Machine Predictions and Causal Explanations: Evidence from a Field Experiment

Xi Kang and Hyunjin Kim *

August 16, 2024

Abstract

A key role that decision-makers play in organizations is to provide causal explanations for decisions. Yet despite growing evidence on how machine predictions impact decisions, there has been less insight on how they affect decision-makers' explanations and how this relates to their impact on decisions. We explore this question in this paper, using a field experiment we designed and ran across mutual fund analysts in a leading investment research firm. We randomly assigned analysts to receive predictions for fund ratings based on a proprietary machine learning algorithm developed within the company, and observed their decisions and causal explanations for those decisions over 6 months. We find that even when predictions improve decisions, they can worsen the reasoning of the causal explanations that decision-makers provide for their decisions. Inexperienced analysts working on simpler decisions are especially affected, leading them to provide worse explanations when working with machine predictions. In contrast, machine predictions appear to help improve explanations for more complex decisions, even when they do not improve decisions. Additionally, we find little evidence that providing explainability for machine predictions improves either decisions or causal explanations.

*Authors contributed equally and are listed alphabetically. Xi Kang: Vanderbilt University, xi.kang@vanderbilt.edu; Hyunjin Kim: INSEAD, hyunjin.kim@insead.edu. We thank Ruijing Chen, Can He, and Xinni Wu for their excellent research assistance. We are grateful for valuable comments from Arnaldo Camuffo, Giada Di Stefano, Dan Gross, Phanish Puranam, Victoria Sevcenko, Jasjit Singh, and participants at the the Strategy Research Forum, Bocconi ION lab, HEC Paris AI and Entrepreneurship Workshop, the MAD conference, Wharton AI and the Future of Work conference, Strategy Science, DRUID, and Academy of Management. We thank our collaborating company, without which this project would not have been possible. We are grateful for financial support from INSEAD and Vanderbilt University. This experiment was approved by the INSEAD IRB office and pre-registered at the AEA RCT Registry. All errors are our own.

1 Introduction

In recent years, firms have become increasingly interested in using predictive algorithms to support strategic decisions across a wide range of domains, including hiring (Cowgill 2019, Dell'Acqua 2022), resource allocation (Kim et al., 2024), and innovation (Agrawal et al. 2018, Lou and Wu 2021). Growing empirical evidence suggests that algorithms can improve decisions through better prediction, with some studies highlighting the importance of overcoming "algorithm aversion" to increase decision-makers' usage of algorithmic predictions (Dietvorst et al. 2015, Allen and Choudhury 2022, Lebovitz et al. 2022, Tong et al. 2021).

In comparison to this focus on decisions, there has been less insight on whether and how relying on algorithmic predictions might affect how decision-makers can explain their decisions. While improved prediction may complement decision-makers' judgment (Agrawal et al. 2018, 2019) and focus their effort on more cognitively intense tasks (Autor 2015, Camuffo et al. 2022), decision-makers may also anchor on predictions and have difficulty reasoning through the decision independently, which could result in better decisions but poor causal explanations behind decisions. This could have potentially long-run implications for organizations, as developing clear, well-reasoned, and persuasive causal explanations to frame and justify decisions is a crucial role that managers play to convince and align employees, shareholders, and other stakeholders across the organization (Felin and Zenger 2009, 2017, Helfat and Peteraf 2015, Mercier and Sperber 2011, Sørensen and Carroll 2021). Moreover, as decision-makers are tasked with reviewing the most challenging cases when algorithms make inaccurate predictions due to biases, over-fitting, or random error (Shrestha et al. 2019, Choudhury et al. 2020, Raisch and Krakowski 2021), maintaining their ability to reason through and explain decisions may become increasingly important.

In this paper, we propose that machine predictions affect how decision-makers explain the decisions they make, separate from their impact on the decisions themselves. We find evidence consistent with the interpretation that even when algorithms improve decisions, they can make the causal reasoning that decision-makers provide for the decision less clear and persuasive (as well as the reverse). This impact varies depending on the type of decision involved and the experience of the decision-maker. Less experienced decision-makers are more impacted, showing worse reasoning on simpler decisions – even though these decisions improve when aided by machine predictions. In contrast, their explanations for more complex decisions improve, although the decisions themselves appear to worsen.

We designed and ran a field experiment for six months across all fund analysts in a leading financial research firm, which provided a unique opportunity to examine the impact of predictive algorithms across decision-makers of varying experience on different types of decisions. Over a 9-month period, we observed all decisions made by analysts, who conducted research on mutual funds worldwide and provided forward-looking ratings and reports for institutional and retail investors. Analysts performed their work on an online interface where they could see their assigned fund's rating from the last evaluation and receive information about the fund for their analysis. They decided on an up-to-date rating for the fund and provided a causal explanation for their rating decision, which was assessed by an expert review committee within the company before being finalized and released.

At the time of the experiment, the company had developed a proprietary machine learning algorithm that provided predictions on mutual fund ratings, which was representative of the most prevalent form of AI adoption in the finance industry. By providing machine predictions, the company sought to shift analyst effort into improving their causal explanations, which were generally valued more highly than the ratings decisions – by investors as well as other financial research companies like our field site who employed analysts. These explanations were sought out from investors for their insights that could help develop a forward-looking outlook about the fund based on the analysts' unique perspectives and contextual insights.

We randomly assigned analysts to control or treatment conditions, where analysts assigned to treatment additionally received a machine prediction for the fund's rating as part of the information provided on the interface. Half of the analysts assigned to treatment additionally received a figure that sought to provide explainability for the machine prediction, using a method at the frontier of the computer science literature at the time based on Shapley values to show the key fund features that contributed most to the prediction (Lundberg and Lee 2017). We first examine how machine predictions affected analysts' likelihood of changing the rating, as a measure of whether algorithms had any effect on their decision-making. We evaluate whether they improved analysts' decisions by assessing whether treated analysts' recommended funds performed better than those in the control group in the subsequent 3-6 months. We then assess the impact of algorithms on analysts' causal explanations by using scores of their reasoning as assessed by an internal expert committee, as well as the text of the explanations using natural language processing techniques.

We find that on average, machine predictions significantly increased the likelihood that analysts changed their fund rating decisions, but with no precise impact on their decision or explanation

quality. However, we find substantial heterogeneity in treatment effects, suggesting that machine predictions impacted decisions and reasoning in opposite directions across decision types, especially for less experienced analysts. New treated analysts improved the quality of their simpler decisions (non-actively managed equity (NAE) funds) – producing ratings that were more predictive of fund returns. However, they received worse reasoning scores on their causal explanations for these decisions as assessed by reasoning scores issued by the expert review committee. On the other hand, on more complex decisions (actively-managed equity (AE) funds), new treated analysts appeared to make worse decisions that were less likely to predict fund returns. Yet, they provided better causal explanations for these decisions. Comparatively, more experienced analysts at the firm were less impacted by machine predictions. Furthermore, we find little evidence that providing algorithm explainability significantly improved decisions or causal explanations.

To understand the mechanisms driving these effects, we analyze the texts of analysts' causal explanations, their perceptions of algorithms, and the impact of providing them with algorithm explainability. We explore a few possible explanations and find some suggestive evidence consistent with the interpretation that analysts were more likely to see algorithms as useful for simpler decisions, and this reliance worsened their reasoning about the decision. We find that analysts wrote shorter, less technical, and less coherent explanations in these cases, suggesting that they either exerted less effort into reasoning or were less able to reason through a clear and persuasive causal explanation when they relied on machine predictions. Moreover, we find that providing algorithm explainability led analysts to rely even more on algorithms for these simpler decisions, possibly leading them to adopt the explanation provided by the explainability treatment in part or in full. Importantly, this difference materialized even though the company valued the causal explanations for both types of funds equally, as every fund was carefully selected to be covered by analysts from the universe of nearly 8,00s0 mutual funds.

Our study contributes to growing research on algorithms and decision-making in organizations, which has provided insight on how the use of algorithms affects firm decisions and performance (Choudhury et al. 2020, Kim et al. 2024, Tong et al. 2021, Brynjolfsson et al. 2021). In this study, we highlight how the use of algorithms may also affect how decision-makers reason about and explain decisions, and suggest that algorithms may worsen their causal explanations even when improving decisions. These findings suggest that organizations may need to be careful about using algorithms to support decisions, especially when decision-makers are less experienced and rely more on algorithms.

Our study also contributes to the literature on how artificial intelligence (AI) may be a complement or substitute to human labor (Acemoglu and Restrepo 2019, Autor and Salomons 2018). Recent work has explored how AI may disrupt labor markets, constructing indices to measure a profession's exposure to algorithms at the industry and country level (Felten et al. 2021, Frey and Osborne 2017). More closely related to our paper is work on how AI affects employee effort across tasks. Grennan and Michaely (2020) examines how potential exposure to AI approximated by industry-level social media data increased stock analysts' exits from the profession and the effort they spent on complementary tasks, such as engaging more with investors. Kang (2023) similarly suggests that AI adoption in the workplace provides implicit incentives for stock analysts to develop skills that are complementary to machine predictions. Our study highlights the importance of examining outcomes beyond short-run decision performance especially for complex, high-stakes decisions, and shows that while providing machine predictions as decision aids may increase incentives when it threatens job security, it may not shift effort into other complementary tasks that involve higher cognitive effort, such as reasoning about and crafting causal explanations for the decisions they make.

This paper also contributes to emerging research on explainable AI. We find suggestive evidence that algorithm explainability may indeed help increase the usage of algorithms, even when it does not provide an accurate or consistent explanation behind machine predictions – as is the case with most methods available today. This suggests that while it may be helpful when seeking to increase adoption of highly predictive algorithms for experienced decision-makers who may be overly averse to machine predictions (Dietvorst et al. 2015, Allen and Choudhury 2022), algorithm explainability should be deployed carefully. Our results also highlight that algorithm explainability may be a double-edged sword, as it can undermine decision-makers' ability to reason through and explain their decisions.

2 Causal explanations and machine predictions

2.1 The importance of causal explanations in strategic decision-making

Communicating causal explanations for decisions is a key component of strategic decision-making. To execute on strategies, managers must persuade and justify their decisions to employees, shareholders, and other stakeholders (Pfeffer 1981, Fiss and Zajac 2006, Helfat and Peteraf 2015). When managers can clearly explain the causal factors underpinning their decisions, it can enable

stakeholders to see the logic behind decisions, thereby facilitating a shared vision, as well as their buy-in and support for execution (MacLennan and Markides 2021, Ehrig and Schmidt 2021). Moreover, well-articulated causal explanations provide a framework for managers to justify their decisions to external stakeholders such as investors, customers, and regulatory bodies. By presenting a causal logic for their strategies, managers can enhance the credibility and legitimacy of their actions. This kind of causal reasoning has been found to be particularly important for firm performance in environments characterized by uncertainty and complexity, where the ability to convincingly explain and justify decisions can be as crucial as the decisions themselves (Bettman and Weitz 1983, Pfeffer 1981). For example, even if a decision did not lead to superior performance, this may be attributed to a bad draw due to uncertainty, and shareholders can trust the decision-maker because they perceived the causal explanation behind the decision to be sound.

Additionally, causal explanations can reflect decision-makers' theories about what drives value creation, which has received increasing attention in strategy research as a key determinant of better strategies (Felin and Zenger 2009, Csaszar 2018, Camuffo et al. 2020, Agarwal et al. 2023). Felin and Zenger (2017) highlight that decision-makers with a unique and clear theory grounded in a causal logic about what leads to superior performance – by providing a hypothesis on the assets, knowledge, and activities that enable a solution for a particular problem – can craft better strategies and superior firm performance. Similarly, Sørensen and Carroll (2021) argue that all great strategies are causal arguments with clear and logical reasoning. Effective strategic decision-making has been also posited to involve theorizing and explaining of the impact of key factors on performance, a process that resembles the elaboration and justification of initial beliefs through logical reasoning (Felin and Zenger 2009). This process of theorizing involves not only the identification of factors that influence performance, but also a detailed explanation of how these factors interact and lead to specific outcomes. Causal explanations can thus represent contextualized statements for a focal problem that can be derived from or embody the forward-looking theories that decision-makers hold.

2.2 The impact of machine predictions on causal explanations

The integration of machine predictions into strategic decision-making processes presents new challenges and opportunities. Agrawal et al. (2018) theorize that AI improves decisions by providing improved predictions, which are complementary to human judgment especially under conditions of uncertainty. Growing research provides supportive empirical evidence (Kleinberg et al. 2017,

Hoffman et al. 2018), suggesting that the use of machine learning algorithms can significantly aid in decision tasks involving prediction. At the same time, recent work has highlighted that AI may not always improve decisions in organizations, as it introduces unique challenges such as biases embedded in data (Choudhury et al. 2020) and the need to consider broader sets of objectives or preferences beyond the algorithm's predictions (Kim et al. 2024).

Despite growing research on the impact of predictive algorithms on decisions, there has been less insight on whether and how machine predictions affect the way that decision-makers causally explain and reason through their decisions. Understanding this question is important, because providing causal explanations – as well as social skills more broadly – has been posited as a key role that managers can play in a world of machines (Deming 2017, Beck and Libert 2017, Luo et al. 2021).

Theoretically, the effects of machine predictions on causal explanations are ambiguous. On the one hand, machine predictions have the potential to improve causal explanations. Building on Agrawal et al. (2019) who highlight the complementarity between predictions and judgment, Athey et al. (2020) argue that AI's augmentation of human activities can be profit-enhancing in complex decision areas, conditional on decision rights and aligned incentives. Moreover, by taking over repetitive tasks, AI can ease the multitasking problem and thus free decision-makers to focus on more cognitively complex tasks that require deeper thought, such as developing causal explanations. This re-allocation of effort can lead to more effective use of time and cognitive capacity (Simon 1947, Autor 2015, Agrawal et al. 2018, Camuffo et al. 2022). Grennan and Michaely (2020) similarly suggest that when stock analysts need to juggle multiple tasks, being exposed to AI advancements in their area of expertise redirects analysts' attention to tasks complementary to AI, such as engaging more with investors and working on more stocks, which raises the possibility that the use of machine predictions may also potentially improve decision-makers' causal explanations of their decisions by reallocating their efforts to reasoning and communication.

On the other hand, using machine predictions may also worsen decision-makers' causal explanations. Dell'Acqua (2022) suggests that decision-makers may become complacent and reduce their effort when working with algorithms, which may also potentially extend to reasoning through causal explanations for the decisions rather than reallocating their efforts to enhancing these explanations. The division of labor induced by AI—by separating prediction from decision-making—may also lead to excessive modularization, which has been theorized to diminish decision-makers' understanding of complex systems (Siggelkow 2002, Ethiraj and Levinthal 2004, Billinger et al. 2014). This challenge may be particularly acute for inexperienced decision-makers who may not have developed

as deep an understanding of the complexity especially in situations where causal explanations are inherently complex, leading them to be unable to develop clear causal models or integrate machine predictions into the causal model they have in mind.

In sum, the development of clear, insightful, and persuasive causal explanations plays a fundamental role in strategic decision-making, as it is crucial for managerial rationalizing and strategy execution in the face of uncertainty and complexity. Theoretically, using machine predictions as decision aids presents both opportunities and challenges in developing causal explanations, and this paper aims to explore these effects empirically as well as the mechanisms through which machine predictions affect decision-makers' causal explanations.

3 Experimental context and setting

3.1 Why study mutual fund analysts?

To examine the impact of machine predictions on causal explanations, we partnered with a leading financial research firm headquartered in the United States to run an experiment across their mutual fund analysts.

Mutual fund analysts produce reports that provide recommendations for investors, with insights on the fund's strategy, holdings, risk-taking, and performance in the context of changing market conditions. Mutual funds are the predominant type of investment vehicle for U.S. households, with 52.3% of households (amounting to 68.7 million households) holding mutual fund investments in 2023. In 2023, the global mutual fund assets market was estimated at approximately $65.54 trillion USD, with close to 8,000 unique funds in the U.S. alone with $18.85 trillion USD in total financial assets (Board of Governors of the Federal Reserve System (US) 2023). Given the large number of mutual funds and the variety of investment strategies available on the market, investors seek out analyst reports to better inform themselves about potential funds to invest in, as well as their broader investment strategy.

Analyst reports contain two key components. The first component is a rating for the mutual fund, which is the analyst's recommendation on whether to invest in the fund based on their forward-looking predictions of fund returns. This recommendation can be in the form of a binary recommendation, or as a rating that scales between "not recommend" to "highly recommend". Previous studies have shown that investors react to these mutual fund ratings, with higher-rated funds attracting significantly higher net cash inflow compared to lower-rated funds (Sirri and Tufano 1998,

Gruber 1996, Goetzmann and Peles 1997), suggesting that the ratings decisions that fund analysts make have broader implications beyond their own company to investors in the market and the efficiency of the mutual fund industry. [1]

The second component of these reports is the analyst's causal explanation for their ratings decision. Analysts specialize in certain types of mutual funds and conduct in-depth quantitative and qualitative research: in addition to tracking changing market conditions, analysts review the fund's investment strategy, documents issued by fund managers, as well as news articles that disclose information related to the fund or its parent firm. They also interview fund managers and other key decision-makers (e.g., managers at the parent firms, traders, and risk managers) to gather first-hand information, cross-validate secondary information, and discuss any concerns. They then piece together all of this research to provide a coherent reasoning to investors about what factors they believe causally drive fund performance and the underlying reasons why they assigned the rating they did for their forward-looking outlook on the fund.

These analyst explanations are generally valued more highly than the ratings themselves – by investors as well as the broader industry of financial research companies who employ analysts. Interviews with analysts as well as industry reports highlight that client demand for fund research is driven by the novel insights and "angles" or causal models they can glean from analyst reasoning to assess the fund, in hopes of better informing their investment strategies. Novel and coherent causal models portrayed by the analysts help persuade investors to incorporate analyst research into their investment decisions. Studies have documented that analyst reports that provide better reasoning to investors gain more visibility and receive larger market reactions (Bradshaw et al. 2021, Campbell et al. 2019). The importance of causal explanations is also reflected in the way that financial research is sold to clients. While all clients can view the ratings, only clients with premium subscriptions have access to detailed analyst reports that offer causal explanations for these ratings. In fact, while ratings can be provided for a larger set of funds, only a small subset—those covered by analysts—include these causal explanations, leading companies to carefully select which funds are in their analyst coverage. This has implications for analyst incentives as well. While analysts generally do not have high-powered incentives, their career concerns are tied to the reputation they can establish from their reports, which can provide them with better outside

[1] However, this is unlikely to introduce the concern that analyst ratings will have a direct impact on fund returns within a short time frame. Fund managers construct portfolios of often a large number of assets. For funds that invest in fixed income assets, the returns of the underlying assets cannot be affected by fund ratings. For actively managed equity funds, a positive rating of the fund is unlikely to provide signals strong enough to affect the performance of individual underlying firms within the portfolio, especially within a 3-month period.

options (Cornaggia et al. 2016, Kempf 2020, Lourie 2019). Star analysts may not always produce ratings that accurately predict future outcomes, which represent a random draw from a distribution with high variance. However, analysts can demonstrate their expertise and research capabilities by presenting and explaining their causal models in their reports, which should account for key sources of uncertainty regardless of what future outcome is realized.

Recent years have seen the growth of the global fund market, which has increased the sheer number of mutual funds and subsequently the demand for fund analyst research. The total number of global open-ended funds increased from 66,362 to 137,892 between 2007 and 2022, making it impossible for research firms to cover the entire universe of funds. For example, the financial research company that we collaborated with for our experiment—one of the largest in the market—was only able to cover about 10% of the mutual funds with their analyst research.

This increasing demand from clients, as well as the growing availability of data, has accelerated the adoption of AI-related technologies in the financial research industry. Both incumbents and startups alike have invested in developing algorithms to perform analytical tasks that historically relied on financial experts, exploring how algorithms can help evaluate financial securities based on big data analytics, leveraging available quantitative data also from alternative sources (Bollaert et al. 2021, Grennan and Michaely 2020). Large banks, including JP Morgan, Citigroup, and Deutsche Bank, were reported to have hired thousands of AI-related roles in 2023 to use AI to model risk and conduct data analytics (Shaw and Gani 2023).

