abbreviated models, we include only a dummy variable for AI as the explanatory factor. We incorporate two sets of additional covariates: one where matching variables used to identify rival funds are excluded, and another where these variables are included. Standard Errors are clustered following Arellano (1987).

	Dependent variable:													
	Fund Activeness		Timing						Picking					
	Short Model	Short Model	Short Model	Short Model	Long w/o MV	Long w/o MV	Long Model	Long Model	Short Model	Short Model	Long w/o MV	Long w/o MV	Long Model	Long Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
ai_fund	-5.4700^{***}		0.0045^{***}		0.0028^{***}		0.0033^{***}		-0.0058^{***}		-0.0034^{*}		-0.0038^{**}	
	(2.1005)		(0.0007)		(0.0009)		(0.0009)		(0.0015)		(0.0018)		(0.0016)	
ai_in_stock_sel		-9.1556^{***}		0.0048^{***}		0.0025^{**}		0.0028^{***}		-0.0059^{***}		-0.0034		-0.0036^{*}
		(1.6843)		(0.0009)		(0.0011)		(0.0010)		(0.0020)		(0.0022)		(0.0020)
quant_strat		7.9256^{*}		0.0038^{***}		0.0041^{***}		0.0050^{***}		-0.0055^{***}		-0.0033^{*}		-0.0043^{**}
		(4.0933)		(0.0010)		(0.0015)		(0.0015)		(0.0015)		(0.0020)		(0.0021)
port_lme							-0.0024^{***}	-0.0024^{***}					0.0018^{***}	0.0018^{***}
							(0.0001)	(0.0001)					(0.0001)	(0.0001)
port_lbm					-0.0041^{***}	-0.0041^{***}	-0.0065^{***}	-0.0065^{***}			0.0008^{***}	0.0008^{***}	0.0026^{***}	0.0026^{***}
					(0.0003)	(0.0003)	(0.0003)	(0.0003)			(0.0002)	(0.0002)	(0.0002)	(0.0002)
port_lmom							-0.0007	-0.0007					0.0037^{***}	0.0037^{***}
							(0.0008)	(0.0008)					(0.0007)	(0.0007)
port_ldy							0.0010^{***}	0.0010^{***}					-0.0003^{*}	-0.0003^{*}
							(0.0002)	(0.0002)					(0.0002)	(0.0002)
lage					0.0006^{***}	0.0006^{***}	0.0007^{***}	0.0007^{***}			0.00002	0.00002	-0.0001	-0.0001
					(0.0001)	(0.0001)	(0.0001)	(0.0001)			(0.0001)	(0.0001)	(0.0001)	(0.0001)
exp_ratio					0.0157	0.0156	-0.0068	-0.0069			-0.0345^{*}	-0.0346^{*}	-0.0103	-0.0103
					(0.0262)	(0.0262)	(0.0247)	(0.0247)			(0.0180)	(0.0180)	(0.0170)	(0.0170)
turn_ratio					-0.0001	-0.0001	-0.0001	-0.0001			-0.0001	-0.0001	-0.0001	-0.0001
					(0.0001)	(0.0001)	(0.0001)	(0.0001)			(0.00005)	(0.00005)	(0.0001)	(0.0001)
Observations	151,456	151,456	485,137	485,137	217,707	217,707	212,512	212,512	485,137	485,137	217,707	217,707	212,512	212,512
R^2	0.00005	0.0001	0.0001	0.0001	0.0037	0.0037	0.0151	0.0151	0.0002	0.0002	0.0003	0.0003	0.0083	0.0083
Adjusted R^2	-0.0002	-0.0002	-0.0001	-0.0001	0.0034	0.0034	0.0147	0.0147	0.0001	0.0001	-0.00003	-0.00004	0.0080	0.0080

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

4 Persistence and Fund Flows

We follow the classical approach put forward by Carhart (1997) to asses the persistence of AI returns. Given the limited number of AI funds, we categorize them into terciles based on the previous month's returns of all funds (including also those funds that are neither AI funds nor rival funds), as opposed to deciles. We then hold the sorts constant for the next period (month, quarter or semester[10]) and compute the returns of equally weighted portfolios of each tercile. This process is repeated, rolling forward until the beginning of the next holding period, yielding a time series of non-overlapping returns for tercile portfolios. Additionally, we evaluate the performance of a zero-investment portfolio that takes long positions in the top-performing tercile and short positions in the bottom-performing tercile. We assess the persistence of the equally weighted tercile portfolios by regressing their excess returns onto K factors:

$$r_{tc,t} - r_t^f = \alpha_{tc} + \sum_{k=1}^{K} \beta_{tc,k} f_{k,t} + \varepsilon_{tc,t}, \tag{6}$$

where $r_{tc,t}$ is the return of the tercile portfolio or a zero-investment portfolio constructed solely with AI funds or rival funds, respectively and r_t^f is the risk-free rate. Similar to Equation 4, we vary the number of factor-mimicking portfolios, K, in $f_{t,k}$.

Busse et al. (2010) conducted similar factor regressions for all U.S. funds and identfied a reversal in the performance of poorly-performing funds and momentum in well-performing funds. Panel A, B and C in Table 4 report the intercepts from the corresponding CAPM, four-factor (FFC4) and six-factor (FFC6) regressions. Columns (1)-(3) are the alphas from sub-sampling solely on AI funds, and columns (4)-(6) filter on rival funds. Throughout both groups, AI and rival funds, and for all variations of considered factors, we observe that the bottom and mid-performing funds yield subsequent significant negative alphas. This finding also holds for all investigated holding periods. This means that, contrary to the results in the overall fund sample of Busse et al. (2010), bad performance on average persists for at least the following six months.

Concerning the top tercile, previous findings of Carhart (1997) and Busse et al. (2010) suggest positive momentum in fund returns. However, for AI funds we only observe positive future semester

[10]We refrain from investigating persistence over the next year or successive years due to the constraints of our sample, as outlined in Section 2. Our sample is restricted by the inception of the first AI fund in 2017, resulting in a limited number of observations. In the case of FFC6, this would entail $K > T$.

alphas within the FFC4 regressions and positive subsequent monthly alphas with six factors. The remaining intercepts of AI-labelled funds are not significant. For the rival funds, we even observe the opposite. The significant intercepts, all at 1Q and 1S horizons exhibit a negative sign, indicating short-term reversal rather than momentum.