This context thus provided a unique opportunity to understand how the use of predictive algorithms might affect decisions and causal explanations. The nature of this work shares many key features with other kinds of complex knowledge work and strategic decision-making, where decisions are forward-looking, complex, and made under uncertainty, with both a predictive component and an explanation component to explain the causal reasoning—making it plausible that insights from our study could potentially generalize to other similar settings that have often been challenging to study empirically due to small sample or measurement issues.

3.2 Experimental setting and data

We collaborated with a leading company in the financial research industry to run an experiment across all decisions made by their full set of mutual fund analysts globally (encompassing the US, EMEA (Europe, the Middle East, and Africa), Asia, and Australia). It was at the forefront of adopting AI-driven technologies to aid the process of financial research, having been one of the first

companies to launch financial research products leveraging predictive algorithms.

At the time of the experiment, the company had developed a machine learning algorithm to help analysts make decisions on mutual funds, but had not yet rolled them out across analysts. The algorithm was a tree-based model trained on analysts' past decisions and all quantifiable data that analysts reviewed to make their decisions, including all available data except private information collected by analysts, such as interviews of managers and any tacit knowledge analysts accumulated from experience (e.g., the style of fund managers and who may be prone to over-optimism). The firm had closely monitored the algorithm's performance since its development and reported that the algorithm clearly separated over- and under-performing funds into categories following the company's rating methodology: highly-recommended funds outperformed their category average by about 1%, and not-recommended funds underperformed their peers by 1% over a 12-month period following the rating. After a few years of observing the performance of algorithmic ratings, piloting changes, and back-testing them to ensure their high level of quality, the firm decided to introduce these machine predictions as a decision aid for analysts. While analysts had not previously used predictions from these algorithms, they were continuously updated on their development and informed of their methodology and overall quality.[2]

We collaborated with the firm at this juncture to randomize the provision of machine predictions across analysts, which provided a unique opportunity to measure how analysts' decisions and causal explanations changed when working with machine predictions.

Decisions: we collected all analysts' fund rating decisions between April 6, 2021 and January 19, 2022, which covered 1,780 rating decisions in total. Mutual fund analysts at the firm specialized in certain asset classes (e.g., equity, fixed income), but covered funds both inside and outside of their specialization.[3] For each fund in their coverage, analysts updated their decisions on the fund's ratings at least once every year. Because these ratings were forward-looking in nature, we also collected data on these funds' returns in the subsequent three months and six months, respectively, which enabled us to assess these decisions by evaluating the extent to which funds recommended by analysts provided higher risk-adjusted returns on average in the months that followed.

Causal Explanations: we collected a measure of the quality of analysts' causal explanations between April 6, 2021 and January 1, 2022.[4] Explanations for decisions are rarely articulated or

[2]Analysts were continuously updated on these algorithms over several years, so we do not expect any anticipatory effects in the months during the experiment.

[3]This coverage list was determined by managers and scheduled ahead of time, and could not be changed spontaneously by analysts.

[4]On January 1, 2022, the company decided to remove the requirement for internal review committees to provide

observable in full in practice, and generally difficult to assess according to stakeholders who would be relevant for the decision. In this company, all analyst ratings and reports had to be approved by an internal expert committee before they could be released publicly, which provided us with a unique opportunity to collect a measure of explanation quality. All committee members were blind to the experiment and assignment to treatment, and the evaluation criteria used by the committees stayed constant throughout this period without any changes. The committees evaluated analysts' causal explanations for each of their fund rating decisions by answering the following questions on a 1 to 10 scale: (1) How would you grade the analyst's knowledge of the strategy? (2) How would you grade the analyst's reasoning behind the ratings? Committees were composed of one to four experienced analysts who managed these teams and held director-level positions. Single member committees were the most common, evaluating 78.48% of all rating proposals. Committees with two, three, or four members evaluated 15.79%, 5.56% and 0.16% proposals, respectively. For each fund rating submitted by an analyst, we standardized evaluator scores and averaged them across the committee. Appendix II.1 provides two examples of these reasoning scores accompanied by the summary section of the corresponding analyst reports, contrasting a report that received a high reasoning score (example 1), with another that received a low score (example 2).

In addition to these reasoning scores, we collected text and survey data to capture richer measures on analysts' causal explanations. When analysts submitted each rating decision and report, they listed the causal drivers that drove their decisions in a survey built into the workflow, which allowed us to capture the number of key factors they proposed as driving their decision. We also leveraged large language models (LLM) to extract causal statements from the text of analysts' reports. We conducted in-context learning in GPT-4 to identify causal statements across the reports, decomposing each extracted statement into its cause and effect (see Appendix II.2 for examples of the causal statements extracted). In addition to these measures, we also directly analyzed the text that analysts wrote in their reports, using various natural language processing techniques to evaluate the length, readability, and coherence of the text. We used two widely used readability measures, Flesch-Kincaid and Coleman-Liau, both of which provide a decreasing measure of readability (the higher the score, the more difficult it is to read, associated with a higher grade level of education). The Flesch-Kincaid grade-level score factors in both the number of words in sentences and the difficulty of words based on the average number of syllables per word (Kincaid et al. 1975).

explicit scores in order to reduce their workload, which meant that they only provided approval decisions and feedback comments back to analysts from this date onwards. This provided us with 1,313 reasoning scores from committees.

The Coleman-Liau index provides an alternative measure based on sentence length and the number of characters per word to approximate readability (Coleman and Liau 1975). Finally, we computed a coherence score based on the coherence of topics discussed in the reports using topic modeling (Röder et al. 2015).

Perceptions of machine predictions: we collected data on analysts' perceptions of machine predictions for those assigned to treatment. For each fund, analysts assigned to treatment rated on a scale of 1-10 how accurate they found the machine prediction to be and the extent to which it contributed to their decision.

We observed each of these measures across all decision-makers and decision types across the company, which spanned a range of analyst experience and decision complexity. The tenure of analysts in the company ranged from 0 to 29 years, with approximately 25% of analysts with more than 10 years of experience in the firm. An important distinction in the company was whether analysts were new to the firm, as the company had a unique methodology and new analysts had to undergo substantial training to learn how to evaluate mutual funds according to this methodology.

The type of decisions analysts made also varied in terms of their complexity based on the fund type, providing variation on decision complexity, a key attribute of strategic decisions (Csaszar 2018). For each fund, analysts assessed the quality of the fund's management team and the fund's investment strategy. Evaluating the team behind the fund included assessing elements such as the ability and experience of the managers, the availability of supporting resources, conflicts of interests, risk-taking tendencies, and fit with the investment strategy. Evaluating the fund's investment strategy involved assessing whether the fund had distinct investment processes that were clearly defined and consistent with the investment process and performance objective, and that produced standout results in a systematic and sustainable way.

This decision process was typically more complex for active equity funds compared to all other fund types, a key distinction made by the company. Active equity (AE) funds are portfolios of stocks actively managed by fund managers, meaning that the fund's managers and investment strategy change more substantially and frequently, the returns tend to be more volatile, and there is a lot of information and news coverage about the stocks in the portfolio that drives these changes. Non-active equity (NAE) funds include passive equity funds, fixed income funds, allocation and alternative funds. These funds change less over time in terms of their management and strategy attributes, and thus generally involve assessing how the market has changed since the last rating update and what this might mean for the performance of the fund. AE funds require extensive

analysis for each rating update, to understand and evaluate changes in the team, investment philosophy, strategy, risk-taking, resources, and execution. Analysts also analyze the portfolio holdings of funds using extensive databases and analytical tools to evaluate each of these changes relative to other similar funds and how they might perform in the current market environment. Decisions on AE funds thus tended to be more complex compared to NAE funds. Importantly, although analysts have a primary fund type that they specialize in (AE or NAE funds), they cover both AE and NAE funds due to capacity and legacy reasons.

These differences in the complexity of analysis do not indicate that ratings decisions or causal explanations on NAE funds are less important than AE funds for the company or its client investors. Passive investment has been on the rise in the past decades due to their relative low cost and higher risk-adjusted returns compared with active investment on average. Assets under management by passively managed US mutual funds and exchange traded funds amount to $13 trillion, which overtook their actively managed counterparts for the first time in December 2023 (Schmitt 2024). The NAE funds that analysts were assigned to cover were thus carefully selected to help guide investors in navigating their choice of investment vehicles among these growing set of choices.

3.3 Causal explanations and decision at baseline

Causal explanations that received higher reasoning scores identified key causal drivers behind the rating to explain and justify the decision. Because possible drivers are numerous, spanning more than fifty high-level categories, the primary goal of the explanation in the analyst report was to pare down the list of drivers to identify ones that mattered most, which could take considerable effort. At baseline, explanations with fewer causal drivers were more likely to receive higher reasoning scores, suggesting that these were better able to explain core reasons that drove the decision (Figure A.1 (b)). We observe this pattern across both AE and NAE funds (Figure A.5). On average, experienced analysts received higher reasoning scores on their causal explanations (Figure A.1 (a)).

We also find that better decisions and causal explanations show a positive correlation at baseline. Recommended funds accompanied by explanations with lower reasoning scores were less likely to observe higher risk-adjusted returns compared to non-recommended funds, whereas recommended funds with higher reasoning scores were more likely to observe at least as high returns compared to non-recommended funds (Figure A.1 (d)) – suggesting a positive association between the performance of ratings decisions and reasoning scores.

4 The machine prediction experiment

4.1 Experimental design

The experiment ran across all 97 mutual fund analysts at the company globally. We randomly assigned half of the analysts (48) to treatment, which provided them with machine predictions on ratings. Half of this treatment group additionally received a figure that provided some algorithm explainability. We implemented the experiment for 6 months from July 18, 2021 to January 19, 2022 and observed a total of 1,780 decisions.

Before the experiment, analysts made decisions without any algorithmic aids. They conducted research using an internal corporate web interface, which displayed information relevant to fund analysis and provided links to other databases (Figure A.2). The top of this interface highlighted the most recent rating that was assigned for that fund, and several tabs of subpages provided data on the fund's historical performance, fee structure, and details about the fund management team and the parent firm. Analysts analyzed this data along with others that they collected on their own (e.g., via interviews) to assess the fund strategy and its potential to outperform its peers. They then synthesized this analysis to rate both the team of fund managers and the process/strategy of the fund, which were then aggregated along with other metrics to classify the fund into one of five categories displayed to investors indicating its investment quality – the top three of which were considered to be "recommended", and the bottom two which were not recommended.[5] The final aggregate rating constituted whether this fund was shown to carry their recommendation. Analysts then elaborated on their reasoning in a report that provided an explanation with relevant analysis on reasons driving the rating decision. These were all submitted via the corporate interface, which were then reviewed by an internal expert review committee, who was blind to the experiment and analyst assignment to treatment.

Within this context, analysts were randomly assigned to one of three experimental groups: (1) a control group, who made decisions according to business-as-usual; (2) a treatment group who received machine predictions on ratings; and (3) a treatment group who received machine predictions on ratings with some form of algorithm explainabilty. The control group saw no changes on the interface and made decisions as usual (Figure 1 (a)). Analysts assigned to either treatment group were shown an additional line of information at the top of the interface that provided a machine prediction for the fund's current rating, which was displayed above the last assigned rating for the

[5] Analysts observed an aggregated rating and adjusted their sub-ratings accordingly until a final decision is made.

fund (Figure 1 (b)). Treatment 1 provided machine predictions only, while Treatment 2 additionally provided a figure that highlighted the ten key features that contributed most to the algorithm's prediction for that fund (Figure 1 (c)).[6] This explanation was based on Shapley values, a concept developed in the computer science literature as one method for "explainable AI", using cooperative game theory to determine the average marginal contribution of each feature value across all possible values in the feature space (Lundberg and Lee 2017). These treatments were extensively tested and piloted with a small group of key stakeholders at the company, which included the team that managed the interface, the team that developed the algorithm, the head of investment analysis, and directors of research teams who managed analysts and previously were senior analysts themselves.

Half of the analysts were randomly assigned to control, and the other half to one of the two treatment conditions using stratified randomization. We stratified on the fund type that they specialized in and analyst tenure. The randomization was coded into the backend of the corporate interface, which enabled us to estimate intent-to-treat effects. Analysts rated and published a total of 1,137 funds during the experimental period, averaging approximately 2 funds per analyst per month on average.

As analysts covered different funds individually, the nature of their work did not involve collaboration, and all analysts worked independently. We did not expect non-compliance, as random assignment was coded in the backend of the corporate web interface, and fully controlled what was shown on the interface for the analyst. However, it is possible that analysts may not pay attention to the treatment intervention, which would bias us against finding a treatment effect. The experiment was implemented during the Covid-19 pandemic, when all analysts worked from home and individually accessed the corporate interface – further limiting the likelihood that other analysts could observe different versions of the interface. The interface was also actively undergoing development during this time, which meant that analysts were aware that certain features were becoming available at different times across different funds. Given these contextual details, we and the key managers at the company did not expect substantial spillover effects across analysts to drive potential treatment effects. Since the machine predictions are specific to each fund, even if analysts in the control learned about others in the treatment group, they would not have access to the machine predictions of the funds they were assigned to work on.

As expected by randomization, experimental groups were well-balanced across baseline indicators

[6]These features were common fund attributes that all analysts were familiar with, such as the fund's fees, its past performance, and fund age.

(Table A.1). There was no attrition.[7]

4.2 Econometric specification

To evaluate how machine predictions affect analyst decisions and causal explanations, we run the following difference-in-difference specification as our base model as pre-registered:

$$y_{ift} = \alpha Treat_{ift} + \beta Post_{ift} * Treat_{ift} + \gamma Post_{ift} + \delta_j + \eta_t + \epsilon_{ift} \quad (1)$$

y_{ift} is the outcome of interest for fund f evaluated by analyst i at month t, which includes: (1) a binary indicator of whether the analyst decided to change the fund's currently assigned rating; (2) the reasoning score for the analyst's causal explanation for the ratings decision as assessed by the internal expert committee; (3) the number of self-reported decision drivers that drove the analyst's decision; (4) the number of causal statements identified in the analyst report. $Treat_{ift}$ is an indicator variable that takes value 1 for analysts assigned to treatment (pooling across both treatment groups) and 0 otherwise. $Post_{ift}$ is an indicator variable that equals 1 for the experimental period and 0 otherwise. While fixed effects are not necessary for identification given that treatment is randomly assigned, we run this specification with randomization strata (δ_j) and month (η_t) fixed effects to account for any random differences across experimental groups and soak up noise. We cluster standard errors at the analyst level, which is the unit of randomization.

β identifies the treatment effect of providing machine predictions, which is the main coefficient of interest. We also use the same specification with separate indicators for each treatment group to compare the two treatments and assess the impact of providing explainability for algorithmic predictions. The key identifying assumption is that analysts assigned to treatment did not have systematically different trajectories of rating decisions and reasoning compared to those in the control group for reasons other than the algorithmic prediction treatment, which was randomized.

We then evaluate whether machine predictions led to better decisions by assessing the extent to which analysts' recommended funds observed higher risk-adjusted returns in the months following their ratings decision. We evaluate fund returns over a three- and six-month period, using a monthly

[7]While we observe no attrition among analysts, the number of observations vary across outcomes. We observe the published ratings of all 1,780 funds. However, the company stopped collecting reasoning scores starting January 2022)– resulting in a total number of observations of 1,313. For fund returns, the sample size was reduced from 1,780 to 1,664 because there are funds with missing data on monthly returns during the prior 36-month period. This should not bias our treatment effects, since for reasoning scores it means that we simply collected data on both groups for a shorter number of months, and for fund returns, the availability of prior 36-month data should not correlate with treatment given the randomization. Indeed, the share or number of missing data for any of these outcomes does not vary between experimental groups (Table A.2).

time series of fund return data from Morningstar Direct and the Center for Research in Security Prices (CRSP) Mutual Fund Database. Following the accounting and finance literature on ratings and fund returns, we calculate risk-adjusted returns to account for variation in risk-taking across funds, using the benchmark Fama and French (2015) five-factor model. We estimate factor loadings by running 36-month rolling regressions where each fund's excess returns over the prior 36 months are regressed on the relevant factor returns. The factor loadings are then used to adjust the raw monthly return data to estimate risk-adjusted returns. We also run the Fama and French (1993) three-factor model and the Carhart (1997) four-factor model to check the robustness of our results, which provide consistent results across risk-adjustment methods.

We regress these risk-adjusted fund returns on a variation of specification (1) above, which includes interactions with a binary indicator of whether a fund was recommended by the analyst. Following best practices in the finance and accounting literature (Armstrong et al. 2019), we also control for key fund-level characteristics, fund age and fund size (proxied by net assets). We thus compare whether recommended funds by analysts randomly assigned to treatment observed higher fund returns compared to those recommended by analysts assigned to the control group.

In addition to average effects, we explore heterogeneity in treatment effects for all outcomes by interacting binary indicators of pre-registered variables. We focus on two key dimensions of heterogeneity: fund type, which determines the complexity of decision-making that analysts engage in, and analyst experience, which is associated with domain expertise. For fund type, we use a binary indicator for AE funds, which represent 45.35% of all fund decisions during the experimental period. For analyst experience, we use a binary indicator that equals 1 for new analysts with less than three years of experience, and 0 otherwise. We use this cutoff as analysts were promoted to an official "analyst" position from the associate level within three years on average.

5 The impact of machine predictions on decisions and causal explanations

5.1 The impact of machine predictions on decisions

To assess whether analysts considered machine predictions in their decisions, we first examine whether treatment impacted analysts' likelihood of changing the rating assigned to the fund.[8] Table

[8]It is worth noting that changing the rating has also been seen as requiring more confidence: analysts are more likely to maintain the status quo unless they have enough confidence that this update is correct (Fernandez and

1 Panel A Columns 1-4 show estimates of the average treatment effect across different specifications, with Column 4 showing that analysts randomly assigned to receive machine predictions were on average 8 percentage points more likely to change the fund rating (p-value=0.01), a 71% increase over 11% of rating decisions that were changed in the control group at baseline. This suggests that machine predictions had a fairly large impact on analyst decisions, as they induced a change in the assigned fund rating. The estimated treatment effect does not vary significantly across decision type or analyst experience (Table 1 Panel A Columns 5-6), suggesting that analysts largely responded to the machine prediction treatment. This treatment impact is unlikely to be driven solely by the fact that analysts receive a rating to anchor on, because all analysts in both control and treatment groups were provided with the live analyst rating for the fund that they assessed, which represents an analyst-generated prediction for all funds.

We then explore the extent to which machine predictions improved decisions, by evaluating the risk-adjusted returns of analysts' recommended funds in the months that followed. To analyze whether treated analysts' ratings performed better, we compare whether the risk-adjusted abnormal returns of funds that treated analysts recommended were higher compared to those of the control group.[9] Column 1 of Table 1 Panel B shows that on average, funds recommended by treated analysts outperformed those of analysts assigned to the control group by 0.89 percentage points (89 basis points), although this estimate is imprecise (p-value=0.77). On average, funds recommended by analysts assigned to the control group observed returns of -4.5%.[10]

These average effects mask some heterogeneity in treatment effects, which appear to vary depending on the type of decision involved – improving simpler decisions while worsening more complex ones. NAE funds recommended by treated analysts observed 9.9 percentage point higher returns (p-value < 0.001) compared to those in the control group – suggesting that these simpler decisions improved as a result of using machine predictions. In contrast, the treatment effect on the returns of more complex AE funds recommended by analysts was 16.8 percentage point lower (p=0.002).

Moreover, the treatment effect appears to be driven by new analysts with less experience. For

Rodrik 1991, Samuelson and Zeckhauser 1988).

[9]We follow the finance literature in analyzing risk-adjusted abnormal returns, which allows us to determine the performance of a financial asset or portfolio when compared to the overall market or a benchmark index. It has been used to help identify an investor's skill on a risk-adjusted basis, finding that high-skill investors consistently generate positive risk-adjusted returns (Soydemir et al. 2014, Bach et al. 2020). Studies in accounting and finance have also analyzed risk-adjusted returns to evaluate the performance of mutual fund ratings (Armstrong et al. 2019, Ertugrul and Hegde 2009, Luo et al. 2015), as ratings that better predict future fund returns generate more profits for investors who build their portfolios following these ratings.