In the model of Farboodi and Veldkamp (2020), higher computational power enables funds to process more information simultaneously, enhancing price informativeness. However, this can lead to diminishing returns as all market participants react to the same information, resulting in a scenario where assets are bought uniformly. As a result, the increasing amount of inflows towards these assets will create upward price pressures towards them. Similarly, Gabaix and Koijen (2022) find that price elasticities in the equity market are small, indicating that flows in and out of the market exert significant price pressure on stocks experiencing substantial flows. The magnitude of this effect is noteworthy, with a $ 1 inflow generating a five-fold impact on the aggregate market. Additionally, findings by Pástor et al. (2021) and Pástor et al. (2022) argue that although green firms typically have low expected returns, they tend to outperform following positive demand for the ESG factor. Moreover, Beck (2021) examines ESG returns at the fund level. Even though ESG funds should not outperform the market, the flow-driven return of such funds is 2.07%. Therefore, we hypothesize that the inelasticity of demand could - analogous to ESG funds - be viable for AI-labelled funds.

To explore this further, we examine the impact of inflows and outflows on the identified AI funds. We suspect that this effect is particularly relevant for larger funds, given their greater visibility to public investors compared to smaller funds. Figure 4 plots the cumulative returns of the top 50 % of AI-labelled funds. The first prominent observation is that the aggregate returns are much larger for the big funds, compared to the cumulative returns in Figure 2. Particularly, the equal-weighted portfolio (shown in the left plot with a solid line) yields higher returns than the equal-weighted market portfolio. This might The poorer performance of the value-weighted equivalent is attributable to the performance of the AIEQ, which at some points constitutes up to 30 % of the aggregate.

The availability of data on fund flows is scarce in the CRSP dataset. In addition, we due to the change in the filing system, the numbers are not comparable. To obtain more reliable data, we proxy fund flows via the change in TNA, corrected for internal portfolio growth and dividends:

Table 4: **Persistence of alphas**: We asses the persistence of returns following Carhart (1997): At the beginning of each month, quarter or semester, we sort each fund portfolio into terciles based on their past-month return. We then aggregate the returns to equally-weighted tercile portfolios. Additionally, we compute returns for a portfolio with a long position in the top tercile and a short position in the bottom portfolio. We re-balance the portfolios at the start of each holding period, meaning monthly, quarterly or half-annually. The table reports the alphas from factor regressions on these portfolio returns. 1M - next month, 1Q - next quarter, 1S - next semester. Newey and West (1994) Standard Errors are in parentheses.

Panel A: CAPM

	AI			Rivals		
	1M	1Q	1S	1M	1Q	1S
	(1)	(2)	(3)	(4)	(5)	(6)
1	-0.0091***	-0.0151**	-0.0317	-0.0041**	-0.0045	-0.0327***
	(0.0032)	(0.0075)	(0.0289)	(0.0019)	(0.0051)	(0.0074)
2	-0.0043	-0.0237***	-0.0247**	-0.0019*	-0.0098***	-0.0188***
	(0.0027)	(0.0055)	(0.0125)	(0.0011)	(0.0021)	(0.0038)
3	0.0024	-0.008	0.0092	-0.0015	-0.0108***	-0.0071
	(0.0034)	(0.0074)	(0.0198)	(0.0021)	(0.0037)	(0.009)
3-1	0.0105***	0.0044	0.0424	0.0016	-0.0098	0.0276*
	(0.004)	(0.0122)	(0.0313)	(0.0035)	(0.0075)	(0.0156)

Panel B: FFC4

	AI			Rivals		
	1M	1Q	1S	1M	1Q	1S
	(1)	(2)	(3)	(4)	(5)	(6)
1	-0.0083***	-0.0084	-0.0178	-0.0039**	-0.0028	-0.0497***
	(0.0029)	(0.0056)	(0.0195)	(0.002)	(0.0046)	(0.0044)
2	-0.0039	-0.0245***	-0.0135*	-0.0017**	-0.0105**	-0.0161***
	(0.0028)	(0.0066)	(0.008)	-8,00E-04	(0.0053)	(0.0034)
3	0.0027	-0.0048	0.0579***	-0.0013	-0.0124**	0.0013
	(0.0028)	(0.0064)	(0.0213)	(0.0025)	(0.0054)	(0.008)
3-1	0.01**	0.0019	0.081**	0.0017	-0.013***	0.0599***
	(0.0044)	(0.009)	(0.0351)	(0.004)	(0.0043)	(0.0131)

Panel C: FFC6

	AI			Rivals		
	1M	1Q	1S	1M	1Q	1S
	(1)	(2)	(3)	(4)	(5)	(6)
1	-0.0071**	-0.0115*	0.0316***	-0.0035*	-0.0052	-0.016
	(0.0029)	(0.0061)	(0.01)	(0.0019)	(0.0067)	(0.0163)
2	-0.0041	-0.0276***	0.0129***	-0.0021***	-0.0112***	-0.0098***
	(0.0032)	(0.0037)	(0.0024)	-8,00E-04	(0.0025)	(0.0014)
3	0.0031	-0.0048	-0.003	-0.0022	-0.0107**	-0.0297***
	(0.0028)	(0.0063)	(0.0049)	(0.0025)	(0.0045)	(0.0055)
3-1	0.0092**	0.0048	-0.0335	4,00E-04	-0.0089***	-0.0062
	(0.0042)	(0.0111)	(0.031)	(0.0038)	(0.003)	(0.0714)