[10]This is not atypical: the literature documents negative fund risk-adjusted returns on average, net of expenses and trading costs (Barras et al. 2010, Carhart 1997, Jensen 1968)

those assigned to treatment, simpler NAE funds recommended by new analysts observed 11.6 percentage point higher returns (p-value = 0.002) compared to those in the control group, and more complex AE funds recommended by new analysts observed 16.8 percentage point lower returns comparatively (p-value=0.04). Treatment effects are more attenuated and less precisely estimated for experienced analysts for both simpler and more complex decisions, with interaction coefficients in the opposite direction compared to new analysts for each decision type.

These treatment effects remain qualitatively robust when analyzing fund returns within 6 months, though attenuated (Table 1 Panel B Columns 4-6). The attenuated fund returns result within 6 months do not necessarily suggest that the treatment effect dissipates over time. Fund returns become more volatile and subject to a larger number of economy-wide factors in the longer run and become more challenging to predict. This is partly why investment research firms often seek to update ratings more frequently.

5.2 The impact of machine predictions on causal explanations

We next evaluate how machine predictions affected analysts' causal explanations, and find treatment effects to be in the opposite direction from decisions. Figure 2(a) and Table 2 Column 1 show that average treatment effects on reasoning scores are not economically or statistically significant (Kolmogorov-Smirnov test p-value=0.316; treatment estimate= 0.01, p-value=0.94). Confidence intervals on these estimates are quite wide, ranging from -0.3 to 0.3 standard deviations in either direction, making it difficult to rule out large effects.

We again find that treatment effects vary by decision type and analyst experience. However, in contrast to the treatment effect on decisions, machine predictions worsened the reasoning scores of analysts' causal explanations for simpler decisions (which improved), while improving their explanations for complex decisions (which worsened) (Figure 2(b) and (c)). Treatment appears to have shifted the distribution of scores in both cases (Kolmogorov-Smirnov test p-values: 0.002 and <0.001, respectively). On average, differences are less precise (Table 2 Column 2) but similar directionally: algorithms deteriorated the reasoning for simpler NAE decisions – which improved – by 0.12 standard deviations (p=0.6), while improving the reasoning of more complex AE decisions – which relatively worsened by 0.3 standard deviations (p-value=0.4).

These treatment effects are more precise when further separated by analyst experience, suggesting that new analysts were again more likely to be impacted. Figure 2(d)-(g) shows that treatment effects appear to be driven most by new analysts, whose reasoning for simpler decisions worsened,

while it improved for complex decisions (Kolmogorov-Smirnov test p-values = 0.02 and 0.001 for new analysts; 0.09 and 0.57 for experienced analysts, respectively). Similarly, Table 2 columns 3-4 show that causal explanations of new treated analysts received approximately 0.7 standard deviations lower reasoning scores on simpler NAE decisions relative to control analysts in the experimental period (p-value=0.06). In contrast, new analysts were more likely to see their reasoning scores improve on more complex AE funds, observing a 0.9 standard deviation higher treatment effect (p-value=0.02) – even though they were more likely to make worse rating decisions. For both decision types, more experienced analysts were again less impacted, with attenuated coefficients in the opposite direction. While we observe a weaker impact of machine predictions on experienced analysts, it is also possible that reasoning scores for experienced analysts varied less if review committees were less sensitive to differences in explanations by experienced analysts given their reputation—which we cannot fully tease apart empirically. Columns 4 and 5 of Table 2 include committee fixed effects, with Column 5 restricting to reasoning scores by single member committees, which approved the majority of ratings during this period, thus providing more precision when including committee fixed effects.[11]

In addition to reasoning scores, we analyze the text of analysts' reports to better understand how their causal explanations change (Table A.4).[12] We use natural language processing techniques to evaluate the readability and coherence of the text, and find directionally similar effects as reasoning scores, though less precise. Treated new analysts produced causal explanations with lower readability and coherence on simpler NAE funds, while showing higher readability and coherence on more complex AE funds, suggesting that they used more precise technical language and showed higher consistency in logic, in line with their higher reasoning scores.

5.3 The impact of algorithm explainability

We further evaluate how providing algorithm explainability affected analysts' decisions and causal explanations. Figure 3(b) suggests that explainability increased the perceived contribution of machine predictions for simpler decisions across all analysts. Moreover, explainability increased reliance on machine predictions especially among experienced analysts, who generally tended to rely less on machine predictions overall, consistent with prior studies (Allen and Choudhury 2022, Kim

[11]Committees with a single reviewer evaluated 20.6 ratings on average, while committees with two reviewers evaluated an average of 3.7 ratings, and committees with three and four reviewers evaluated 1.4 and 1 ratings on average, respectively.

[12]This current draft includes results on all published reports for U.S.-based funds.

et al. 2024).

However, we observe limited differential treatment effects of explainability on decisions and causal explanations (Table 4). Across all outcomes, estimates for the explainability treatment are generally qualitatively similar to those for the treatment without explainability, across both average and heterogeneous effects. This suggests that providing explainability may have limited effects on decisions and causal explanations. Of course, it is possible that the results we observe are driven by the particular method of algorithm explainability that we examine. Machine learning algorithms are black boxes that are difficult, if not impossible, to fully explain. However, the Shapley values used in our setting are one of the most commonly used of methods at the frontier of research on explainable AI so far by scholars and engineers in practice (Lundberg and Lee 2017), which seek to provide a practical way to offer goal-oriented, or "teleological," explanations to inherently uninterpretable algorithms (Tomaino et al. 2020). This suggests that explainability may not be the silver bullet as policymakers have hoped, and provide some support to recent studies that challenge the benefits of creating explanations for black box models and raise concerns that such practices may in some cases do more harm than good (Rudin 2019).

Together, these results suggest that while algorithms can help improve the decisions that they aid, they may deteriorate the causal explanation behind the decision and possibly introduce a tradeoff between the two. The findings raise the possibility that organizations may face a key challenge as they use algorithms to support managerial decisions, especially for less experienced managers. For those who make simpler decisions, algorithms may worsen their reasoning even when they improve the decision, while for those who make more complex decisions, algorithms may worsen decisions even while improving the reasoning behind them.

6 Mechanisms

These findings suggest that providing machine predictions does not always improve decisions, and that even when they do, they can worsen the decision-makers' causal explanations for the decision. Why might this be?

One possibility is that machine predictions change not only the information available but also the nature of work and the incentives involved in complex ways when they are introduced into decision processes in organizations (Athey et al. 2020). Relying on machine predictions may fundamentally change the nature of the work, by outsourcing prediction from the decision process. Analysts

need to develop a clear and coherent causal explanation without explicitly making predictions, or reconcile the causal model they have in mind with the prediction from the algorithm—which may require broader and more complex models that can integrate new considerations, or a higher ability to reason flexibly through the decision. Similarly, introducing machine predictions may affect incentives in unintended ways. If analysts are engaged in a multitasking problem where they are rewarded for both better decisions and better explanations subject to their effort constraint, we might expect that providing a machine prediction that can substitute for some of the effort in crafting a rating decision may result in improved efficiency gains, with effort reallocated to causal explanations—a cognitively more intensive task (see also Camuffo et al. (2022) for a similar argument applied to strategic decisions). However, it may be that providing better machine predictions that can substitute for effort conversely leads them to shirk overall, rather than reallocate their effort. This explanation would further extend results in Dell'Acqua (2022) that when managers perceive algorithms to be of higher quality, they decrease their efforts on the prediction task by following the algorithm—raising the possibility that they may not only shirk from the prediction task but also on adjacent tasks not directly informed by the algorithm.

This explanation would predict that treated analysts working on simpler NAE decisions—which resulted in worsened reasoning scores—were more likely to rely on algorithms than when working on more complex AE decisions. We find evidence consistent with this explanation from survey questions built in the workflow during the experiment, which provides insight into analysts' perceptions of machine predictions. We observe that treated analysts when working on simpler NAE funds stated that machine predictions contributed more to their decisions compared to decisions on more complex AE funds (Figure 3(a)). Analysts' reported contribution of machine predictions to NAE decisions (standardized) was on average 0.14 (with 95% confidence intervals between 0.02 and 0.26), while that of AE decisions was -0.19 (with 95% confidence intervals between -0.32 and -0.07), and this difference is statistically significant—suggesting that analysts relied more on algorithms when their reasoning deteriorated. Furthermore, while providing algorithm explainability has limited differential treatment effects on decisions and causal explanations overall (Table 4), it appears to have increased the likelihood that these new analysts relied on algorithms for NAE decisions (Figure 3(b) and Figure 5), suggesting that they may have relied more on algorithmic guidance for their causal explanations, potentially adopting the provided explanation in part or in full rather than their own reasoning through the decision.

We also find evidence consistent with the interpretation that these NAE analysts were less able

to reason or allocated less effort to reasoning through their causal explanations (Table A.4 Columns 10-12). The primary goal of these explanations is to identify key causal drivers behind the rating to explain and justify the decision, which can take considerable effort, and baseline data shows that higher reasoning scores are associated with fewer drivers. We find that new NAE analysts who received worse reasoning scores were also more likely to report a larger number of causal drivers compared to the control group, suggesting that they may have not been able or willing to reason deeply about the decision when they relied more on algorithms (Figure 4(d), Table 3). In comparison, for AE decisions where they relied less on machine predictions, new treated analysts were more likely to report fewer causal drivers (Figure 2(e)), consistent with the higher reasoning scores they receive. We find additional supportive evidence when counting the number of causal statements made in their report (Table 3). We find that new NAE analysts assigned to treatment who relied more on machine predictions made 3% more causal statements in their report, while new AE analysts assigned to treatment made 21% fewer causal statements. Finally, NAE analysts assigned to treatment who relied more on machine predictions and observed lower reasoning scores were more likely to write shorter explanations of their causal reasoning behind decisions as measured by word count, writing reports that were 40 words or more shorter than their counterparts assigned to the control group. Intriguingly, experienced analysts appear to have been able to identify key causal drivers (Figure 4(f) and (g)) and not increase their following of the machine prediction (Figure 5 even when they found algorithm explainability to be valuable (Figure 3(b)). This raises the possibility that experience may enable decision-makers to be better able to integrate or reconcile machine predictions into their causal models—building on prior findings that domain expertise may help decision-makers better judge predictions (Tranchero 2024).

There may also be other possible reasons. Another potential explanation may be that different fund types simply have different incentives regardless of machine predictions, and the underlying incentives were not strong enough for analysts to reallocate their effort to causal explanations on simpler NAE decisions. One may argue that causal explanations matter less for NAE funds compared to AE funds, as macroeconomic changes are more difficult to reason about. If this were the case, analysts may have been motivated to prioritize causal explanations on more complex AE decisions, thus reallocating their effort, while they were less motivated on simpler NAE decisions and thus shirked from providing good explanations.

However, we do not observe supportive evidence. This explanation would predict that analysts should have higher reasoning scores on AE decisions than NAE on average, which we do not observe

in the baseline data (Figure A.6). We also do not observe detrimental effects on causal explanations for experienced analysts on NAE decisions who appear to rely less on the machine predictions (Table 2). Furthermore, our conversations with industry experts including those within the company suggest that this may be less likely. Indeed, if analyst reasoning were less useful for NAE funds, the company would benefit from dropping their analyst coverage of these funds to increase coverage of more AE funds that they lack bandwidth to cover. Experts explained that while NAE funds on average display lower volatility, the value of analyst explanations extends beyond predicting the performance of the funds. Particularly in the context of NAE funds, keeping both explicit costs (e.g., the expense ratio) and implicit costs (e.g., the cost of portfolio turnover) at a minimum is paramount (Morningstar 2018). The landscape of NAE funds is also diverse: many do not track simple and transparent indices, and make bets against the market that may risk underperformance relative to the broader market – which increases the potential role that analyst explanations about these funds can play.

Together, these results provide some suggestive evidence that the impact of machine predictions on decisions and causal explanations—and the tradeoff between the two—may have possibly been driven by machine predictions changing the nature of the work and incentives involved, leading analysts to over-rely on machine predictions and become less able to reason through their causal explanations. They also suggest that while algorithm explainability can help the use of algorithms in decisions in some cases, it may also amplify the tradeoffs at work.

7 Discussion and conclusion

In this paper, we designed and ran a field experiment across mutual fund analysts to understand how machine predictions affect decision-makers' causal explanations in addition to the decisions they support. We find that even when algorithms improve decisions, they can deteriorate the causal explanations that decision-makers can provide for their decisions.

As firms become increasingly interested in using algorithms to support various types of managerial decisions across the organization, our findings highlight the importance of understanding how the use of algorithms affects not only the decisions they support, but also the decision-makers who use them. Despite the promise of freeing up cognitive capacity, we find that algorithms may introduce complex and potentially detrimental effects on decision-makers—especially those who are less experienced. This raises the concern that the increasing adoption of algorithms may potentially

hinder new decision-makers' capability to learn from doing and accumulate domain expertise over time. This risk of skill degradation suggests that firms interested in algorithms may need to be cautious about where and how they adopt algorithms into the organization, and highlights that understanding how the use of algorithms affects the learning of decision-makers in organizations is an important direction for future work.

A key limitation of our study is that we study decision-makers in a single organization who engage in a particular type of decision. Although we have separated decisions in our context to relatively simpler and more complex decisions, both are inherently complex and require extensive research by experts with domain expertise. Hence, the insights learned from our study may be most generalizable to other similar decisions made by expert knowledge workers. Furthermore, we are not able to observe how the effect of machine predictions evolves over the long run beyond six months of our experiment. This suggests that future work on how machine predictions impact decision-makers over a longer period and their ability to reason through different decisions in different contexts may provide valuable complementary insights.

Finally, our findings suggest that algorithm explainability may not be a silver bullet. Much research has focused on improving decision-makers' uptake of machine predictions, and highlighted the potential role that explainability may play in increasing this usage (Lebovitz et al. 2022, Kim et al. 2024). While our results provide preliminary evidence in line with this idea, they also suggest that increasing the use of algorithms is likely to come with tradeoffs and that explainability can also amplify some of their detrimental effects.

References

Acemoglu, D. and P. Restrepo (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives 33*(2), 3–30.

Agarwal, R., F. Bacco, A. Camuffo, A. Coali, A. Gambardella, H. Msangi, S. T. Sonka, A. Temu, B. Waized, and A. Wormald (2023). Does a theory-of-value add value? Evidence from a randomized control trial with Tanzanian entrepreneurs. *Evidence from a Randomized Control Trial with Tanzanian Entrepreneurs (April 6, 2023)*.

Agrawal, A., J. Gans, and A. Goldfarb (2018, May). Prediction, judgment, and complexity: A theory of decision-making and artificial intelligence. In *The Economics of Artificial Intelligence: An Agenda*, pp. 89–110. University of Chicago Press.

Agrawal, A., J. S. Gans, and A. Goldfarb (2019, June). Exploring the impact of artificial Intelligence: Prediction versus judgment. *Information Economics and Policy 47*, 1–6.

Agrawal, A., J. McHale, and A. Oettl (2018). Finding needles in haystacks: Artificial intelligence and recombinant growth. In *The economics of artificial intelligence: An agenda*, pp. 149–174. University of Chicago Press.

Allen, R. and P. R. Choudhury (2022, January). Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. *Organization Science 33*(1), 149–169.

Armstrong, W. J., E. Genc, and M. Verbeek (2019). Going for gold: An analysis of Morningstar analyst ratings. *Management Science 65*(5), 2310–2327.

Athey, S. C., K. A. Bryan, and J. S. Gans (2020, May). The allocation of decision authority to human and artificial intelligence. *AEA Papers and Proceedings 110*, 80–84.

Autor, D. and A. Salomons (2018). Is automation labor-displacing? Productivity growth, employment, and the labor share. (National Bureau of Economic Research).

Autor, D. H. (2015, August). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives 29*(3), 3–30.

Bach, L., L. E. Calvet, and P. Sodini (2020, September). Rich pickings? Risk, return, and skill in household wealth. *American Economic Review 110*(9), 2703–2747.

Barras, L., O. Scaillet, and R. Wermers (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The journal of finance 65*(1), 179–216.

Beck, M. and B. Libert (2017). The rise of AI makes emotional intelligence more important. *Harvard Business Review 15*(1-5).

Bettman, J. R. and B. A. Weitz (1983, June). Attributions in the board room: Causal reasoning in corporate annual reports. *Administrative Science Quarterly 28*(2), 165.

Billinger, S., N. Stieglitz, and T. R. Schumacher (2014, February). Search on rugged landscapes: An experimental study. *Organization Science 25*(1), 93–108.

Board of Governors of the Federal Reserve System (US) (2023, December). Mutual funds; Total financial assets, market value levels [BOGZ1LM654090000Q]. https://fred.stlouisfed.org/series/BOGZ1LM654090000Q.

Bollaert, H., F. Lopez-de Silanes, and A. Schwienbacher (2021). Fintech and access to finance. *Journal of corporate finance 68*, 101941.

Bradshaw, M. T., B. Lock, X. Wang, and D. Zhou (2021). Soft information in the financial press and analyst revisions. *The accounting review 96*(5), 107–132.

Brynjolfsson, E., W. Jin, and K. McElheran (2021). The power of prediction: predictive analytics, workplace complements, and business performance. *Business Economics 56*, 217–239.

Campbell, D., M. Loumioti, and R. Wittenberg-Moerman (2019). Making sense of soft information: Interpretation bias and loan quality. *Journal of Accounting and Economics 68*(2-3), 101240.

Camuffo, A., A. Cordova, A. Gambardella, and C. Spina (2020, February). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science 66*(2), 564–586.

Camuffo, A., A. Gambardella, and A. Pignataro (2022). Framing strategic decisions in the digital world. *Strategic Management Review*. Forthcoming.

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

Choudhury, P., E. Starr, and R. Agarwal (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal 41*(8), 1381–1411.

Coleman, M. and T. L. Liau (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology 60*(2), 283.

Cornaggia, J., K. J. Cornaggia, and H. Xia (2016). Revolving doors on wall street. *Journal of Financial Economics 120*(2), 400–419.

Cowgill, B. (2019). Bias and productivity in humans and machines. *SSRN Electronic Journal*.

Csaszar, F. A. (2018, December). What makes a decision strategic? strategic representations. *Strategy Science 3*(4), 606–619.

Dell'Acqua, F. (2022). Falling asleep at the wheel: Human/AI collaboration in a field experiment on HR recruiters. Working Paper. Havard Business School, Boston.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics 132*(4), 1593–1640.

Dietvorst, B. J., J. P. Simmons, and C. Massey (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General 144*(1), 114–126.

Ehrig, T. and J. Schmidt (2021, May). Making biased but better predictions: The trade-offs strategists face when they learn and use heuristics. *Strategic Organization 19*(2), 263–284.

Ertugrul, M. and S. Hegde (2009, March). Corporate governance ratings and firm performance. *Financial Management 38*(1), 139–160.

Ethiraj, S. K. and D. Levinthal (2004, February). Modularity and innovation in complex systems. *Management Science 50*(2), 159–173.

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

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of financial economics 116*(1), 1–22.

Felin, T. and T. R. Zenger (2009). Entrepreneurs as theorists: on the origins of collective beliefs and novel strategies. *Strategic Entrepreneurship Journal 3*(2), 127–146.

Felin, T. and T. R. Zenger (2017, December). The theory-based view: economic actors as theorists. *Strategy Science 2*(4), 258–271.

Felten, E., M. Raj, and R. Seamans (2021). Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses. *Strategic Management Journal 42*(12), 2195–2217.

Fernandez, R. and D. Rodrik (1991). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty. *The American economic review*, 1146–1155.

Fiss, P. C. and E. J. Zajac (2006). The symbolic management of strategic change: Sensegiving via framing and decoupling. *Academy of management journal 49*(6), 1173–1193.

Frey, C. B. and M. A. Osborne (2017). The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change 114*, 254–280.

Goetzmann, W. N. and N. Peles (1997). Cognitive dissonance and mutual fund investors. *Journal of financial Research 20*(2), 145–158.

Grennan, J. and R. Michaely (2020). Artificial intelligence and high-skilled work: Evidence from analysts. *Swiss Finance Institute Research Paper* (20-84).

Gruber, M. J. (1996). Another puzzle: The growth in actively managed mutual funds. *The journal of finance 51*(3), 783–810.

Helfat, C. E. and M. A. Peteraf (2015, June). Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal 36*(6), 831–850.