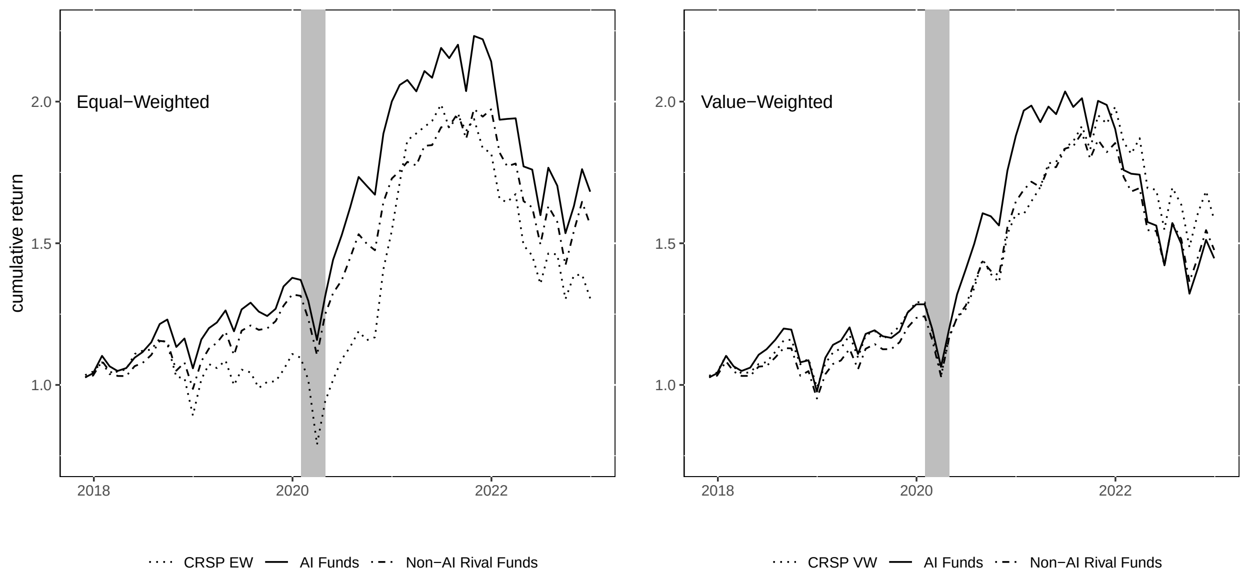

Figure 4: **Top 50 % (TNA) AI funds, cumulative returns:** We compute the cumulative returns for AI-labelled funds, same as for Figure 2 for the funds with the largest 50 % (measured in TNA) AI-labelled funds. The solid line depicts the returns of an equal and value-weighted portfolio of AI funds respectively. The dash-dotted line is the portfolio of rival funds, matched adapting the methodology in Hoberg et al. (2018). A detailed description is in Section 2. The dotted line represents the cumulative returns of CRSP equity market portfolios. The grey shaded area are recessions as reported by NBER.

$$flow_{j,t}^{abs} = NAV_{j,t} - NAV_{j,t-1} * (1 + r_{j,t}) \tag{7a}$$

$$flow_{j,t}^{rel} = \frac{NAV_{j,t} - NAV_{j,t-1} * (1 + r_{j,t})}{NAV_{j,t-1}}, \tag{7b}$$

where $NAV_{j,t}$ is the aggregated net asset value of portfolio j at time t and $(1 + r_{j,t})$ denotes the investment-weighted portfolio returns of portfolio j from time $t-1$ to t. Berk and Tonks (2007) argue that Equation 7b does not correctly capture the relative in- or outflow towards funds. The authors argue that this computation does not perfectly control the change of flows for internal growth. Alternatively, they propose the following measure for relative fund flows

$$flow_{j,t}^{rel,BT07} = \frac{NAV_{j,t} - NAV_{j,t-1} * (1 + r_{j,t})}{NAV_{j,t-1} * (1 + r_{j,t})} \tag{7c}$$

To analyse the implications of fund flows into AI and rival funds, we conduct fixed effects regressions with three types of regressors: a short model with a dummy variable for AI funds as the sole explanatory variable, one resembling the specifications of Miguel and Chen (2021), and a longer one where we include additional explanatory variables. The corresponding results are printed in Table 5. Contrary to Miguel and Chen (2021), we do not find robust evidence for a strong sensitivity of AI or rival fund flows to past returns. Although past-month performance exhibits significance and a positive coefficient in the long models (columns (3), (6), and (9)), the dummies indicating whether a portfolio was in the top or bottom tercile of fund returns remain largely insignificant across most model specifications. Only in column (2) does the dummy representing past performance in the lowest tercile show significance, albeit with a positive coefficient. This finding appears counterintuitive, especially considering the positive and significant past four-factor alpha. Such a result would imply that poorly performing funds (in the bottom tercile) and well-performing ones (with higher alphas) receive positive absolute inflows, which seems paradoxical. In favour of the idea would on the other hand be that both, past timing and past picking skill, are significant and positive for all three measurements of fund flows. Moreover, for absolute flows, past alpha is significant, positive and high for both specifications in which it is considered. Nonetheless, the AI dummy variable proves to be irrelevant in all but one model, indicating that the utilization of AI is not a determinant for future fund flows. Consequently, the results suggest that AI funds are not as strongly influenced by fund flows as ESG funds.

5 Further Analyses and Robustness

So far, we have restricted our analysis by several assumptions. However, these design choices might be influential for the subsequent results. Below, we, therefore, investigate the robustness of our results towards some alternations in the analysis.

5.1 Keep tiny funds

For the main analysis, we removed all funds with a fund size below five million. Appendix B repeats the analysis including those tiny funds and presents the results. We find that the results are similar to the main analysis, though AI-labelled funds now perform slightly worse but still not significantly worse than their non-AI peers.

Table 5: **Fund flows**: We run Fixed Effects Regressions with fixed time effects to investigate flows towards funds. In the short model, we only include a dummy for AI funds. In the second specification, we further the variables used in Miguel and Chen (2021): That includes two further dummy variables that are one if the past performance was in the top or bottom tercile of all funds respectively. Furthermore, we include past performance, past alphas from a four-factor regression and the prevalent standard deviation of the fund-specific flow computed with rolling regressions. The last specification adds additional fund-specific control variables. In the table, we abbreviate Berk and Tonks (2007) with BT07. Standard Errors are clustered following Arellano (1987).