Hoffman, M., L. B. Kahn, and D. Li (2018, May). Discretion in hiring. *The Quarterly Journal of Economics 133*(2), 765–800.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. *The Journal of finance 23*(2), 389–416.

Kang, X. (2023). Humans and algorithms: Knowledge workers' adaptation to artificial intelligence. Working Paper. Vanderbilt University, Nashville.

Kempf, E. (2020). The job rating game: Revolving doors and analyst incentives. *Journal of Financial Economics 135*(1), 41–67.

Kim, H., E. L. Glaeser, A. Hillis, S. D. Kominers, and M. Luca (2024). Decision authority and the returns to algorithms. *Strategic Management Journal 45*(4), 619–648.

Kincaid, J. P., R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Institute for Simulation and Training, University of Central Florida.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2017, August). Human decisions and machine predictions. *The Quarterly Journal of Economics*.

Lebovitz, S., H. Lifshitz-Assaf, and N. Levina (2022, January). To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis. *Organization Science 33*(1), 126–148.

Lou, B. and L. Wu (2021, September). AI on drugs: Can artificial intelligence accelerate drug development? Evidence from a large-scale examination of bio-pharma firms. *MIS Quarterly 45*(3), 1451–1482.

Lourie, B. (2019). The revolving door of sell-side analysts. *The Accounting Review 94*(1), 249–270.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems 30*.

Luo, X., M. S. Qin, Z. Fang, and Z. Qu (2021). Artificial intelligence coaches for sales agents: Caveats and solutions. *Journal of Marketing 85*(2), 14–32.

Luo, X., H. Wang, S. Raithel, and Q. Zheng (2015, January). Corporate social performance, analyst stock recommendations, and firm future returns: Research Notes and Commentaries. *Strategic Management Journal 36*(1), 123–136.

MacLennan, A. F. and C. C. Markides (2021, August). Causal mapping for strategy execution: Pitfalls and applications. *California Management Review 63*(4), 89–122.

Mercier, H. and D. Sperber (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and brain sciences 34*(2), 57–74.

Morningstar (2018). Morningstar's guide to passive investing. Report, Morningstar.

Pfeffer, J. (1981). Management as symbolic action: the creation and maintenance of organizational paradigm. *Research in organizational behavior 3*, 1–52.

Raisch, S. and S. Krakowski (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of management review 46*(1), 192–210.

Rudin, C. (2019, May). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence 1*(5), 206–215.

Röder, M., A. Both, and A. Hinneburg (2015). Exploring the space of topic coherence measures. pp. 399–408.

Samuelson, W. and R. Zeckhauser (1988). Status quo bias in decision making. *Journal of risk and uncertainty 1*, 7–59.

Schmitt, W. (2024, 18 Jan). Passive eclipses active in us fund market as assets swell to $13.3tn. *Financial Times*. Available at: `https://www.ft.com/content/faf74f66-c4d6-45aa-bb30-0e73d523c547` (Accessed: March 1st, 2024).

Shaw, W. and A. Gani (2023, May). Wall street banks are using AI to rewire the world of finance. *Bloomberg.com*.

Shrestha, Y. R., S. M. Ben-Menahem, and G. Von Krogh (2019). Organizational decision-making structures in the age of artificial intelligence. *California management review 61*(4), 66–83.

Siggelkow, N. (2002, July). Misperceiving interactions among complements and substitutes: Organizational consequences. *Management Science 48*(7), 900–916.

Simon, H. (1947). *Administrative behavior: A study of decision-making processes in administrative organizations.* New York: Macmillan.

Sirri, E. R. and P. Tufano (1998). Costly search and mutual fund flows. *The journal of finance 53*(5), 1589–1622.

Sørensen, J. B. and G. R. Carroll (2021, June). Why Good Arguments Make Better Strategy. *MIT Sloan Management Review 62*(4).

Soydemir, G., J. Smolarski, and S. Shin (2014). Hedge funds, fund attributes and risk adjusted returns. *Journal of Economics and Finance 38*, 133–149.

Tomaino, G., H. Abdulhalim, P. Kireyev, and K. Wertenbroch (2020). Denied by an (Unexplainable) Algorithm: Teleological Explanations for Algorithmic Decisions Enhance Customer Satisfaction. *SSRN Electronic Journal*.

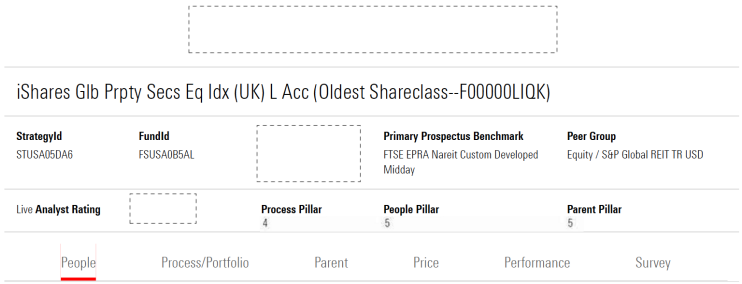
Tong, S., N. Jia, X. Luo, and Z. Fang (2021). The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. *Strategic Management Journal 42*(9), 1600–1631.

Tranchero, M. (2024, August). Finding Diamonds in the Rough: Data-Driven Opportunities and Pharmaceutical Innovation. *Academy of Management Proceedings 2024*(1), 13751.

Figures and Tables

Figure 1: Design of control and treatment conditions

(a) Control



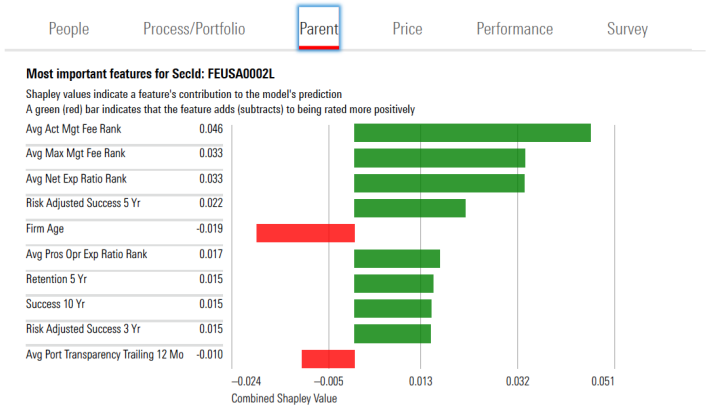
(b.1) Treatment



(b.2) Treatment cont.



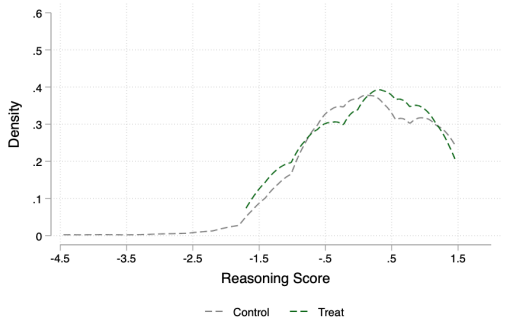
(c) Algorithm explainability using Shapley values



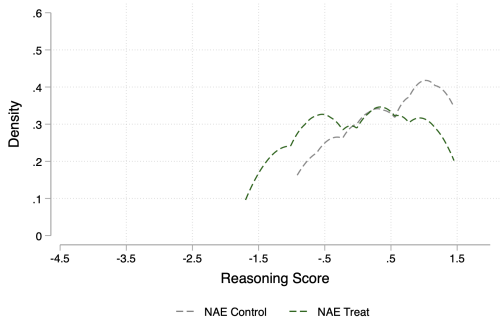
Notes: This set of figures shows what analysts saw on the research platform across different experimental conditions. The control group continued to do research as usual, and saw no changes on the interface ((a)). Analysts assigned to either treatment group saw an additional line of information at the top of the interface that provided a machine prediction for the fund ((b.1-2)). (a) shows an example of a non-actively managed fund, while (b) shows an example of an active equity fund. Analysts in treatment group 1 received algorithmic predictions only, while treatment 2 additionally saw a figure that highlighted the ten key features that contributed most to the algorithm's prediction for that fund ((c)), which was based on Shapley values.

Figure 2: Distributions of reasoning scores by treatment

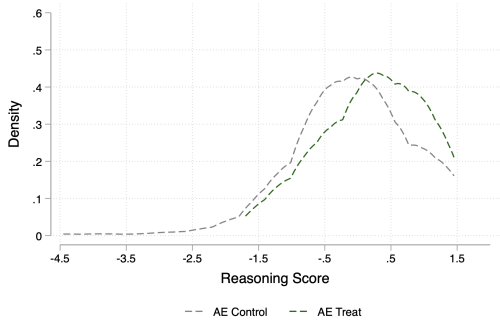
(a) Average treatment effects



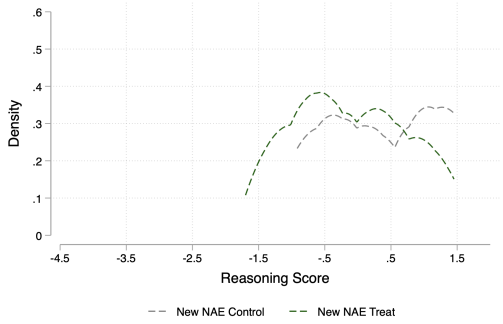
(b) Simpler decisions



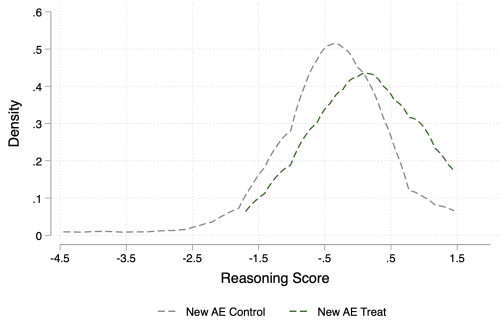
(c) Complex decisions



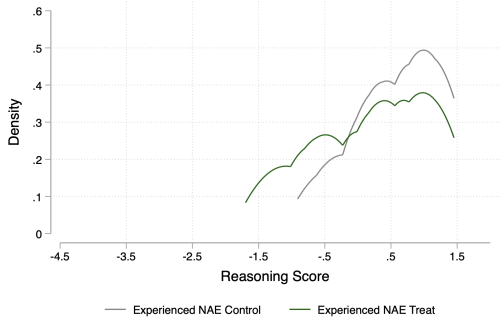
(d) New analysts on simpler decisions



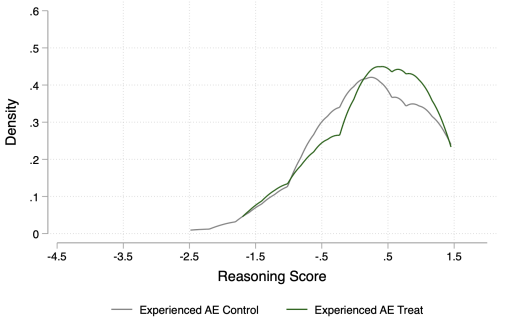
(e) New analysts on complex decisions



(f) Experienced analysts on simpler decisions



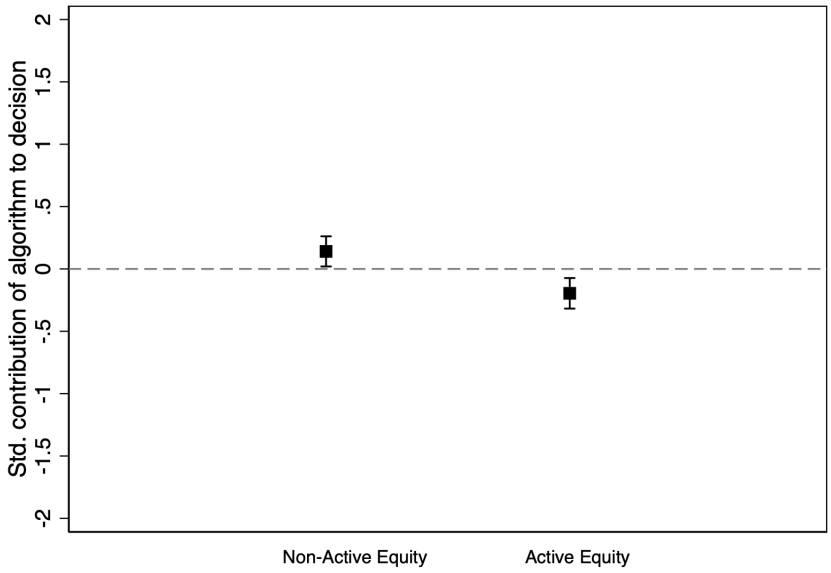
(g) Experienced analysts on complex decisions



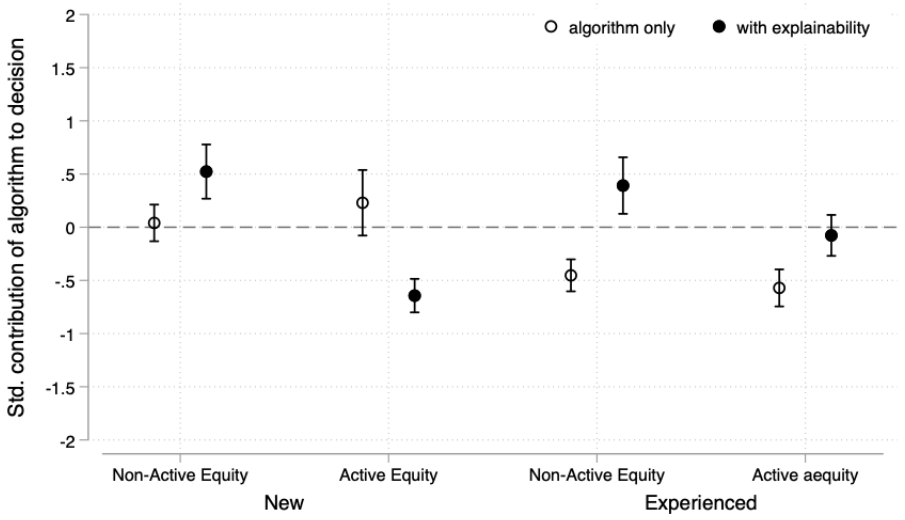
Notes: These figures show kernel density plots of reasoning scores by treatment and control conditions on average ((a)) and by subgroups ((b)-(g)). Kolmogorov-Smirnov test *p-values*: (a) 0.316, (b) 0.002, (c) <0.001, (d) 0.011, (e) 0.001, (f) 0.075, (g) 0.529.

Figure 3: Analyst perceptions of how machine predictions contributed to their decisions

(a) Simpler vs. more complex decisions



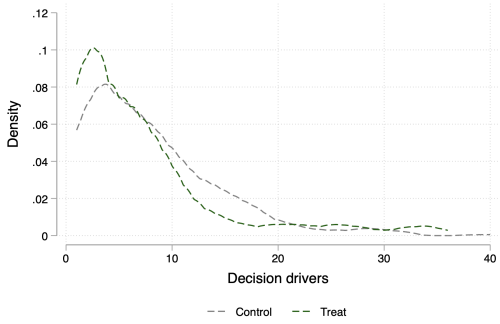
(b) Responses by decision type, analyst experience, and algorithm explainability



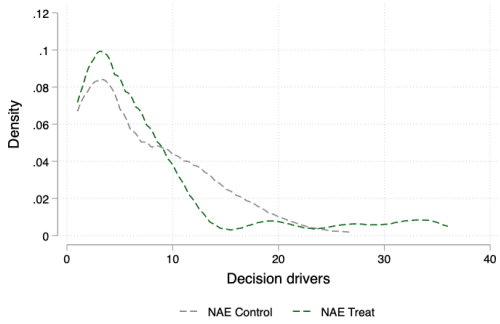
Notes: These figures show means and confidence intervals of analysts' reported contribution of machine predictions to their decisions. Analysts reported the extent to which machine predictions contributed to their rating decisions on a scale between 1 and 10. (a) shows the differential levels of perceived contribution across NAE and AE analysts, and (b) additionally presents the differential levels of contribution by analyst experience and algorithm explainability.

Figure 4: Distributions of decision drivers by treatment

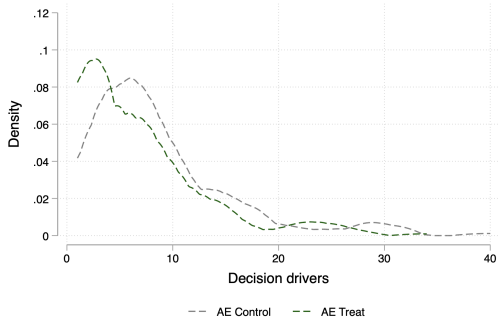
(a) Average treatment effects



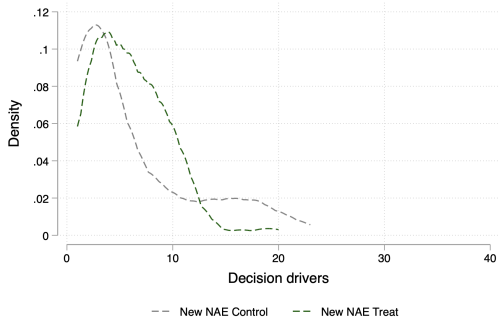
(b) Simpler decisions



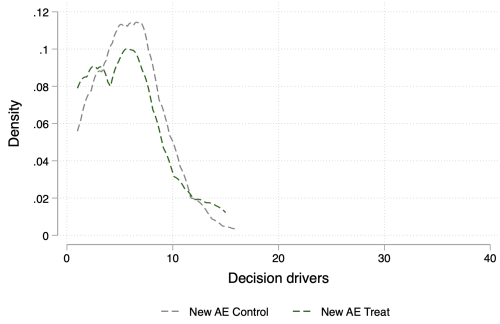
(c) Complex decisions



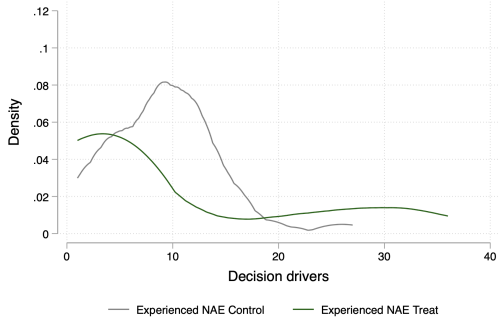
(d) New analysts on simpler decisions



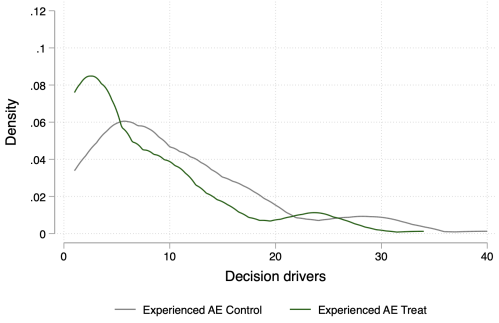
(e) New analysts on complex decisions



(f) Experienced analysts on simpler decisions

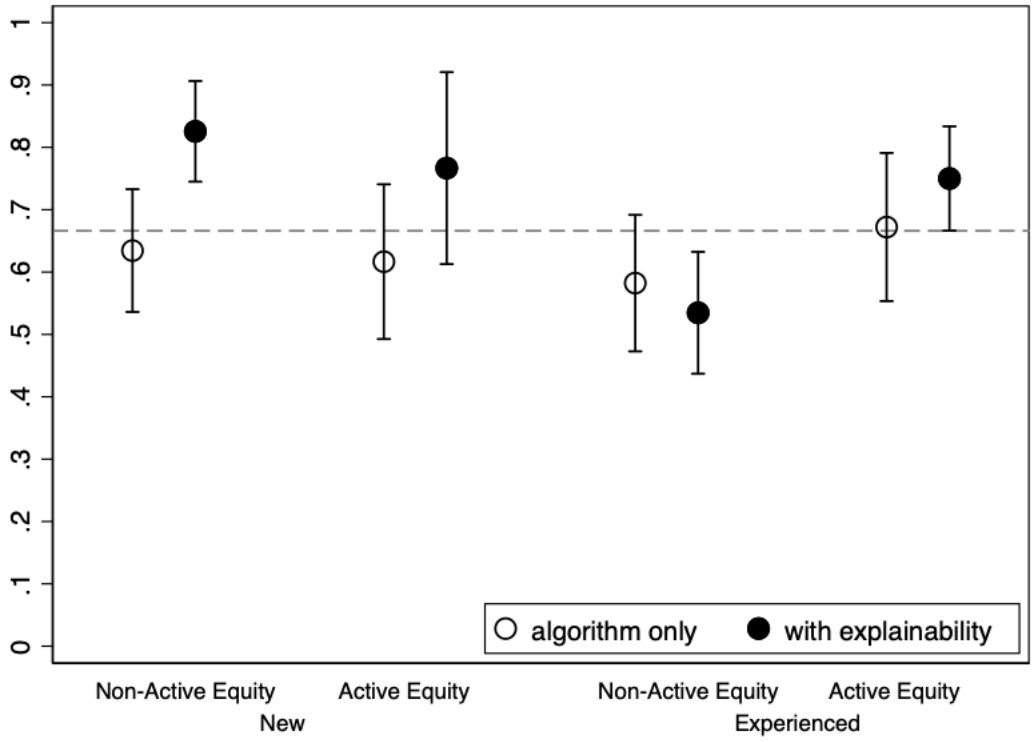


(g) Experienced analysts on complex decisions



Notes: These figures show kernel density plots of the number of decision drivers by treatment and control conditions on average ((a)) and by subgroups ((b)-(g)). Kolmogorov-Smirnov test *p-values*: (a) 0.009, (b) 0.218, (c) <0.001, (d) 0.017, (e) 0.183, (f) <0.001, (g) <0.001.

Figure 5: Share of analyst decisions that matched machine predictions



Notes: This figure shows the percentage of analyst decisions that matched machine predictions by fund type, analyst experience, and machine explainability.

Table 1: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating					
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.074**	0.075**	0.080**	0.080**	0.080*	0.114**
	(0.033)	(0.033)	(0.031)	(0.031)	(0.043)	(0.057)
Treat × Post × Active Equity					-0.000	-0.039
					(0.071)	(0.122)
Treat × Post × Experienced						-0.042
						(0.084)
Treat × Post × Active Equity × Experienced						0.043
						(0.151)
Strata FE	No	Yes	No	Yes	Yes	Yes
Month FE	No	No	Yes	Yes	Yes	Yes
Observations	1780	1780	1780	1780	1780	1780
Mean (control)	0.113	0.113	0.113	0.113	0.113	0.113
SD (control)	0.317	0.317	0.317	0.317	0.317	0.317

Panel B. Decision performance

	Fund returns in 3 months			Fund returns in 6 months		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post × Recommend	0.894	9.886***	11.557***	1.130	7.949***	6.741*
	(3.033)	(2.585)	(3.600)	(2.165)	(2.487)	(3.761)
Treat × Post × Recommend × Active Equity		-16.822***	-16.747**		-11.458***	-8.323
		(5.268)	(7.857)		(4.191)	(6.471)
Treat × Post × Recommend × Experienced			-5.959			-0.604
			(5.031)			(4.510)
Treat × Post × Recommend × Active Equity × Experienced			5.281			-0.092
			(10.532)			(8.418)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1668	1668	1668
Mean (control)	-4.467	-4.467	-4.467	-4.832	-4.832	-4.832
SD (control)	13.865	13.865	13.865	11.525	11.525	11.525

Notes: Panel A shows treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on *Fund returns in 3 months*, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on *Fund returns in 6 months*, the risk-adjusted return of a fund in 6 months after its rating is published. *Treatment* is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. All regressions include a full set of strata and month fixed effects, and regressions in Panel B additionally control for fund age and size. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 2: The impact of machine predictions on reasoning

	All committees				Single
	(1)	(2)	(3)	(4)	(5)
Treat × Post	0.012	-0.124	-0.661*	-0.564*	-0.698**
	(0.169)	(0.210)	(0.345)	(0.322)	(0.339)
Treat × Post × Active Equity		0.297	0.907**	0.609	1.151**
		(0.351)	(0.395)	(0.387)	(0.446)
Treat × Post × Experienced			0.828*	0.668*	0.786*
			(0.438)	(0.374)	(0.398)
Treat × Post × Active Equity × Experienced			-0.950	-0.796*	-1.293**
			(0.607)	(0.465)	(0.529)
Strata FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	Yes
Observations	1313	1313	1313	1207	901
Mean (control)	0.238	0.238	0.238	0.238	0.286
SD (control)	0.928	0.928	0.928	0.928	0.968

Notes: All regressions include a full set of strata and month fixed effects. Regressions in columns (4)-(5) further include review committee fixed effects. *Reasoning score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. Regressions in columns (1)-(4) include committees of all sizes, which vary between 1 reviewer and 4 reviewers per committee. Column (5) analyzes the subsample of committees with one reviewer only. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 3: The number of decision drivers and causal statements by experimental condition

	Decision drivers			Causal statements		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	-0.332*	0.033	0.478	-0.065	-0.114**	0.026
	(0.187)	(0.259)	(0.321)	(0.040)	(0.052)	(0.082)
Treat × Post × Active Equity		-0.610*	-1.378**		0.110	-0.194*
		(0.352)	(0.542)		(0.080)	(0.111)
Treat × Post × Experienced			-0.893*			-0.240**
			(0.459)			(0.104)
Treat × Post × Active Equity × Experienced			1.331**			0.496***
			(0.669)			(0.142)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1317	1317	1317	948	948	948
Mean (control)	1.728	1.728	1.728	2.104	2.104	2.104
SD (control)	0.896	0.896	0.896	0.274	0.274	0.274

Notes: All regressions include a full set of strata and month fixed effects. *Decision drivers* measure the number of reported decision drivers by the analyst and is natural log transformed. *Causal statements* measure the number of causal statements in the analyst report as identified by GPT-4 and is natural log transformed. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. Post takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 4: The impact of algorithm explainability on decisions and reasoning

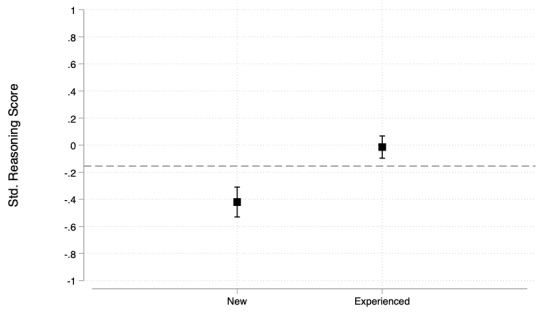
Panel A: Decision change						
	Change rating			Reasoning score		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.120***	0.097**	0.081	0.005	-0.362	-0.788**
	(0.041)	(0.048)	(0.063)	(0.235)	(0.223)	(0.358)
Explainability Treat × Post	0.040	0.062	0.151**	0.020	0.112	-0.554
	(0.034)	(0.050)	(0.064)	(0.194)	(0.278)	(0.407)
Treat × Post × Active Equity		0.057	0.005		0.736	0.981**
		(0.098)	(0.152)		(0.453)	(0.434)
Explainability Treat × Post × Active Equity		-0.049	-0.025		-0.181	0.984**
		(0.069)	(0.093)		(0.379)	(0.391)
Treat × Post × Experienced			0.063			0.496
			(0.093)			(0.424)
Explainability Treat × Post × Experienced			-0.150			1.099*
			(0.094)			(0.556)
Treat × Post × Active Equity × Experienced			0.051			-0.246
			(0.186)			(0.804)
Explainability Treat × Post × Active Equity × Experienced			0.017			-1.666***
			(0.128)			(0.626)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1780	1780	1780	1313	1313	1313
Panel B: Decision performance						
	Fund returns in 3 months			Fund returns in 6 months		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post × Recommend	0.002	10.461***	15.499***	1.834	8.679***	9.799**
	(4.065)	(3.656)	(4.443)	(2.547)	(3.195)	(4.478)
Explainability Treat × Post × Recommend	2.235	10.209***	11.637**	0.322	7.318**	3.152
	(3.357)	(3.347)	(5.127)	(3.118)	(3.272)	(4.878)
Treat × Post × Recommend × Active Equity		-18.858***	-20.091**		-10.709**	-12.275
		(6.436)	(9.755)		(5.026)	(7.462)
Explainability Treat × Post × Recommend × Active Equity		-15.289**	-19.980***		-12.109**	-2.050
		(6.671)	(6.769)		(6.028)	(7.529)
Treat × Post × Recommend × Experienced			-10.428*			-3.031
			(6.013)			(5.443)
Explainability Treat × Post × Recommend × Experienced			-7.221			0.178
			(8.201)			(6.744)
Treat × Post × Recommend × Active Equity × Experienced			5.981			6.060
			(11.883)			(9.287)
Explainability Treat × Post × Recommend × Active Equity × Experienced			12.747			-5.065
			(11.874)			(11.021)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1668	1668	1668

Notes: *Treat* is an indicator variable that takes value 1 for analysts assigned to treatment 1 who receive only the algorithmic prediction. *Explainability* is a dummy variable that takes the value 1 for analysts assigned to treatment 2 who receive algorithmic predictions along with key fund features that contributed most to the prediction based on Shapley values. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. All regressions include a full set of strata and month fixed effects, and regressions in Panel B additionally control for fund age and size. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

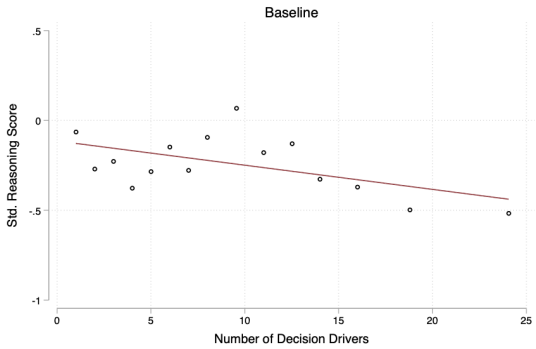
Appendix

Figure A.1: Baseline descriptives on causal explanations

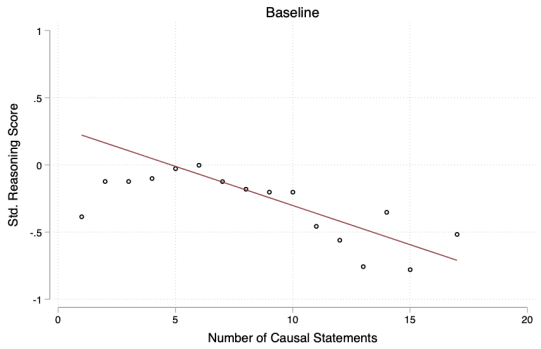
(a) Baseline reasoning scores by analyst experience



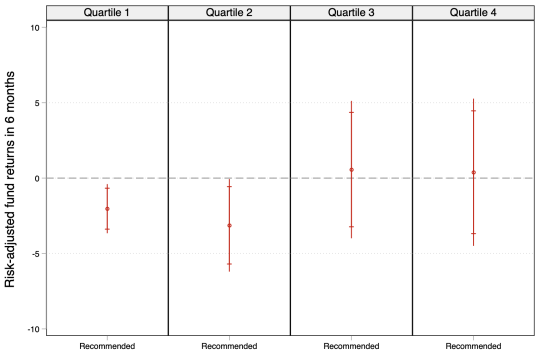
(b) Baseline correlation between reasoning scores and decision drivers



(c) Baseline correlation between reasoning scores and causal statements

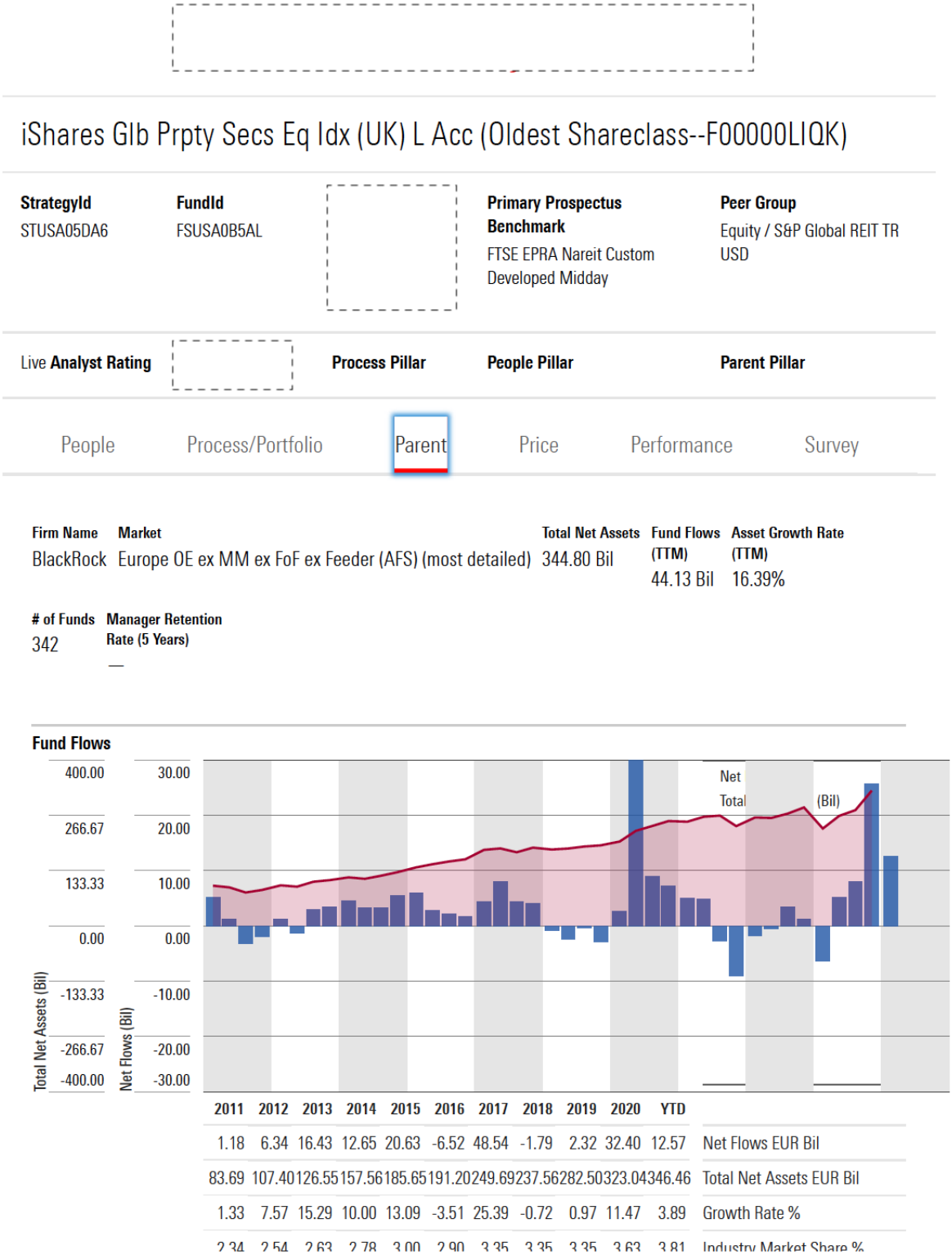


(d) Fund returns by quartiles of reasoning scores



Notes: Figure (a) shows the relationship between reasoning scores and the number of decision drivers at baseline. *Reasoning score* is constructed by standardizing reasoning scores (evaluated on a scale of 1 to 10) across individual expert committee members and taking an average across them for each fund. The *number of decision drivers* counts the number of factors that analysts report as driving their rating decisions. Figure (b) shows the average reasoning score by analyst experience. Experienced analysts are those that have more than three years of rating experience at the company.

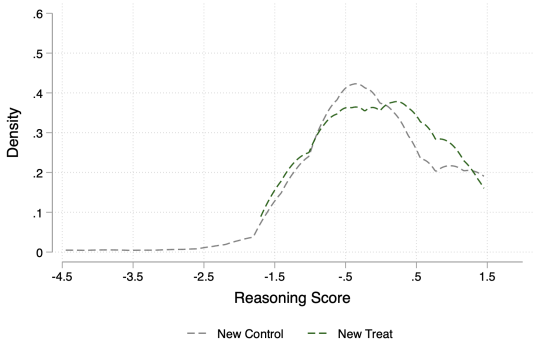
Figure A.2: Online interface for analysts



Notes: This figure shows a screenshot of the online research interface where analysts performed their work. Analysts saw the fund's currently "live" rating from the last evaluation and received information about the fund for their analysis. On this interface, analysts decided on an up-to-date rating for the fund and delineated their causal explanation for the rating.

Figure A.3: Distributions of reasoning scores by treatment and analyst tenure

(a) New analysts



(b) Experienced analysts

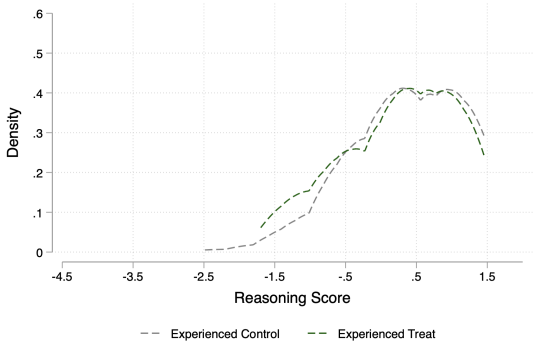
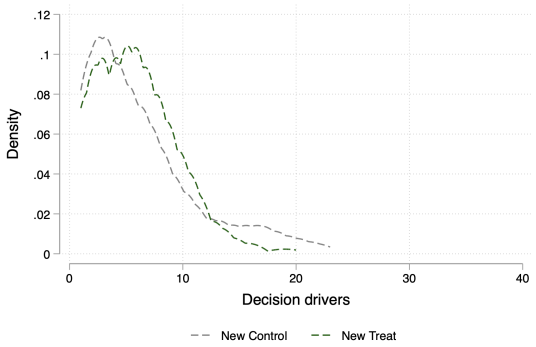