	Dependent variable:								
	fund flows (abs)			fund flows (rel)			fund flows (rel) BT07		
	Short Model	Middle Model	Long Model	Short Model	Middle Model	Long Model	Short Model	Middle Model	Long Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ai_fund	0.385	0.324	−17.023***	0.007	−0.004	0.008	0.006	−0.005	0.005
	(0.757)	(0.834)	(4.201)	(0.011)	(0.009)	(0.014)	(0.011)	(0.009)	(0.013)
d_bot_perf_l1		4.716***	−4.647		0.042	−0.001		0.041	−0.001
		(1.698)	(5.233)		(0.037)	(0.005)		(0.036)	(0.004)
mretal1		8.980	113.176***		−0.428	0.216***		−0.450	0.223***
		(12.966)	(22.120)		(0.441)	(0.071)		(0.436)	(0.072)
d_top_perf_l1		−0.205	2.247		−0.023	0.002		−0.023	0.002
		(1.420)	(4.818)		(0.032)	(0.004)		(0.032)	(0.004)
alphal1		156.294**	670.653***		1.019	0.290		1.039	0.241
		(61.867)	(158.604)		(1.103)	(0.201)		(1.089)	(0.206)
sd_flow_abs_l1			−0.101***						
			(0.018)						
sd_flow_rel_l1						0.001			
						(0.002)			
sd_flow_rel_BT_l1									0.001
									(0.002)
fa_l1			0.058**			0.00003			0.00003
			(0.026)			(0.00003)			(0.00003)
tmg_l1			135.743**			0.150***			0.146***
			(54.078)			(0.055)			(0.054)
pkg_l1			201.034***			0.355***			0.331***
			(39.758)			(0.068)			(0.066)
lage_l1			−4.907			−0.013***			−0.013***
			(3.354)			(0.002)			(0.002)
exp_ratio_l1			−3,071.852***			−0.180			−0.208
			(992.097)			(0.337)			(0.332)
turn_ratio_l1			0.110			0.003			0.003
			(0.222)			(0.002)			(0.002)
port_lme_l1			1.326			−0.001			−0.0005
			(1.378)			(0.002)			(0.002)
port_lbm_l1			1.693			0.005			0.004
			(3.233)			(0.004)			(0.004)
port_lmom_l1			1.986			0.041**			0.040**
			(6.178)			(0.017)			(0.017)
port_ldy_l1			1.256			−0.005			−0.005
			(1.946)			(0.003)			(0.003)
Observations	732,794	655,206	130,422	732,794	655,206	130,422	732,794	655,206	130,422
R^2	0.000	0.00001	0.017	0.000	0.00000	0.001	0.000	0.00001	0.001
Adjusted R^2	−0.0001	−0.0001	0.017	−0.0001	−0.0001	0.001	−0.0001	−0.0001	0.001
Note:									*p<0.1; **p<0.05; ***p<0.01

5.2 Drop orthogonalization and z-scores

The analysis' results rely heavily on the choice of the rival funds. Especially the orthogonalization of the z-scores puts a lot of emphasis on the order of the variables in the characteristics vector $C_{f,t}$. Matching variables that are placed in the first elements of the vector receive much higher importance than those which are in the latter elements of the matching vector. We, therefore, re-match the rival funds based on z-score without orthogonalization. We abbreviate these results with *zs*. Alternatively, instead of using z-scores, we use the ranks in the $C_{f,t}$ vector. We construct ranks, following the first matching methodology of Hoberg et al. (2018), as the percentile of a stock's characteristic in the distribution of all NYSE stocks with share code 11 or 12. We then aggregate the ranks of all stock holdings at the portfolio level using the relative portfolio weights of the constituents. *pcrk* flag all results based on these percentile ranks throughout this paper.

Figure 5 plots the analogous returns to Figure 2 but with alternative rival funds, matched without orthogonalizing either ranks or z-scores. Panel A shows results computed based on ranking the variables. Panel B depicts the respective plot based on z-scores instead. The solid lines represent the funds relying on AI, whereas the dash-dotted lines depict the non-AI peers. A grey line indicates an aggregation where each portfolio receives an equal weight, a black line on the other hand aggregates relative to the value managed within the respective portfolio. Comparing the results between ranks and z-scores, we see no substantial differences. The graphs paint a very similar image. Similar to the rival funds from the main analysis, the AI-managed funds perform better than their peers in the beginning. However, depending on the set of matching variables the performance of the AI-labelled funds deteriorates in the second half of the observation period. In some cases even substantially. We, further, observe that the result varies across the set of variables used to match the funds. Therefore, we conclude that while z-scores or ranks do not greatly affect the set of rivals, the orthogonalization and the set of variables in the matching vector might play a role in matching though.

5.3 Alternative variables

From the results in Figure 5 we concluded that the set of variables in the matching vector might play a substantial role in finding benchmark funds for the investigated AI funds. One could, for in-

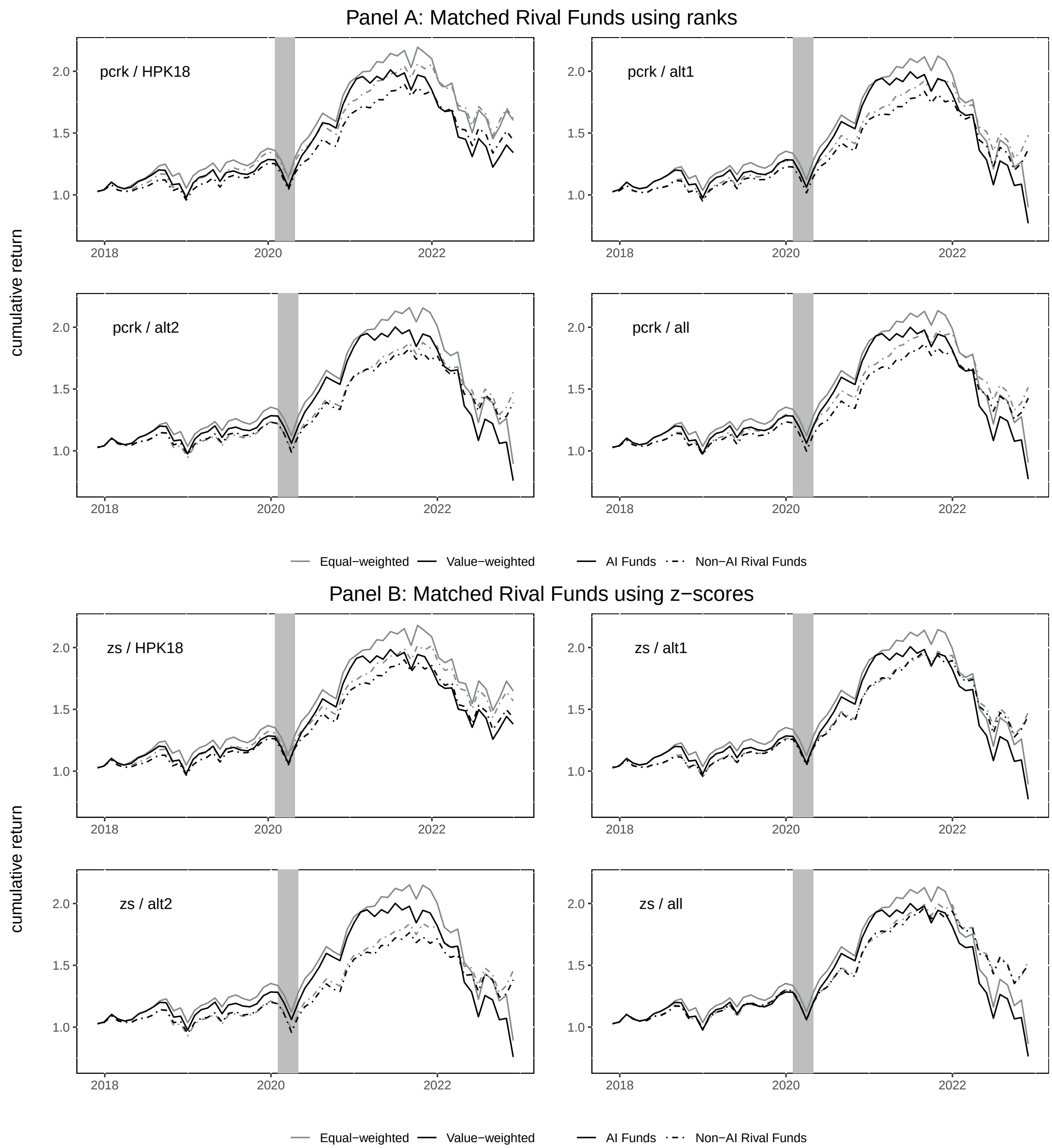

Figure 5: **Return over time using alternative variables to match rival funds:** We use the same methodology as previously to match (portfolios of) rival funds. However, this time we drop orthogonalization and use both, ranks (*pcrk*) and z-scores (*zs*). Furthermore, we use alternative variables in $C_{f,t}$. For *HKP18*, we adopt $C_{f,t}^{HPK18} = \big(lme \quad lbm \quad lmom \quad ldy\big)$, for *alt1*, $C_{f,t}^{alt1} = \big(lme \quad fa \quad lbm\big)$, *alt2* uses $C_{f,t}^{alt2} = \big(lage \quad ltnr \quad lep\big)$, and *all* combines all previous variables and adds one extra $C_{f,t}^{all} = \big(lme \quad lbm \quad lmom \quad ldy \quad lage \quad fa \quad lep \quad ltnr \quad lexr\big)$, where lme is the log size, lbm is log book-to-market, $lmom$ is log momentum, ldy is log dividend yield, $lage$ is the log fund age in years, fa represents a measure of fund activeness, lep is log earings/price, $ltnr$ is the log turnover ratio, and $lexr$ is the log of reported expense ratio. Grey lines indicate equal weights in portfolios, black lines use market weights. Solid lines represent AI funds, dot-dashed lines depict rival funds. Periods of recessions (as reported by NBER) are shaded in grey.

stance, argue that momentum as a matching variable may lead to a self-fulfilling prophecy of similar subsequent fund returns. We, therefore, alter the variables in the $C_{f,t}$ vector. *HKP18* adopts the variables used in the 4D network of Hoberg et al. (2018), namely log size, log book-to-market ratio, log past returns and log dividend yield. *alt1* picks up the argument from before and replaces the momentum characteristic from the main analysis with fund activeness. *alt2* tries to find alternatives for all three variables from the main analysis. Fund age replaces size, as longer existing funds should also have a higher TNA. Log turnover poses the alternative to momentum, following the idea of fund activeness in Pástor et al. (2020) that one has to trade away from a given benchmark to create outperformance, leading to positive momentum. The log earings-to-price ratio is used as an alternative multiple for valuation. Furthermore, earnings should sooner or later transform into cash dividends, which would lead us back to the dividend yield. Finally, *all* combines all previously mentioned variables into the spacial basis. The graphs in Figure 5 show that while the cumulative returns for *HPK18* are similar to our baseline model in the main analysis, the cumulative returns of the other alternatives are much worse than for their benchmarks in the second half of the investigated period. These results are similar for equal and value weights. However, equal weights perform better than value weights, further indicating that AIEQ receives too much weight and returns yields below its rivals.

6 Conclusion

We evaluate funds that claim to enhance their portfolio management through the implementation of AI and related methods, such as machine learning or deep learning. In summary, funds labelled as AI-driven perform comparably to non-AI funds, showing neither significantly superior nor inferior performance. However, they still underperform in comparison to the market portfolio. When compared to traditional funds, AI-driven funds demonstrate lower activity levels and exhibit greater timing ability. However, their stock selection skills fall short when compared to their counterparts. Additionally, our findings reveal persistent negative performance in both, AI and rival funds, with weak evidence towards momentum in well-performing AI funds and reversal for rival funds.

Although our sample contains more data than the analysis conducted by Chen and Ren (2022), which closely aligns with ours, our data availability remains limited. The integration of AI into portfolio management processes began in 2017 and is therefore relatively recent. In the coming years, with more return observations and a greater number of AI funds in the cross-section, more conclusive findings are likely to be drawn. Future research could also add the CSMAR and RESSET databases to the analysis and broaden the scope of analysis from U.S. to Chinese mutual funds.

References

Arellano, M. (1987). "Computing Robust Standard Errors for Within-Groups Estimators". In: *Oxford Bulletin of Economics and Statistics* 49.4, pp. 431–434.

Avramov, D., S. Cheng, and L. Metzker (2023). "Machine Learning vs. Economic Restrictions: Evidence from Stock Return Predictability". In: *Management Science* 69.5, pp. 2587–2619.