Figure A.4: Distributions of decision drivers by treatment and analyst tenure

(a) New analysts



(b) Experienced analysts

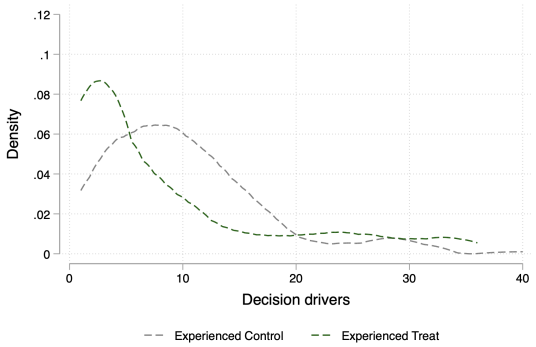
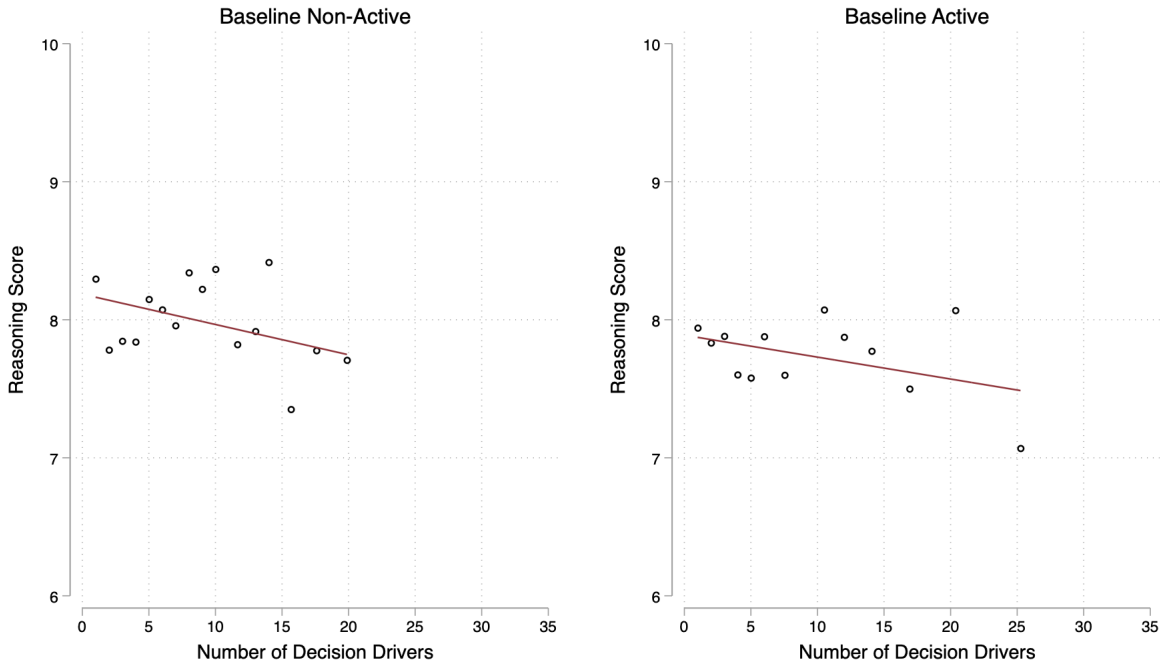
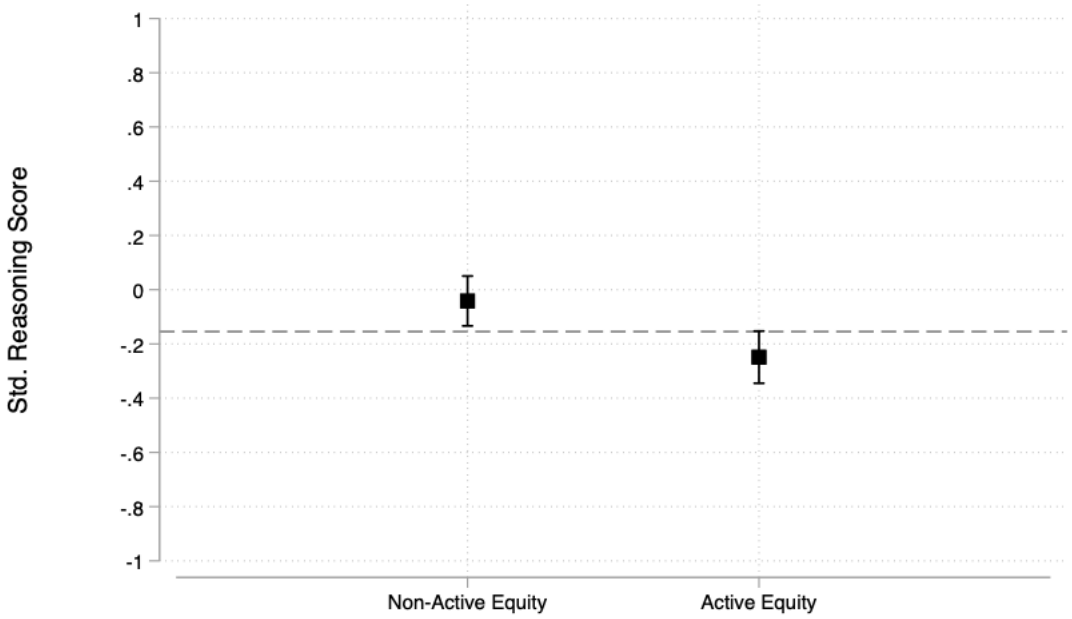


Figure A.5: Baseline correlation between reasoning scores and decision drivers by fund type



Notes: The figures show binned scatterplots of reasoning scores on the number of decision drivers at baseline by fund type (active equity or non-active equity funds). *Reasoning score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. The *number of decision drivers* counts the number of factors that analysts report as driving their rating decisions.

Figure A.6: Baseline reasoning scores by fund type



Notes: This figure shows means and 95% confidence intervals of reasoning scores by fund type at baseline. *Reasoning score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. Fund type is separated into active equity funds and non-active equity funds.

Table A.1: Summary statistics and balance of variables

Panel A. Analyst fund coverage

	Full Sample				Control	Treatment	Difference
	Mean	SD	Min.	Max.	Mean	Mean	p-value
Percentage of active equity funds	0.483	0.468	0.000	1.000	0.498	0.466	0.372
Percentage of fund recommendations	0.550	0.301	0.000	1.000	0.540	0.561	0.630
Average fund age	19.743	9.254	6.622	71.540	20.642	18.762	0.166
Average fund size (billion)	2.627	3.741	0.000	20.879	2.732	2.509	0.391
Average fund returns in 3 months	-4.131	8.700	-33.434	31.432	-5.315	-2.840	0.913
Average fund returns in 6 months	-4.947	6.296	-19.003	14.749	-5.143	-4.732	0.622

Panel B. Analyst characteristics

	Full Sample				Control	Treatment	Difference
	Mean	SD	Min.	Max.	Mean	Mean	p-value
Gender	0.701	0.460	0.000	1.000	0.673	0.729	0.723
Tenure (log)	1.562	0.933	-0.436	3.384	1.568	1.556	0.474
Time in current position (log)	1.419	0.932	-0.436	3.384	1.406	1.431	0.552
Manager	0.093	0.292	0.000	1.000	0.102	0.083	0.377
Professional certificate	0.309	0.465	0.000	1.000	0.347	0.271	0.211
Postgrad	0.433	0.498	0.000	1.000	0.429	0.438	0.535
Observations	97				49	48	97

Notes: Panel A shows the summary statistics and balance of variables about the analysts' fund coverage. The *percentage of active equity funds* indicates the share of active equity funds covered by an analyst, which ranges between 0 and 1. The *percentage of fund recommendations* measures the percentage of funds that an analyst recommends out of all funds covered. The *average fund age* is the average age of funds covered by an analyst. The *average fund size* is the average size of funds covered by an analyst. The *average fund return in 3 months* is the average of risk-adjusted return of funds covered by an analyst 3 months after their ratings are published. The *average fund return in 6 months* is the average of risk-adjusted return of funds covered by an analyst 6 months after their ratings are published.
Panel B shows the summary statistics and balance of variables about analyst characteristics. *Gender* takes value 1 if the analyst is male and 0 if the analyst is female. *Tenure* measures the number of years since the analyst join the firm and is natural log transformed. *Time in current position* measures the number of years that the analyst has held the current position and is natural log transformed. *Manager* is an indicator variable that takes the value 1 if the analyst holds an managerial position. *Professional certificate* is an indicator variable that takes the value 1 if the analyst holds any professional certificate, such as CPA, CFA, CMA, etc. *Postgrad* is an indicator variable that takes the value 1 if the analyst has postgraduate degrees, including Master's degrees and Ph.D. degrees. The number of funds covered measures the workload of an analyst by counting the number of funds they cover during the experiment.

Table A.2: Details on missing data

	Treatment		Control		p-value
	Number of observations	% of missing observations	Number of observations	% of missing observations	
Reasoning Score	608	32.294	664	24.717	0.258
Fund Returns in 3 Months	892	0.668	873	1.02	0.842
Fund Returns in 6 Months	892	0.668	870	1.361	0.687

Notes: The *reasoning score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. The *fund returns in 3 months* measures the risk-adjusted return of a fund 3 months after its rating is published. The *fund returns in 6 months* measures the risk-adjusted return of a fund 6 months after its rating is published.

Table A.3: Separating knowledge and reasoning components of reasoning scores

	Knowledge component score				Logic component score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat × Post	0.004	-0.172	-0.742**	-0.638**	0.021	-0.076	-0.579*	-0.490
	(0.169)	(0.215)	(0.352)	(0.315)	(0.173)	(0.210)	(0.345)	(0.337)
Treat × Post × Active Equity		0.357	1.020**	0.706*		0.237	0.794**	0.513
		(0.353)	(0.417)	(0.403)		(0.355)	(0.387)	(0.380)
Treat × Post × Experienced			0.853*	0.679*			0.802*	0.656*
			(0.444)	(0.369)			(0.440)	(0.391)
Treat × Post × Active Equity × Experienced			-1.006	-0.796*			-0.895	-0.796*
			(0.617)	(0.478)			(0.611)	(0.468)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	No	No	No	Yes
Observations	1313	1313	1313	1207	1313	1313	1313	1207
Mean (control)	0.220	0.220	0.220	0.220	0.256	0.256	0.256	0.256
SD (control)	0.975	0.975	0.975	0.975	0.933	0.933	0.933	0.933

Notes: All regressions include a full set of strata and month fixed effects. Regressions in columns (4) and (8) further include review committee fixed effects. The *knowledge component scores* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. The *reasoning component score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise.
Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.4: Text analysis of analyst reports

	Flesch-Kincaid			Coleman-Liau			Coherence			Word count		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treat × Post	0.052	0.006	-0.475	-0.012	-0.056	-0.463*	-0.005	-0.005	-0.020	-29.273*	-40.091**	-43.299
	(0.153)	(0.242)	(0.399)	(0.118)	(0.177)	(0.259)	(0.005)	(0.007)	(0.014)	(15.070)	(17.768)	(26.702)
Treat × Post × Active Equity		0.042	0.829		0.066	0.754**		0.009	0.028		-0.711	14.184
		(0.321)	(0.499)		(0.267)	(0.323)		(0.009)	(0.018)		(28.413)	(32.799)
Treat × Post × Experienced			0.841*			0.641*			0.027*			23.788
			(0.501)			(0.336)			(0.015)			(35.630)
Treat × Post × Active Equity × Experienced			-1.336**			-1.075**			-0.036*			-29.261
			(0.636)			(0.464)			(0.020)			(48.683)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	984	984	984	984	984	984	984	984	984	984	984	984
Mean (control)	11.193	11.193	11.193	13.632	13.632	13.632	0.233	0.233	0.233	1533.126	1533.126	1533.126
SD (control)	1.085	1.085	1.085	0.777	0.777	0.777	0.037	0.037	0.037	86.421	86.421	86.421

Notes: All regressions include a full set of strata and month fixed effects. For all measures except Word count, we first processed the text by lemmatizing and removing stop words before constructing the variables. The *Flesch-Kincaid score* in columns (1)-(3) captures the readability of reports based on the number of words in sentences and the average number of syllables per word. The *Coleman-Liau index* in columns (4)-(6) provides another measure of readability based on sentence length and the average number of characters per word. *Coherence* in columns (7)-(9) measures the coherence of topics in the report based on topic modeling. Word count in columns (10)-(12) measures the number of words in the analyst report. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.5: The differential effect of explainability on decisions and reasoning

Panel A. Decision change

	Change rating			Reasoning score		
	(1)	(2)	(3)	(4)	(5)	(6)
Explainability × Post	-0.077*	-0.030	0.066	0.192	0.407*	0.175
	(0.044)	(0.048)	(0.062)	(0.171)	(0.228)	(0.287)
Explainability × Post × Active Equity		-0.115	-0.076		-0.418	0.703**
		(0.098)	(0.145)		(0.349)	(0.327)
Explainability × Post × Experienced			-0.197**			0.578
			(0.087)			(0.404)
Explainability × Post × Active Equity × Experienced			0.011			-1.533**
			(0.176)			(0.603)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	Yes	Yes
Observations	898	898	898	549	549	549
Mean (control)	0.163	0.163	0.163	0.127	0.127	0.127
SD (control)	0.370	0.370	0.370	0.887	0.887	0.887

Panel B. Decision performance

	Fund returns in 3 months			Fund returns in 6 months		
	(1)	(2)	(3)	(4)	(5)	(6)
Explainability × Post × Recommend	3.080	0.700	-2.311	-0.328	0.009	-5.087
	(4.422)	(4.987)	(5.423)	(3.358)	(4.089)	(4.491)
Explainability × Post × Recommend × Active Equity		3.345	0.583		-2.083	10.452
		(8.121)	(10.804)		(6.754)	(8.235)
Explainability × Post × Recommend × Experienced			-0.651			-0.146
			(9.743)			(7.361)
Explainability × Post × Recommend × Active Equity × Experienced			8.986			-9.248
			(14.636)			(12.236)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	823	823	823	823	823	823
Mean (control)	-4.467	-4.467	-4.467	-4.832	-4.832	-4.832
SD (control)	13.865	13.865	13.865	11.525	11.525	11.525

Notes: Columns (1)-(3) of Panel A show treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Columns (4)-(6) of Panel A show treatment effects on *reasoning score*, constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. Columns (1)-(3) of Panel B show treatment effects on *Fund returns in 3 months*, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on *Fund returns in 6 months*, the risk-adjusted return of a fund in 6 months after its rating is published. *Explainability* is an indicator variable that takes the value 1 for analysts assigned to the treatment 2 and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise.
Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.6: Separating knowledge and reasoning components by algorithm explainability

	Knowledge component score				Logic component score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treat × Post	-0.009	-0.383	-0.877**	-0.957***	0.019	-0.342	-0.699*	-0.808**
	(0.229)	(0.233)	(0.361)	(0.364)	(0.245)	(0.218)	(0.361)	(0.373)
Explainability × Post	0.018	0.038	-0.632	-0.384	0.023	0.185	-0.476	-0.218
	(0.195)	(0.274)	(0.406)	(0.268)	(0.197)	(0.287)	(0.415)	(0.332)
Treat × Post × Active Equity		0.735	1.058**	0.851*		0.738	0.903**	0.697
		(0.454)	(0.452)	(0.445)		(0.458)	(0.426)	(0.432)
Explainability × Post × Active Equity		-0.056	1.245***	1.272***		-0.306	0.723*	0.966**
		(0.386)	(0.403)	(0.318)		(0.380)	(0.400)	(0.415)
Treat × Post × Experienced			0.581	0.840**			0.411	0.714*
			(0.435)	(0.418)			(0.423)	(0.426)
Explainability × Post × Experienced			1.073*	0.517			1.126*	0.570
			(0.550)	(0.377)			(0.573)	(0.435)
Treat × Post × Active Equity × Experienced			-0.369	-0.710			-0.124	-0.590
			(0.803)	(0.571)			(0.817)	(0.599)
Explainability × Post × Active Equity × Experienced			-1.791***	-1.524***			-1.540**	-1.533***
			(0.630)	(0.466)			(0.642)	(0.539)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	No	No	No	Yes
Observations	1313	1313	1313	1207	1313	1313	1313	1207
Mean (control)	0.220	0.220	0.220	0.220	0.256	0.256	0.256	0.256
SD (control)	0.975	0.975	0.975	0.975	0.933	0.933	0.933	0.933

	Knowledge component score				Logic component score			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Explainability × Post	0.121	0.525*	0.127	0.149	0.095	0.633**	0.221	0.200
	(0.269)	(0.283)	(0.398)	(0.247)	(0.291)	(0.306)	(0.435)	(0.345)
Explainability × Post × Active Equity		-0.848	0.464	0.716**		-1.110**	-0.026	0.689*
		(0.506)	(0.472)	(0.311)		(0.530)	(0.510)	(0.389)
Explainability × Post × Experienced			0.841	0.558			0.890	0.597
			(0.555)	(0.374)			(0.605)	(0.453)
Explainability × Post × Active Equity × Experienced			-2.027**	-1.444**			-1.832*	-1.623**
			(0.883)	(0.559)			(0.987)	(0.674)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	No	No	No	Yes
Observations	629	629	629	549	629	629	629	549
Mean (control)	0.104	0.104	0.104	0.104	0.150	0.150	0.150	0.150
SD (control)	0.927	0.927	0.927	0.927	0.899	0.899	0.899	0.899

Notes: The *knowledge component scores* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. The *reasoning component score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. *Treat* is a dummy variable that takes the value 1 for analysts assigned to treatment 1 and 0 otherwise. *Explainability* is a dummy variable that takes the value 1 for analysts assigned to treatment 2 and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.7: Text analysis of analysts' reports by treatment condition

	Flesch-Kincaid			Coleman-Liau			Coherence			Word count		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treat × Post	-0.046	-0.160	-0.577	-0.002	-0.094	-0.351	-0.004	-0.007	-0.032**	-24.079	-31.942*	-6.354
	(0.158)	(0.258)	(0.467)	(0.109)	(0.162)	(0.261)	(0.007)	(0.008)	(0.014)	(16.867)	(18.223)	(25.261)
Explainability × Post	0.177	0.206	-0.410	-0.030	-0.008	-0.537*	-0.007	-0.002	-0.006	-36.869*	-48.790**	-56.692**
	(0.237)	(0.361)	(0.456)	(0.187)	(0.288)	(0.278)	(0.012)	(0.007)	(0.012)	(20.315)	(20.982)	(27.669)
Treat × Post × Active Equity		0.322	0.754		0.242	0.577*		0.013	0.054***		-17.205	-43.490
		(0.298)	(0.554)		(0.230)	(0.323)		(0.011)	(0.019)		(30.765)	(29.539)
Explainability × Post × Active Equity		-0.310	0.934*		-0.164	0.950***		-0.011	0.027		14.012	44.041
		(0.485)	(0.532)		(0.396)	(0.331)		(0.033)	(0.025)		(38.270)	(30.191)
Treat × Post × Experienced			0.712			0.383			0.042***			-15.298
			(0.567)			(0.338)			(0.014)			(35.917)
Explainability × Post × Experienced			1.101**			0.915**			0.006			42.263
			(0.526)			(0.353)			(0.015)			(33.222)
Treat × Post × Active Equity × Experienced			-0.774			-0.606			-0.072***			33.172
			(0.681)			(0.442)			(0.021)			(51.800)
Explainability × Post × Active Equity × Experienced			-2.030***			-1.667***			-0.063			-64.348
			(0.690)			(0.506)			(0.047)			(52.801)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	984	984	984	984	984	984	986	986	986	984	984	984
Mean (control)	11.193	11.193	11.193	13.632	13.632	13.632	0.233	0.233	0.233	1533.126	1533.126	1533.126
SD (control)	1.085	1.085	1.085	0.777	0.777	0.777	0.037	0.037	0.037	86.421	86.421	86.421

Notes: All regressions include a full set of strata and month fixed effects. For all measures except Word count, we first processed the text by lemmatizing and removing stop words before constructing the variables. The *Flesch-Kincaid score* in columns (1)-(3) captures the readability of reports based on the number of words in sentences and the average number of syllables per word. The *Coleman-Liau index* in columns (4)-(6) provides another measure of readability based on sentence length and the average number of characters per word. *Coherence* in columns (7)-(9) measures the coherence of topics in the report based on topic modeling. Word count in columns (10)-(12) measures the number of words in the analyst report. *Treat* is a dummy variable that takes the value 1 for analysts assigned to treatment 1 and 0 otherwise. *Explainability* is a dummy variable that takes the value 1 for analysts assigned to treatment 2 and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise.

Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.8: The impact of algorithmic predictions on standardized stated confidence

	Confidence before interview			Confidence after interview		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.247**	0.375*	0.256	0.147	0.296	0.101
	(0.124)	(0.205)	(0.165)	(0.136)	(0.222)	(0.189)
Treat × Post × Active Equity		-0.315	-0.578**		-0.394	-0.451
		(0.257)	(0.265)		(0.278)	(0.371)
Treat × Post × Experienced			0.256			0.360
			(0.372)			(0.400)
Treat × Post × Active Equity × Experienced			0.229			-0.033
			(0.466)			(0.543)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1887	1887	1887
Mean (control)	-0.071	-0.071	-0.071	-0.041	-0.041	-0.041
SD (control)	0.885	0.885	0.885	0.920	0.920	0.920

	Confidence before interview			Confidence after interview		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.121	0.193	0.231	0.060	0.143	0.102
	(0.111)	(0.150)	(0.167)	(0.113)	(0.150)	(0.176)
Explainability × Post	0.394*	0.549	0.161	0.263	0.463	0.065
	(0.212)	(0.370)	(0.174)	(0.235)	(0.401)	(0.279)
Treat × Post × Active Equity		-0.161	-0.401		-0.219	-0.261
		(0.222)	(0.280)		(0.240)	(0.369)
Explainability × Post × Active Equity		-0.428	-0.677**		-0.577	-0.666
		(0.433)	(0.284)		(0.457)	(0.427)
Treat × Post × Experienced			-0.142			-0.026
			(0.259)			(0.264)
Explainability × Post × Experienced			0.728			0.743
			(0.613)			(0.697)
Treat × Post × Active Equity × Experienced			0.310			0.035
			(0.404)			(0.462)
Explainability × Post × Active Equity × Experienced			0.037			-0.120
			(0.709)			(0.809)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1887	1887	1887
Mean (control)	-0.071	-0.071	-0.071	-0.041	-0.041	-0.041
SD (control)	0.885	0.885	0.885	0.920	0.920	0.920

Notes: The *confidence before interview* is the standardized stated confidence in their decisions reported by analysts on a scale of 1-10 before they conduct interview and additional research. The *confidence after interview* the standardized stated confidence in their decisions reported by analysts on a scale of 1-10 before they conduct interview and additional research. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Treat* is a dummy variable that takes the value 1 for analysts assigned to treatment 1 and 0 otherwise. *Explainability* is a dummy variable that takes the value 1 for analysts assigned to treatment 2 and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.9: The impact of machine predictions on decisions and reasoning

	Change rating			Reasoning score		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.080**	0.062	0.093**	-0.103	-0.059	-0.054
	(0.031)	(0.045)	(0.046)	(0.118)	(0.154)	(0.153)
Treat × Post × High Baseline Confidence		0.056			-0.003	
		(0.066)			(0.213)	
Treat × Post × High Baseline Reasoning Score			-0.030			-0.090
			(0.066)			(0.235)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	Yes	Yes
Observations	1780	1780	1780	1207	1207	1207
Mean (control)	0.113	0.113	0.113	0.238	0.238	0.238
SD (control)	0.317	0.317	0.317	0.928	0.928	0.928

Notes: All regressions include a full set of strata and month fixed effects.
All regressions include a full set of strata and month fixed effects. Regressions in columns (4)-(6) further include review committee fixed effects. Columns (1)-(3) show treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Columns (4)-(6) show treatment effects on *reasoning score*, constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *High baseline confidence* takes value 1 if the analyst has an average stated confidence after interview that is above average at baseline and 0 otherwise. *High baseline reasoning score* takes value 1 if the analyst has an average reasoning score that is above average at baseline and 0 otherwise.