Azevedo, V., C. Hoegner, and M. Velikov (2023a). "The Expected Returns on Machine-Learning Strategies". In: *Available at SSRN 4702406*.

Azevedo, V., G. S. Kaiser, and S. Mueller (2023b). "Stock Market Anomalies and Machine Learning Across the Globe." In: *Journal of Asset Management* 24, pp. 419–441.

Barras, L., O. Scaillet, and R. Wermers (2010). "False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas". In: *Journal of Finance* 65.1, pp. 179–216.

Beck, P. van der (2021). "Flow-driven ESG returns". In: *Swiss Finance Institute Research Paper* 21–71. Available at SSRN: https://ssrn.com/abstract=3929359.

Begenau, J., M. Farboodi, and L. Veldkamp (2018). "Big data in Finance and the Growth of Large Firms". In: *Journal of Monetary Economics* 97, pp. 71–87.

Berk, J. B. and I. Tonks (2007). "Return Persistence and Fund Flows in the Worst Performing Mutual Funds". NBER Working Paper 13042.

Berk, J. B. and J. H. Van Binsbergen (2015). "Measuring Skill in the Mutual Fund Industry". In: *Journal of Financial Economics* 118.1, pp. 1–20.

Busse, J. A., A. Goyal, and S. Wahal (2010). "Performance and Persistence in Institutional Investment Management". In: *Journal of Finance* 65.2, pp. 765–790.

Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance". In: *Journal of Finance* 52.1, pp. 57–82.

Chen, A. Y. and M. Velikov (2023). "Zeroing in on the Expected Returns of Anomalies". In: *Journal of Financial and Quantitative Analysis* 58.3, pp. 968–1004.

Chen, L., M. Pelger, and J. Zhu (2024). "Deep Learning in Asset Pricing". In: *Management Science* 70.2, pp. 714–750.

Chen, R. and J. Ren (2022). "Do AI-powered Mutual Funds Perform Better?" In: *Finance Research Letters* 47, p. 102616.

Chinco, A., A. D. Clark-Joseph, and M. Ye (2019). "Sparse Signals in the Cross-Section of Returns". In: *Journal of Finance* 74.1, pp. 449–492.

Chordia, T., A. Subrahmanyam, and Q. Tong (2014). "Have Capital Market Anomalies Attenuated in the Recent Era of High Liquidity and Trading Activity?" In: *Journal of Accounting and Economics* 58.1, pp. 41–58.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2021). "AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI". In: *Available at SSRN 3554486*.

Dugast, J. and T. Foucault (2018). "Data Abundance and Asset Price Informativeness". In: *Journal of Financial Economics* 130.2, pp. 367–391.

Farboodi, M. and L. Veldkamp (2020). "Long-Run Growth of Financial Data Technology". In: *American Economic Review* 110.8, pp. 2485–2523.

Freyberger, J., A. Neuhierl, and M. Weber (2020). "Dissecting Characteristics Nonparametrically". In: *Review of Financial Studies* 33.5, pp. 2326–2377.

Gabaix, X. and R. S. J. Koijen (2022). "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis". Swiss Finance Institute Research Paper No. 20-91, Available at SSRN: https://ssrn.com/abstract=3686935.

Grennan, J. and R. Michaely (2021). "Fintechs and the Market for Financial Analysis". In: *Journal of Financial and Quantitative Analysis* 56.6, pp. 1877–1907.

Gu, S., B. Kelly, and D. Xiu (2020). "Empirical Asset Pricing via Machine Learning". In: *Review of Financial Studies* 33.5, pp. 2223–2273.

Hoberg, G., N. Kumar, and N. Prabhala (2018). "Mutual Fund Competition, Managerial Skill, and Alpha Persistence". In: *Review of Financial Studies* 31.5, pp. 1896–1929.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp (2014). "Time-varying Fund Manager Skill". In: *Journal of Finance* 69.4, pp. 1455–1484.

Kelly, B., S. Malamud, and K. Zhou (2024). "The virtue of Complexity in Return Prediction". In: *Journal of Finance* 79.1, pp. 459–503.

Ledoit, O. and M. Wolf (2008). "Robust Performance Hypothesis Testing with the Sharpe Ratio". In: *Journal of Empirical Finance* 15.5, pp. 850–859.

Linnainmaa, J. T., B. T. Melzer, and A. Previtero (2021). "The Misguided Beliefs of Financial Advisors". In: *Journal of Finance* 76.2, pp. 587–621.

Martin, I. W. and S. Nagel (2022). "Market Efficiency in the Age of Big Data". In: *Journal of Financial Economics* 145.1, pp. 154–177.

McLean, R. D. and J. Pontiff (2016). "Does Academic Research Destroy Stock Return Predictability?" In: *Journal of Finance* 71.1, pp. 5–32.

Miguel, A. F. and Y. Chen (2021). "Do Machines Beat Humans? Evidence from Mutual Fund Performance Persistence". In: *International Review of Financial Analysis* 78, p. 101913.

Newey, W. K. and K. D. West (1994). "Automatic Lag Selection in Covariance Matrix Estimation". In: *Review of Economic Studies* 61.4, pp. 631–653.

Pástor, L., R. F. Stambaugh, and L. A. Taylor (2015). "Scale and Skill in Active Management". In: *Journal of Financial Economics* 116.1, pp. 23–45.

— (2020). "Fund Tradeoffs". In: *Journal of Financial Economics* 138.3, pp. 614–634.

— (2021). "Sustainable Investing in Equilibrium". In: *Journal of Financial Economics* 142.2, pp. 550–571.

— (2022). "Dissecting Green Returns". In: *Journal of Financial Economics* 146.2, pp. 403–424.

Walter, D., R. Weber, and P. Weiss (2023). "Non-standard Errors in Portfolio Sorts". In: *Available at SSRN 4164117*.

Zhang, J., Z. Peng, Y. Zeng, and H. Yang (2023). "Do Big Data Mutual Funds Outperform?" In: *Journal of International Financial Markets, Institutions and Money* 88, p. 101842.

Zhu, C. (2019). "Big Data as a Governance Mechanism". In: *Review of Financial Studies* 32.5, pp. 2021–2061.

A Identifying AI(-labelled) funds

At the heart of our study lies the identification of funds that feature some sort of artificial intelligence (AI) in the portfolio construction process. To achieve this classification, we rely on the Summary Prospectus Type 497K which we obtain from the Securities and Exchange Commission's (SEC's) Electronic Data Gathering, Analysis, and Retrieval (EDGAR) filing system. Chen and Ren (2022) report that the first AI-managed fund was the *ETF Managers Trust: AI Powered Equity ETF* (AIEQ)[11]. It was issued on 19/10/2017. They find 15 AI-managed funds until the end of 2019. We, therefore, adopt their identified funds and start the construction of our dataset from 2020 forward. We download the Summary Prospectus Type 497K for all CIK starting from 2019. In total, we download filings for 35,887 funds from the beginning of 2019 until the end of 2022. We subsequently apply a number of filters. Table A1 outlines this process of arriving from all fund filings to the final identified AI-managed funds. In the first step, we search for AI-related keywords in the filings. Out of the 35,887 funds, 318 contain any of the keywords in their filings. However, most of these "hits" are misleading. The Summary Prospectus also contains information for instance about the fund manager. If they have a master's degree in "Machine learning", this has nothing to do with the fund itself or how the fund is being managed. To give another example of such non-intended hits, many funds are AI-sector funds and somehow contain variations of our search-words in their title or fund description. Nevertheless, we are not interested in funds that invest in companies that predominantly operate within or prominently focus their R&D on AI. We are keen on solely finding those funds with AI in their portfolio management and optimization process. Therefore, we review the descriptions of the remaining filings manually and decide whether or not - given the information in the Summary Prospectus - the fund uses AI in some stage of the portfolio construction. In this step, we also classify whether a fund uses AI in the step of selecting the investable universe or whether AI is used in the stage of computing a suitable weighting. This leaves us with 93 funds that use AI at some stage of the asset management process to come up with a portfolio. Given that the filings are identified by the company's CIK, we merge the list of identified funds with the CRSP Survivor-Bias-Free U.S. Mutual Fund Database. However, we encounter 12 funds without matching CIKs in the CRSP database, resulting in a final sample of 81 funds.

[11]CUSIP: 26924G813

Table A1: **Selecting Criteria to identify AI-managed funds:** We obtain Summary Prospectus 497K from EDGAR. Panel A shows the filters that we use to identify which of them only use AI keywords in their name or descriptions, which of them only invest in companies that operate predominantly in AI and which funds use AI somewhere along their investment process. Panel B reports the number of matches in the filings that we obtain with our keywords.

Panel A: Selecting Funds	**N**
Number of screened funds	35887
Filtering Step 1:	
Fund Filing contains AI keywords	318
Filtering Step 2:	
Fund uses AI at some stage of their investment process	93
Filtering Step 3:	
Fund has a connected CRSP Portfolio Number	81
Filtering Step 4:	
Account for subclasses of funds	70

Panel B: Used Keywords in Step 1	**N**
Natural Language Processing	31
Artificial Intelligence	25
Machine Learning	25
AI	6
AI-enhanced	4
Deep Learning	4
AI powered	2
Neural Networks	2
Cluster Analysis	1

Pástor et al. (2015) note that many mutual funds hold the same portfolio. These funds represent different share classes with different fee structures. We aggregate sub-classes of the same fund in two ways. First, we connect funds, if they share the same portfolio number. Pástor et al. (2015) already mention that many mutual funds offer multiple share classes, which in reality only represent a claim onto the same underlying portfolio. These funds then carry different fund identifiers, but the same portfolio identifier. Where possible, we, therefore, run our analysis based on portfolio level rather than fund level. A mutual fund that offers the same portfolio with multiple fee structures would otherwise then lead to multiple versions of the same portfolio, which in turn would result in a higher weight for that specific portfolio in the aggregate analysis. This step already finds most of these connected funds. However, there are still some unmatched funds left. For example, CRSP Portno 1023744 (*Voya Small Company Fund*) and 1024790 (*Voya Small Company Portfolio*) hold a close to similar portfolio. We link these remaining funds based on correlation. If the time-series correlation of daily returns between two funds is above 99.9 % and if they are issued by the same company, we consider them the same portfolio. We require at least 252 days of pairwise complete observations for the correlations. In this fourth filtering step, we reduce our sample to 70 unique portfolios that are managed entirely or in part with the use of artificial intelligence.

We want to highlight again that this methodology has a major drawback: As already discussed in Section 1 and 2 this procedure of identifying AI-managed funds is rather a way to identify "AI-managed"-labelled funds. Both interpretations would actually be to some extent true. On the one hand, we do identify funds that intend to enhance their portfolio management process through the application of AI, ML, and DL. On the other hand, we miss out on those funds that do not mention that they apply AI. Furthermore, some keywords are ambiguous: Some funds claim to do "statistical analysis", others report applying "quantitative methods" without providing further insights into what this analysis specifically includes or which methods they use. Since we only keep those portfolios with unambiguous cases, we have to drop all those and lose out on several potentially AI-applying portfolios. Throughout the paper we use the two terms "AI-labelled" and "AI-enhanced" or "AI managed" interchangeably and leave the decision on how to coin the interpretation to the reader.

For the benchmark funds we then match funds in the CRSP Survivor-Bias-Free U.S. Mutual Fund Database based on their fund characteristics. Contrary to the datasets of previous research (c.f. e.g. Berk and Van Binsbergen, 2015 or Pástor et al., 2015), we keep ETFs, sector and index funds in the dataset. Most AI-managed funds are offered through the fund construct of ETFs or use AI methods, such as NLP, to identify the investment universe for an improved version of a predefined sector fund - filtering them would eliminate potential (suitable) benchmark funds.

B Repeat analysis including tiny funds

We re-conduct the analysis from Section 3 without removing funds with TNA below 5 million from our sample. Figure A1 draws the cumulative returns, as we do in Figure 2 in the main analysis. The results are very similar to the ones where we eliminate the tiny funds. Equally-weighted AI funds and their conventional fund benchmark outpace the market benchmark. Value-weighted, the graph shows close to no alteration to the original plot, as the tiny funds receive close to no weight anyway. Nevertheless, the tiny funds do alter the results in Figure A2. While the *Selection* subportfolio exhibited returns close to the ew market portfolio, the tiny funds that use AI to screen for investment opportunities seem to deteriorate the performance. Again, the value-weighted graph does not alter substantially, due to the small weights assigned to the newly included small funds.