Table A.10: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating		Reasoning score	
	(1)	(2)	(3)	(4)
Treat × Post	0.080**	0.094***	-0.103	-0.161
	(0.031)	(0.033)	(0.118)	(0.118)
Treat × Post × New coverage		-0.133***		0.531
		(0.048)		(0.418)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Committee FE	No	No	Yes	Yes
Observations	1780	1780	1207	1207
Mean (control)	0.113	0.113	0.238	0.238
SD (control)	0.317	0.317	0.928	0.928

Panel B. Decision performance

	Fund returns in 3 months		Fund returns in 6 months	
	(1)	(2)	(3)	(4)
Treat × Post × Recommend	0.894	-0.070	1.130	0.555
	(3.033)	(2.954)	(2.165)	(2.284)
Treat × Post × Recommend × New Coverage		15.054		12.369
		(11.808)		(8.223)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	1670	1670	1668	1668
Mean (control)	-4.467	-4.467	-4.832	-4.832
SD (control)	13.865	13.865	11.525	11.525

Notes: Panel A shows treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on *Fund returns in 3 months*, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on *Fund returns in 6 months*, the risk-adjusted return of a fund in 6 months after its rating is published. *Treatment* is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. *New coverage* takes value 1 if it is the first time that an analyst covers a given fund and 0 otherwise. All regressions include a full set of strata and month fixed effects, and regressions in Panel B additionally control for fund age and size.
Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.11: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating		Reasoning score	
	(1)	(2)	(3)	(4)
Treat × Post	0.080**	0.166***	-0.103	0.197
	(0.031)	(0.049)	(0.118)	(0.244)
Treat × Post × Male		-0.115*		-0.401
		(0.065)		(0.286)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Committee FE	No	No	Yes	Yes
Observations	1780	1780	1207	1207
Mean (control)	0.113	0.113	0.238	0.238
SD (control)	0.317	0.317	0.928	0.928

Panel B. Decision performance

	Fund returns in 3 months		Fund returns in 6 months	
	(1)	(2)	(3)	(4)
Treat × Post × Recommend	0.894	1.203	1.130	5.424
	(3.033)	(6.518)	(2.165)	(3.463)
Treat × Post × Recommend × Male		-0.025		-5.239
		(7.387)		(4.313)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	1670	1670	1668	1668
Mean (control)	-4.467	-4.467	-4.832	-4.832
SD (control)	13.865	13.865	11.525	11.525

Notes: Panel A shows treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on *Fund returns in 3 months*, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on *Fund returns in 6 months*, the risk-adjusted return of a fund in 6 months after its rating is published. *Treatment* is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. *Male* takes value 1 if the analyst is male and 0 if the analyst is female. All regressions include a full set of strata and month fixed effects, and regressions in Panel B additionally control for fund age and size.
Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.12: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating		Reasoning score	
	(1)	(2)	(3)	(4)
Treat × Post	0.080**	0.080**	-0.103	-0.071
	(0.031)	(0.032)	(0.118)	(0.122)
Treat × Post × Manager		-0.024		-0.525
		(0.101)		(0.433)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Committee FE	No	No	Yes	Yes
Observations	1780	1780	1207	1207
Mean (control)	0.113	0.113	0.238	0.238
SD (control)	0.317	0.317	0.928	0.928

Panel B. Decision performance

	Fund returns in 3 months		Fund returns in 6 months	
	(1)	(2)	(3)	(4)
Treat × Post × Recommend	0.894	1.748	1.130	1.338
	(3.033)	(3.219)	(2.165)	(2.285)
Treat × Post × Recommend × Manager		-15.320**		-4.750
		(6.379)		(4.432)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	1670	1670	1668	1668
Mean (control)	-4.467	-4.467	-4.832	-4.832
SD (control)	13.865	13.865	11.525	11.525

Notes: Panel A shows treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on *Fund returns in 3 months*, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on *Fund returns in 6 months*, the risk-adjusted return of a fund in 6 months after its rating is published. *Treatment* is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. *Male* takes value 1 if the analyst is male and 0 otherwise. All regressions include a full set of strata and month fixed effects, and regressions in Panel B additionally control for fund age and size.
Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.13: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating		Reasoning score	
	(1)	(2)	(3)	(4)
Treat × Post	0.080**	0.088**	-0.103	-0.120
	(0.031)	(0.044)	(0.118)	(0.173)
Treat × Post × Postgraduate		-0.010		0.025
		(0.065)		(0.227)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Committee FE	No	No	Yes	Yes
Observations	1780	1780	1207	1207
Mean (control)	0.113	0.113	0.238	0.238
SD (control)	0.317	0.317	0.928	0.928

Panel B. Decision performance

	Fund returns in 3 months		Fund returns in 6 months	
	(1)	(2)	(3)	(4)
Treat × Post × Recommend	0.894	-5.856	1.130	-3.132
	(3.033)	(3.598)	(2.165)	(2.372)
Treat × Post × Recommend × Postgraduate		15.359**		8.961*
		(6.443)		(4.728)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	1670	1670	1668	1668
Mean (control)	-4.467	-4.467	-4.832	-4.832
SD (control)	13.865	13.865	11.525	11.525

Notes: Panel A shows treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on *Fund returns in 3 months*, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on *Fund returns in 6 months*, the risk-adjusted return of a fund in 6 months after its rating is published. *Treatment* is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. *Postgraduate* takes value 1 if the analyst has a post-graduate degree and 0 otherwise. All regressions include a full set of strata and month fixed effects, and regressions in Panel B additionally control for fund age and size.
Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.14: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating					
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.074**	0.075**	0.080**	0.080**	0.080*	0.067
	(0.033)	(0.033)	(0.031)	(0.031)	(0.043)	(0.060)
Treat × Post × Active Equity					-0.000	0.007
					(0.071)	(0.106)
Treat × Post × Experienced						0.033
						(0.090)
Treat × Post × Active Equity × Experienced						-0.029
						(0.144)
Strata FE	No	Yes	No	Yes	Yes	Yes
Month FE	No	No	Yes	Yes	Yes	Yes
Observations	1780	1780	1780	1780	1780	1780
Mean (control)	0.113	0.113	0.113	0.113	0.113	0.113
SD (control)	0.317	0.317	0.317	0.317	0.317	0.317

Panel B. Decision performance

	Fund returns in 3 months			Fund returns in 6 months		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post × Recommend	0.894	9.886***	13.499***	1.130	7.949***	8.885**
	(3.033)	(2.585)	(3.440)	(2.165)	(2.487)	(3.620)
Treat × Post × Recommend × Active Equity		-16.822***	-14.568*		-11.458***	-8.163
		(5.268)	(7.764)		(4.191)	(6.830)
Treat × Post × Recommend × Experienced			-9.966*			-3.784
			(5.374)			(4.691)
Treat × Post × Recommend × Active Equity × Experienced			-0.198			-1.867
			(11.332)			(9.125)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1668	1668	1668
Mean (control)	-4.467	-4.467	-4.467	-4.832	-4.832	-4.832
SD (control)	13.865	13.865	13.865	11.525	11.525	11.525

Notes: Panel A shows treatment effects on Change rating, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on Fund returns in 3 months, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on Fund returns in 6 months, the risk-adjusted return of a fund in 6 months after its rating is published. Treatment is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. Post takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than the medium rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. All regressions include a full set of strata and month fixed effects, and regressions in Panel B additionally control for fund age and size.
Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.15: The impact of machine predictions on reasoning

	All committees				Single
	(1)	(2)	(3)	(4)	(5)
Treat × Post	0.012	-0.124	-0.292	-0.366	-0.423
	(0.169)	(0.210)	(0.307)	(0.249)	(0.277)
Treat × Post × Active Equity		0.297	0.440	0.301	0.707*
		(0.351)	(0.364)	(0.299)	(0.376)
Treat × Post × Experienced			0.410	0.520	0.533
			(0.445)	(0.333)	(0.373)
Treat × Post × Active Equity × Experienced			-0.478	-0.537	-0.821
			(0.645)	(0.438)	(0.526)
Strata FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	Yes
Observations	1313	1313	1313	1207	901
Mean (control)	0.238	0.238	0.238	0.238	0.286
SD (control)	0.928	0.928	0.928	0.928	0.968

Notes: All regressions include a full set of strata and month fixed effects. Regressions in columns (4)-(5) further include review committee fixed effects. *Reasoning score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. Regressions in columns (1)-(4) include committees of all sizes, which vary between 1 reviewer and 4 reviewers per committee. Column (5) analyzes the subsample of committees with one reviewer only. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than the medium rating experience at the company and 0 otherwise.
Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.16: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating				
	(1)	(2)	(3)	(4)	(5)
Treat × Post	0.074**	0.075**	0.080**	0.080**	
	(0.033)	(0.033)	(0.031)	(0.031)	
Treat x Post x New x Non-Active Equity					0.076
					(0.046)
Treat x Post x New x Active Equity					0.045
					(0.097)
Treat x Post x Experienced x Non-Active Equity					0.079
					(0.054)
Treat x Post x Experienced x Active Equity					0.100*
					(0.057)
Strata FE	No	Yes	No	Yes	Yes
Month FE	No	No	Yes	Yes	Yes
Observations	1780	1780	1780	1780	1780
Mean (control)	0.113	0.113	0.113	0.113	0.113
SD (control)	0.317	0.317	0.317	0.317	0.317

Panel B. Decision performance

	Fund returns in 3 months		Fund returns in 6 months	
	(1)	(2)	(3)	(4)
Treat × Post × Recommend	0.894		1.130	
	(3.033)		(2.165)	
Treat x Post x Recommend x New x Non-Active Equity		9.933***		6.359
		(3.439)		(4.085)
Treat x Post x Recommend x New x Active Equity		-5.204		0.718
		(6.912)		(4.783)
Treat x Post x Recommend x Experienced x Non-Active Equity		6.418*		6.077**
		(3.536)		(2.879)
Treat x Post x Recommend x Experienced x Active Equity		-6.328		-4.788
		(5.267)		(3.973)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	1670	1670	1668	1668
Mean (control)	-4.467	-4.467	-4.832	-4.832
SD (control)	13.865	13.865	11.525	11.525

Notes: Panel A shows treatment effects on Change rating, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on Fund returns in 3 months, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on Fund returns in 6 months, the risk-adjusted return of a fund in 6 months after its rating is published. Treatment is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. Post takes the value 1 after the treatment is implemented and 0 otherwise. Active Equity takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. Experienced takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. Recommend takes the value 1 for funds that analysts recommend and 0 otherwise. All regressions include a full set of strata and month fixed effects, and regressions in Panel B additionally control for fund age and size.
Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.17: The impact of machine predictions on reasoning

	All committees			Single
	(1)	(2)	(3)	(4)
Treat × Post	0.012			
	(0.169)			
Treat x Post x New x Non-Active Equity		-0.509*	-0.368*	-0.231
		(0.261)	(0.221)	(0.213)
Treat x Post x New x Active Equity		0.660**	0.366*	0.545**
		(0.259)	(0.215)	(0.247)
Treat x Post x Experienced x Non-Active Equity		-0.047	-0.018	-0.040
		(0.249)	(0.194)	(0.201)
Treat x Post x Experienced x Active Equity		-0.024	-0.295	-0.152
		(0.352)	(0.187)	(0.224)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Committee FE	No	No	Yes	Yes
Observations	1313	1313	1207	901
Mean (control)	0.238	0.238	0.238	0.286
SD (control)	0.928	0.928	0.928	0.968

Notes: All regressions include a full set of strata and month fixed effects. Regressions in columns (4)-(5) further include review committee fixed effects. *Reasoning score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. Regressions in columns (1)-(4) include committees of all sizes, which vary between 1 reviewer and 4 reviewers per committee. Column (5) analyzes the subsample of committees with one reviewer only. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise.
Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.18: The number of decision drivers and causal statements by experimental condition

	Decision drivers		Causal statements	
	(1)	(2)	(3)	(4)
Treat × Post	-0.332*		-0.065	
	(0.187)		(0.040)	
Treat x Post x New x Non-Active Equity		-0.047		-0.016
		(0.348)		(0.050)
Treat x Post x New x Active Equity		-0.547		-0.184***
		(0.409)		(0.052)
Treat x Post x Experienced x Non-Active Equity		-0.492		-0.157**
		(0.404)		(0.068)
Treat x Post x Experienced x Active Equity		-0.487		0.079
		(0.295)		(0.063)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	1317	1317	948	948
Mean (control)	1.728	1.728	2.104	2.104
SD (control)	0.896	0.896	0.274	0.274

Notes: All regressions include a full set of strata and month fixed effects. *Decision drivers* measure the number of reported decision drivers by the analyst and is natural log transformed. *Causal statements* measure the number of causal statements in the analyst report as identified by GPT-4 and is natural log transformed. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. Post takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise.
Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.19: The impact of algorithm explainability on decisions and causal explanations

Panel A: Decision change				
	Change rating		Reasoning score	
	(1)	(2)	(3)	(4)
Treat × Post	0.120***		0.005	
	(0.041)		(0.235)	
Explainability × Post	0.040		0.020	
	(0.034)		(0.194)	
Treat x Post x New x Non-Active Equity		0.038		-0.602**
		(0.056)		(0.253)
Treat x Post x New x Active Equity		0.059		0.603**
		(0.130)		(0.291)
Treat x Post x Experienced x Non-Active Equity		0.146**		-0.483**
		(0.063)		(0.230)
Treat x Post x Experienced x Active Equity		0.224***		0.295
		(0.074)		(0.616)
Explainability x Post x New x Non-Active Equity		0.107*		-0.411
		(0.055)		(0.357)
Explainability x Post x New x Active Equity		0.098*		0.832***
		(0.057)		(0.270)
Explainability x Post x Experienced x Non-Active Equity		0.007		0.290
		(0.060)		(0.361)
Explainability x Post x Experienced x Active Equity		0.018		-0.283
		(0.045)		(0.280)
Strata FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	1780	1780	1313	1313
Panel B: Decision performance				
	Fund returns in 3 months		Fund returns in 6 months	
	(1)	(2)	(3)	(4)
Treat × Post	-1.278		-1.941	
	(2.072)		(2.296)	
Explainability × Post	-3.100		-0.092	
	(2.120)		(2.694)	
Treat × Post × Recommend	0.002	-9.388*	1.834	-1.865
	(4.065)	(5.120)	(2.547)	(3.704)
Explainability × Post × Recommend	2.235	-3.169	0.322	-6.295
	(3.357)	(6.759)	(3.118)	(5.915)
Treat x Post x New x Non-Active Equity		-1.605		1.247
		(2.614)		(3.696)
Treat x Post x New x Active Equity		4.428		-2.080
		(5.208)		(7.220)
Treat x Post x Experienced x Non-Active Equity		-2.542		-2.996
		(1.958)		(1.879)
Treat x Post x Experienced x Active Equity		-3.432		-3.822
		(2.679)		(3.001)
Explainability x Post x New x Non-Active Equity		-5.419		5.917
		(3.958)		(5.600)
Explainability x Post x New x Active Equity		-2.238		-9.630**
		(7.288)		(4.086)
Explainability x Post x Experienced x Non-Active Equity		-5.452*		-4.355**
		(2.870)		(2.100)
Explainability x Post x Experienced x Active Equity		0.285		3.651
		(4.472)		(5.070)
Treat x Post x Recommend x New x Non-Active Equity		23.282***		11.422*
		(6.431)		(5.931)

Table A.20: The depth and breadth of decision drivers and causal statements by experimental condition

	Breadth of decision drivers			Depth of decision drivers		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	-0.386	-0.027	0.134	-0.331	0.257	1.239**
	(0.251)	(0.336)	(0.433)	(0.448)	(0.556)	(0.600)
Treat × Post × Active Equity		-0.613	-0.936		-0.922	-2.524*
		(0.444)	(0.616)		(0.867)	(1.308)
Treat × Post × Experienced			-0.135			-1.659
			(0.574)			(1.065)
Treat × Post × Active Equity × Experienced			0.263			2.540
			(0.779)			(1.709)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1317	1317	1317	1317	1317	1317
Mean (control)	2.596	2.596	2.596	4.003	4.003	4.003
SD (control)	1.216	1.216	1.216	2.404	2.404	2.404

Notes: All regressions include a full set of strata and month fixed effects. *Breadth of decision drivers* measures to what extend the reported decision drivers by the analyst span across five major categories of decision drivers. *Depth of decision drivers* measures depth of decision drivers by counting the number of decision drivers in the category where an analyst reports the largest number of decision drivers. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. Post takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise.
Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.21: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating		
	(1)	(2)	(3)
Treat × Post	0.075**	0.070	0.109*
	(0.033)	(0.045)	(0.057)
Treat × Post × Active Equity		0.006	-0.047
		(0.072)	(0.119)
Treat × Post × Experienced			-0.045
			(0.086)
Treat × Post × Active Equity × Experienced			0.060
			(0.149)
Strata FE	Yes	Yes	Yes
Observations	1780	1780	1780
Mean (control)	0.113	0.113	0.113
SD (control)	0.317	0.317	0.317

Panel B. Decision performance

	Fund returns in 3 months			Fund returns in 6 months		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post × Recommend	1.247	12.367***	14.468***	1.123	8.543***	6.704*
	(3.178)	(2.654)	(3.365)	(2.250)	(2.563)	(3.679)
Treat × Post × Recommend × Active Equity		-21.020***	-19.903***		-12.566***	-8.136
		(5.161)	(7.204)		(4.313)	(6.443)
Treat × Post × Recommend × Experienced			-7.326			0.148
			(4.892)			(4.396)
Treat × Post × Recommend × Active Equity × Experienc			4.800			-1.575
			(10.114)			(8.415)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1668	1668	1668
Mean (control)	-4.467	-4.467	-4.467	-4.832	-4.832	-4.832
SD (control)	13.865	13.865	13.865	11.525	11.525	11.525

Notes: Panel A shows treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on *Fund returns in 3 months*, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on *Fund returns in 6 months*, the risk-adjusted return of a fund in 6 months after its rating is published. *Treatment* is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. All regressions include strata fixed effects, and regressions in Panel B additionally control for fund age and size. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.22: The impact of machine predictions on reasoning

	All committees				Single
	(1)	(2)	(3)	(4)	(5)
Treat × Post	0.007	-0.131	-0.663*	-0.542*	-0.661*
	(0.166)	(0.211)	(0.343)	(0.317)	(0.333)
Treat × Post × Active Equity		0.301	0.910**	0.599	1.124**
		(0.349)	(0.398)	(0.384)	(0.445)
Treat × Post × Experienced			0.814*	0.652*	0.758*
			(0.435)	(0.370)	(0.393)
Treat × Post × Active Equity × Experienced			-0.939	-0.799*	-1.300**
			(0.603)	(0.462)	(0.530)
Strata FE	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	Yes
Observations	1313	1313	1313	1207	901
Mean (control)	0.238	0.238	0.238	0.238	0.286
SD (control)	0.928	0.928	0.928	0.928	0.968