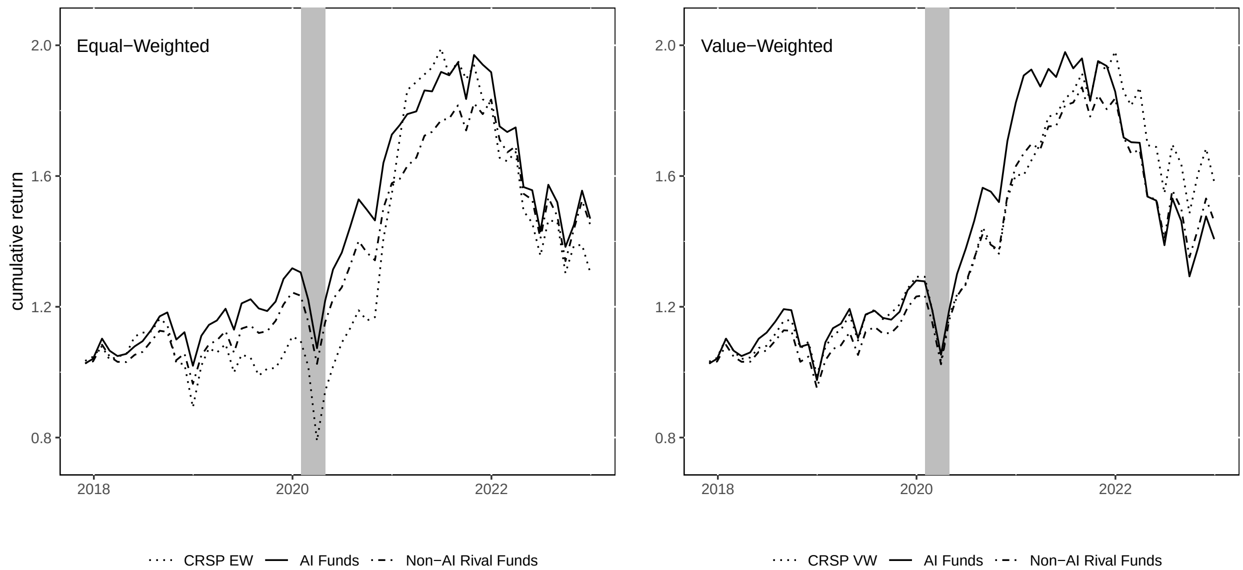

Figure A1: **Return over time incl. tiny funds:** Analogous to Figure 2, we aggregate the returns of AI-labelled funds and their peers. However, now we also include funds with TNA below five million, which we eliminated in the main analysis. The rivals are matched based on z-scores as outlined in Section 2, based on Hoberg et al. (2018). The solid line represents AI funds, dash-doted plots the cumulative returns of the peers, and the dotted line is the equal and value-weighted return of the CRSP market. The grey-shaded area is a recession period, defined by NBER.

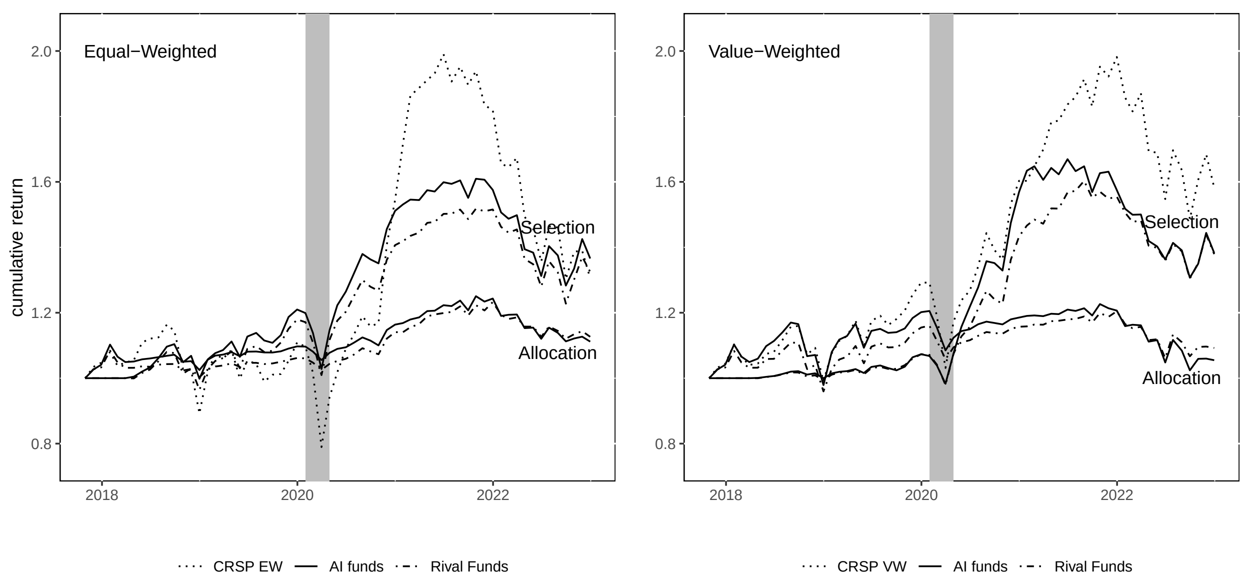

Figure A2: **Subgroups of AI funds, incl. tiny funds:** Again, we sort the funds from Figure A1 into the two categories of Figure 3: Selection an Allocation. The latter employs AI to weigh assets in the portfolio. Selection uses AI at an earlier stage of portfolio management to skim the investment universe for potentially profitable constituents. The solid line depicts the AI-labelled funds, combined into an AI fund portfolio with equal weights in the left panel, and TNA weights in the right plot. Solid lines represent sub-portfolios of AI funds. Dash-dotted lines are respective sub-portfolios of rival funds and the dotted lines are the equal and value-weighted market portfolios of CRSP.