Notes: All regressions include strata fixed effects. Regressions in columns (4)-(5) further include review committee fixed effects. *Reasoning score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. Regressions in columns (1)-(4) include committees of all sizes, which vary between 1 reviewer and 4 reviewers per committee. Column (5) analyzes the subsample of committees with one reviewer only. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.23: The number of decision drivers and causal statements by experimental condition

	Decision drivers			Causal statements		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	-0.319*	0.040	0.443	-0.061	-0.103*	0.041
	(0.191)	(0.260)	(0.320)	(0.041)	(0.053)	(0.083)
Treat × Post × Active Equity		-0.609*	-1.374**		0.097	-0.200*
		(0.352)	(0.543)		(0.080)	(0.115)
Treat × Post × Experienced			-0.776*			-0.243**
			(0.459)			(0.107)
Treat × Post × Active Equity × Experienced			1.266*			0.481***
			(0.666)			(0.147)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1317	1317	1317	948	948	948
Mean (control)	1.728	1.728	1.728	2.104	2.104	2.104
SD (control)	0.896	0.896	0.896	0.274	0.274	0.274

Notes: All regressions include strata fixed effects. *Decision drivers* measure the number of reported decision drivers by the analyst and is natural log transformed. *Causal statements* measure the number of causal statements in the analyst report as identified by GPT-4 and is natural log transformed. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. Post takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.24: The impact of algorithm explainability on decisions and reasoning

Panel A: Decision change						
	Change rating			Reasoning score		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.111**	0.082	0.070	0.004	-0.359	-0.812**
	(0.043)	(0.051)	(0.067)	(0.233)	(0.225)	(0.361)
Explainability × Post	0.039	0.056	0.150***	0.010	0.094	-0.540
	(0.037)	(0.051)	(0.057)	(0.189)	(0.277)	(0.395)
Treat × Post × Active Equity		0.067	-0.007		0.727	1.009**
		(0.102)	(0.150)		(0.449)	(0.440)
Explainability × Post × Active Equity		-0.045	-0.006		-0.162	0.954**
		(0.068)	(0.084)		(0.385)	(0.391)
Treat × Post × Experienced			0.060			0.553
			(0.097)			(0.425)
Explainability × Post × Experienced			-0.153			1.023*
			(0.093)			(0.552)
Treat × Post × Active Equity × Experienced			0.087			-0.323
			(0.185)			(0.789)
Explainability × Post × Active Equity × Experienced			-0.000			-1.565**
			(0.122)			(0.632)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1780	1780	1780	1313	1313	1313
Panel B: Decision performance						
	Fund returns in 3 months			Fund returns in 6 months		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post × Recommend	0.418	13.524***	18.246***	2.187	10.252***	10.589**
	(4.193)	(3.436)	(3.948)	(2.655)	(3.175)	(4.391)
Explainability × Post × Recommend	2.524	12.003***	14.319***	-0.123	6.919**	2.976
	(3.684)	(3.737)	(5.247)	(3.118)	(3.279)	(4.755)
Treat × Post × Recommend × Active Equity		-24.363***	-22.685***		-13.090**	-12.380
		(5.966)	(8.499)		(5.051)	(7.527)
Explainability × Post × Recommend × Active Equity		-18.022**	-23.613***		-11.938**	-3.070
		(6.931)	(7.945)		(6.002)	(7.170)
Treat × Post × Recommend × Experienced			-10.135*			-2.229
			(5.604)			(5.178)
Explainability × Post × Recommend × Experienced			-11.585			-1.177
			(8.956)			(6.589)
Treat × Post × Recommend × Active Equity × Experienc			1.595			3.223
			(10.856)			(9.114)
Explainability × Post × Recommend × Active Equity ×			17.415			-2.408
			(13.136)			(10.804)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1668	1668	1668

Notes: *Treat* is an indicator variable that takes value 1 for analysts assigned to treatment 1 who receive only the algorithmic prediction. *Explainability* is a dummy variable that takes the value 1 for analysts assigned to treatment 2 who receive algorithmic predictions along with key fund features that contributed most to the prediction based on Shapley values. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. All regressions include strata fixed effects, and regressions in Panel B additionally control for fund age and size. Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

Table A.25: The impact of machine predictions on decisions

Panel A. Decision change

	Change rating		
	(1)	(2)	(3)
Treat × Post	0.080**	0.074*	0.105*
	(0.031)	(0.044)	(0.059)
Treat × Post × Active Equity		0.003	-0.034
		(0.071)	(0.126)
Treat × Post × Experienced			-0.035
			(0.086)
Treat × Post × Active Equity × Experienced			0.039
			(0.154)
Month FE	Yes	Yes	Yes
Observations	1780	1780	1780
Mean (control)	0.113	0.113	0.113
SD (control)	0.317	0.317	0.317

Panel B. Decision performance

	Fund returns in 3 months			Fund returns in 6 months		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post × Recommend	1.183	10.044***	10.746***	1.094	7.943***	5.693
	(3.101)	(2.777)	(3.820)	(2.239)	(2.441)	(3.968)
Treat × Post × Recommend × Active Equity		-17.437***	-16.506**		-12.119***	-7.769
		(5.344)	(7.706)		(4.083)	(6.303)
Treat × Post × Recommend × Experienced			-3.264			2.721
			(5.430)			(4.939)
Treat × Post × Recommend × Active Equity × Experienc			3.338			-2.799
			(10.535)			(8.445)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1668	1668	1668
Mean (control)	-4.467	-4.467	-4.467	-4.832	-4.832	-4.832
SD (control)	13.865	13.865	13.865	11.525	11.525	11.525

Notes: Panel A shows treatment effects on *Change rating*, which is an indicator variable that takes the value 1 if the rating decision by the analyst differs from the currently assigned rating. Panel B shows whether recommended funds by treated analysts observe higher fund returns in subsequent months. Columns (1)-(3) of Panel B show treatment effects on *Fund returns in 3 months*, the risk-adjusted return of a fund 3 months after its rating is published. Columns (4)-(6) of Panel B show treatment effects on *Fund returns in 6 months*, the risk-adjusted return of a fund in 6 months after its rating is published. *Treatment* is an indicator variable that takes the value 1 for analysts assigned to the treatment group and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. All regressions include month fixed effects, and regressions in Panel B additionally control for fund age and size. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.26: The impact of machine predictions on reasoning

	All committees				Single
	(1)	(2)	(3)	(4)	(5)
Treat × Post	-0.002	-0.055	-0.613*	-0.510*	-0.603*
	(0.177)	(0.218)	(0.316)	(0.292)	(0.318)
Treat × Post × Active Equity		0.188	0.785**	0.528	1.005**
		(0.354)	(0.358)	(0.350)	(0.407)
Treat × Post × Experienced			0.818*	0.658*	0.726*
			(0.418)	(0.352)	(0.381)
Treat × Post × Active Equity × Experienced			-0.885	-0.778*	-1.220**
			(0.590)	(0.437)	(0.498)
Month FE	Yes	Yes	Yes	Yes	Yes
Committee FE	No	No	No	Yes	Yes
Observations	1313	1313	1313	1207	901
Mean (control)	0.238	0.238	0.238	0.238	0.286
SD (control)	0.928	0.928	0.928	0.928	0.968

Notes: All regressions include month fixed effects. Regressions in columns (4)-(5) further include review committee fixed effects. *Reasoning score* is constructed at the committee level by standardizing individual scores and taking an average across all reviewers in the committee. Regressions in columns (1)-(4) include committees of all sizes, which vary between 1 reviewer and 4 reviewers per committee. Column (5) analyzes the subsample of committees with one reviewer only. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.27: The number of decision drivers and causal statements by experimental condition

	Decision drivers			Causal statements		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	-0.319	0.082	0.476	-0.067	-0.123**	0.033
	(0.199)	(0.297)	(0.330)	(0.043)	(0.055)	(0.075)
Treat × Post × Active Equity		-0.718*	-1.419**		0.142*	-0.201*
		(0.386)	(0.552)		(0.084)	(0.107)
Treat × Post × Experienced			-0.873			-0.252**
			(0.547)			(0.099)
Treat × Post × Active Equity × Experienced			1.330*			0.539***
			(0.730)			(0.140)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1317	1317	1317	948	948	948
Mean (control)	1.728	1.728	1.728	2.104	2.104	2.104
SD (control)	0.896	0.896	0.896	0.274	0.274	0.274

Notes: All regressions include month fixed effects. *Decision drivers* measure the number of reported decision drivers by the analyst and is natural log transformed. *Causal statements* measure the number of causal statements in the analyst report as identified by GPT-4 and is natural log transformed. *Treatment* is a dummy variable that takes the value 1 for analysts assigned to treatment and 0 otherwise. Post takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. Standard errors are in parentheses, clustered at analyst level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A.28: The impact of algorithm explainability on decisions and reasoning

Panel A: Decision change						
	Change rating			Reasoning score		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post	0.119***	0.092*	0.072	-0.085	-0.383	-0.854**
	(0.040)	(0.049)	(0.067)	(0.245)	(0.244)	(0.355)
Explainability × Post	0.040	0.055	0.138**	0.084	0.259	-0.394
	(0.034)	(0.051)	(0.065)	(0.197)	(0.277)	(0.352)
Treat × Post × Active Equity		0.058	0.012		0.681	0.958**
		(0.099)	(0.158)		(0.460)	(0.414)
Explainability × Post × Active Equity		-0.046	-0.024		-0.323	0.810**
		(0.069)	(0.096)		(0.371)	(0.336)
Treat × Post × Experienced			0.070			0.602
			(0.094)			(0.448)
Explainability × Post × Experienced			-0.143			0.979*
			(0.097)			(0.501)
Treat × Post × Active Equity × Experienced			0.041			-0.289
			(0.192)			(0.806)
Explainability × Post × Active Equity × Experienced			0.023			-1.544***
			(0.129)			(0.582)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1780	1780	1780	1313	1313	1313
Panel B: Decision performance						
	Fund returns in 3 months			Fund returns in 6 months		
	(1)	(2)	(3)	(4)	(5)	(6)
Treat × Post × Recommend	0.098	10.143***	14.597***	1.436	7.988**	8.540*
	(4.094)	(3.746)	(4.758)	(2.644)	(3.221)	(4.729)
Explainability × Post × Recommend	2.739	10.658***	10.634*	0.711	7.911**	2.127
	(3.361)	(3.659)	(5.384)	(2.961)	(3.008)	(5.122)
Treat × Post × Recommend × Active Equity		-19.162***	-19.567**		-11.025**	-11.302
		(6.465)	(9.712)		(4.842)	(7.280)
Explainability × Post × Recommend × Active Equity		-15.794**	-19.994***		-12.782**	-1.779
		(6.773)	(6.870)		(5.735)	(7.809)
Treat × Post × Recommend × Experienced			-9.257			-1.349
			(6.376)			(5.863)
Explainability × Post × Recommend × Experienced			-0.672			7.629
			(8.116)			(6.457)
Treat × Post × Recommend × Active Equity × Experienc			5.594			4.785
			(11.913)			(9.261)
Explainability × Post × Recommend × Active Equity ×			7.538			-11.419
			(11.782)			(10.914)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1670	1670	1670	1668	1668	1668

Notes: *Treat* is an indicator variable that takes value 1 for analysts assigned to treatment 1 who receive only the algorithmic prediction. *Explainability* is a dummy variable that takes the value 1 for analysts assigned to treatment 2 who receive algorithmic predictions along with key fund features that contributed most to the prediction based on Shapley values. *Post* takes the value 1 after the treatment is implemented and 0 otherwise. *Active Equity* takes value 1 if the fund is classified as an actively managed equity fund and 0 otherwise. *Experienced* takes the value 1 if the analyst has more than three years of rating experience at the company and 0 otherwise. *Recommend* takes the value 1 for funds that analysts recommend and 0 otherwise. All regressions include month fixed effects, and regressions in Panel B additionally control for fund age and size. Standard errors are in parentheses, clustered at analyst level. * p<0.05, ** p<0.01, *** p<0.001.

A.1 Examples of analyst reports

A.1.1 The summary section of an analyst report that received a high reasoning score

[redacted]'s seasoned management team deftly executes a research-driven, flexible process that delivers income while limiting downside risk. It earns a [redacted] for its cheapest share classes and [redacted] on its more expensive shares. [redacted] brought an income-focused approach to this strategy in late 2011, and after a decade it's clear he has delivered on this strategy's objective. Over his tenure through October 2022, the strategy's 12-month yield averaged 4.8%, impressive compared with the category peer average yield of 2.2%. Though [redacted] takes the lead here, he is well supported. [redacted] joined the manager roster with [redacted], and [redacted] was added in March 2015. The trio's collaboration and ability to tap the firm's extensive resources and strong fundamental teams gives them an edge relative to other multi-asset income peers. Around 17 distinct, well-regarded [redacted] teams run sleeves for the strategy, including the high-yield team that manages [redacted]. Alongside an experienced team is an income approach that differentiates itself through its emphasis on downside protection. Rather than loading up on a mix of risky asset classes that tend to be correlated during stress periods, the team focuses on the correlation of its underlying sleeves to build a well-diversified portfolio. The managers seek to keep the fund's volatility below that of a blended benchmark (50% [redacted] World Index and 50% [redacted] Bond Index) but are not constrained to a strategic allocation. This gives the team a great amount of flexibility, but it has deliberately abstained from making quick and drastic allocation changes that could lead to a whirlwind experience for investors. Instead, the team, with input from the underlying sleeve managers, focuses on its medium- and long-term views to gradually tilt and adjust the portfolio's composition. Management has been able to limit the risk of drawdowns over [redacted]'s tenure. The strategy's downside-capture ratio of 93% relative to the category index compares favorably with the allocation—30% to 50% equity category average of 108%. Investors should be wary, however, of this strategy's credit risk, which may cause bouts of volatility. Around 66% of its fixed-income exposure resides in securities rated BB or below versus around 25% for category peers.

A.1.2 The summary section of an analyst report that received a low reasoning score

[redacted] is a systematic strategy that tilts toward smaller value names, resulting in a portfolio that can exhibit factor sensitivity. With low fees and a strong parent in [redacted], the sole Investor share class retains its [redacted].

The investment team employs a sensible systematic process that seeks to identify stocks that will grow earnings faster than industry peers and that are trading at reasonable valuations. The starting investment universe is its prospectus benchmark, the [redacted] Index. The model looks for stocks that score well on five factors–valuation, growth, quality, momentum, and management decisions–

relative to industry peers. The fund does not make sector bets versus the index but can have certain size and style tilts. The model generally finds more attractive opportunities in smaller value names, relative to the mid-blend category benchmark, the [redacted] Index.

This strategy launched in 1995 as a quantitative fund managed by [redacted], and over time it has adopted a more team-based management approach, similar to other systematic strategies. Currently, the fund is managed by a nine-person alpha equity team, which is part of [redacted]'s 30-person quant equity group. Leadership has traditionally been long-tenured [redacted] professionals, but in 2021, managers [redacted] and [redacted] retired, and [redacted], who joined the group in April 2020, became the sole lead in October 2021. He joined the team with solid quant investment experience, serving as the head of U.S. research at [redacted], where he was responsible for developing, testing, and implementing quant models.

The markets in 2021 have been very favorable to this strategy; in the first 11 months, the fund returned 23.8%, significantly outperforming the category benchmark's 17.8%. This comes after several challenging years when the fund underperformed both its prospectus and the category benchmarks for four consecutive calendar years. In the volatile first quarter of 2020, it underperformed the category index by almost 5 percentage points, owing primarily to an overweighting in small-value names, which as a group declined by about 50%.

A.1.3 Examples of causal statements identified using LLMs

Causal statements	**Cause**	**Effect**
Excluding securitized bonds tilts the fund toward Treasuries and corporates.	Excluding securitized bonds	The fund is tilted toward Treasuries and corporates
The portfolio also tends to have a lower duration than its category index and peers, which makes it less sensitive to changes in interest rates.	The portfolio tends to have a lower duration than its category index and peers	The fund is less sensitive to changes in interest rates
The speed and fluidity at which opinions are translated into prices makes it difficult to consistently beat the market.	The speed and fluidity at which opinions are translated into prices	it is difficult to consistently beat the market
Two comanager departures and a team addition bring some uncertainty to the strategy.	Two comanager departures and a team addition	Uncertainty to the strategy
The strategy's disciplined and conservative credit-driven process has demonstrated its value through time, but the analyst churn casts a shadow on its execution and puts a lid on our confidence level, supporting an Average Process Pillar rating.	The analyst churn	Casts a shadow on the strategy's execution and puts a lid on our confidence level
The China stake rose to 42% from 32% in the quarter ended March 2020, contributing to excellent returns for the remainder of 2020 as China successfully dealt with the virus and outperformed other emerging markets.	The China stake rose to 42% from 32% in the quarter ended March 2020 and China successfully dealt with the virus and outperformed other emerging markets.	This contributed to excellent returns for the remainder of 2020.

A.2 Pre-registration differences

This study was pre-registered in the AEA Randomized Controlled Trials registry with a pre-analysis plan. All aspects of the experimental sample, design, outcomes, and analyses were implemented without deviation other than the following key differences:

- The pre-analysis plan specified approximately 99 analysts to be part of the experiment, subject to turnover within the company during the experiment. Two analysts left the company before the start of the experiment, and resulted in a sample of 97 analysts.
- We run the same econometric specification as in the pre-analysis plan, and run additional specifications with committee fixed effects for reasoning scores. We observed that committees were heterogeneous in how they used the reasoning score scale, so included these fixed effects to reduce noise.
- The pre-analysis plan describes all possible primary and secondary outcomes, some of which we noted as potentially not being available due to partner and budget constraints. We were indeed not able to obtain some of the outcomes. We also report an additional outcome that was not available pre-experiment and transform one of the pre-registered variables:
  - We pre-specified an outcome measuring reasoning scores additionally provided by MBA students. Due to budget constraints accompanied by the number of reports we would have had to have MBA students score to sufficiently reduce noise, we were not able to undertake this effort. We were also not able to obtain the internal algorithm recommendation quality for each fund rating, although we verified the quality on average.
  - Instead, we obtained an additional variable to measure causal reasoning in the reports, by leveraging Large Language Models to identify causal statements in the draft. We included this result as it provides more insight into how analysts wrote their explanations.
  - We pre-specified an outcome called "complexity of decision drivers". We ask analysts "what are the key data points that most helped make your decision?" To answer this survey question, analysts can multi-select from key data points that contribute to their decision. This provided us a measure of the number of key causal drivers behind their decision. We later learned from the company, after we pre-registered the experiment analysis, that reports that clearly articulate its logical path and reasoning process tend to focus on a more selective set of causal drivers instead of having a laundry list of factors. Leveraging this insight from the field, we focus on the number of decision drivers as another way to provide deeper insight into causal explanations, which we natural-log transformed given the skewness in the distribution. In addition to the first order analysis on the number of decision drivers, we further created two measures that capture the "complexity of decision drivers", per specified in the pre-analysis plan. The key decision drivers reported by analysts were labeled into categories, which could be segmented into breadth and depth of decision drivers.
  - We report all possible exploratory analyses on outcomes in the appendix, as well as heterogeneous treatment effects on gender and prior experience with security. We could not report on two variables that we pre-specified – search for external knowledge and committee request to collect additional information. For the former, analyst responses varied substantially and could not be meaningfully compared quantitatively. For example, analysts' responses ranged from high level answers (e.g."manager interview" and "firm report"), to detailed description of factors and their sources. For the second variable, there were less than 1% of cases where the committee requested to collect additional information, so we did not analyze or report this variable.
- We pre-specified confidence as a primary outcome in addition to reasoning, but received feedback that the measures may not necessarily reflect confidence as they are self-reported. We thus report all of our results on stated confidence in the appendix, which shows that treatment on average boosted stated confidence prior to conducting interviews, but not after.
- We specified cutting tenure at the median in the pre-analysis plan, but cut tenure by new analysts in the main paper results. We learned after we received the data that three years was the cutoff for new

analysts to be promoted officially to analyst levels, as well as the difficulty of learning the company methodology. We report all results cut at the median in the appendix.