Journal of Financial Economics 160 (2024) 103910

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/finec

From Man vs. Machine to Man + Machine: The art and AI of stock analyses ☆

Sean Cao [a], Wei Jiang [b,*], Junbo Wang [c], Baozhong Yang [d]

[a] Robert H. Smith School of Business, University of Maryland, United States of America
[b] Goizueta Business School, Emory University, United States of America
[c] E.J. Ourso College of Business, Louisiana State University, United States of America
[d] J. Mack Robinson College of Business, Georgia State University, United States of America

ARTICLE INFO

Dataset link: Data_Submit_2.7z (Original data)

JEL classification:
C55
G12
G14
G17
G24
O33

Keywords:
Artificial intelligence
Machine learning
FinTech
Stock analyst
Alternative data
Disruptive innovation

ABSTRACT

An AI analyst trained to digest corporate disclosures, industry trends, and macroeconomic indicators surpasses most analysts in stock return predictions. Nevertheless, humans win "Man vs. Machine" when institutional knowledge is crucial, e.g., involving intangible assets and financial distress. AI wins when information is transparent but voluminous. Humans provide significant incremental value in "Man + Machine", which also substantially reduces extreme errors. Analysts catch up with machines after "alternative data" become available if their employers build AI capabilities. Documented synergies between humans and machines inform how humans can leverage their advantage for better adaptation to the growing AI prowess.

1. Introduction

Since the inception of artificial intelligence (AI) and as it continues to rise, AI has constantly made human beings rethink their own roles. While AI is meant to augment human intelligence, concerns abound that it could replace humans in increasingly skilled tasks and thus displace jobs currently performed by better paid and highly educated workers (Muro et al., 2019). Such concern and the associated debates have motivated a rapidly growing literature. Recent work by Webb

☆ Dimitris Papanikolaou was the editor for this article. The authors are grateful to helpful suggestions from the editor and two anonymous referees. We have benefited from discussions with Simona Abis (discussant), Jules van Binsbergen (discussant), Francesco D'Acunto (discussant), Svetlana Bryzgalova (discussant), Si Cheng (discussant), Jungho Choi (discussant), Will Cong, Jin-chuan Duan, Michael Gofman (discussant), Paul Goldsmith-Pinkham (discussant), Jillian Grennan, Bing Han, Jingyu He (discussant), Gerard Hoberg, Byoung-Hyoun Hwang, Mark Liu (discussant), Bin Li (discussant), Roxana Mihet (discussant), Markus Pelger, Uday Rajan, Siew Hong Teoh, Paul Tetlock, Cristian Tiu (discussant), Eric Yeung (discussant), Haifeng You (discussant), Rodrigo Verdi, Joe Weber, Michael Weber, Hong Zhang, Xiaoyan Zhang, Guofu Zhou, Luofeng Zhou, and Christina Zhu (discussant), and comments and suggestions from seminar/conference participants at the NBER Economics of AI Conference, the Stanford Engineering AI & Big Data in Finance Research Forum (ABFR) webinar, AFA, AllianceBernstein, Bank of Japan, Baruch, Case Western, CICF, CFEA, CFRC, CFTRC, CKGSB, Copenhagen Business School, CUHK-Shenzhen, Econometric Society China Meetings, ESSEC, FARS, Fordham, Frankfurt School of Finance & Management, Fintech: Innovation, Inclusion and Risks Conference, HARC, ITAM, Louisiana State University, MIT Sloan, National University of Singapore, Northern Illinois University, Stockholm School of Economics AI conference, Texas A&M, Tsinghua PBC, University of Cambridge, University of Florida Machine learning Conference, University of Mississippi, University of New Orleans, the Annual Conference on Digital Economics, the Quantitative Work Alliance at Boston (QWAFAFEW), the 28th Finance Forum at the Nova School of Business and Economics, NEOMA, the Midwest Finance Annual Conference, the Northern Finance Association conference, Penn State Empirical Accounting Conference, 3rd Shanghai Financial Forefront Symposium, SFS Cavalcade Asia, and the UT Dallas Finance Conference. We appreciate the computing resources offered by the Advanced Research Computing Technology and Innovation Core at GSU, High Performance Computing from LSU and excellent research assistance by Soyun Nam and Meng Wang.

* Corresponding author.
E-mail addresses: scao824@umd.edu (S. Cao), wei.jiang@emory.edu (W. Jiang), junbowang@lsu.edu (J. Wang), bzyang@gsu.edu (B. Yang).

https://doi.org/10.1016/j.jfineco.2024.103910
Received 21 June 2022; Received in revised form 6 July 2024; Accepted 9 July 2024
Available online 22 July 2024

(2020), Acemoglu et al. (2022), Babina et al. (2024), and Jiang et al. (2023) has all conducted large-sample analyses on the extent of job exposure and vulnerability to AI-related technology, as well as the consequences for employment and productivity.

The existing literature has mostly been focusing on characterizing the type of jobs that are vulnerable to disruption by AI's evolution, as well as those it could create. In other words, the sentiment of existent studies mostly involves a theme of "Man versus Machine", which characterizes the contest between humans and AI, explores the ways humans adapt and predicts the resulting job redeployments. In such settings, human beings are often rendered passive or reactive, dealing with disruptions and looking for new opportunities defined by the AI landscape. There has been relatively little research devoted to prescribing how skilled human workers could tap into higher potential with enhancement from AI technology, which is presumably the primary goal for humans to design and develop AI in the first place. This study aims to connect the contest of "Man versus Machine" ("Man vs. Machine" hereafter) to a potential equilibrium of "Man plus Machine" ("Man + Machine" hereafter).

Our study could be motivated by the experience of chess grandmaster Garry Kasparov. The story that IBM's Deep Blue beat the then reigning grand master in 1997 is well known. Afterwards, multiple contests repeated in a similar setting killed any remaining suspense for the outcome of Man vs. Machine in chess playing. What is far less known is that humans, despite having lost interest in man-versus-machine chess contests, have not lost interest in either the game or the machine. In fact, the encounter with Deep Blue was a catalyst for people like Kasparov to pioneer the concept of Man + Machine matches, in which a chess player equipped with AI assistance (a "centaur" player) competes against AI. Up to today, the centaur has kept an upper hand against machines; and even more encouragingly, there have been more and better human chess players with the advent of affordable AI-based chess programs.[1]

If AI can help more humans become better chess players, it stands to reason that it can help more of us become better at many skilled jobs, including pilots, medical doctors, and investment advisors. In this study, we zoom in on the profession of stock analysis, whose data availability allows us to calibrate both Man vs. Machine and Man + Machine. Stock analysts are among the most important information intermediaries in the marketplace (e.g., Brav and Lehavy, 2003; Jegadeesh et al., 2004; Crane and Crotty, 2020). Their job, which requires both institutional knowledge and data analytics, has not been spared by AI, as making powerful and fast predictions at a relatively low cost is at the heart of the technology (Agrawal and Goldfarb, 2018). More and more investors have begun to heed AI-powered recommendations about stock picking and portfolio formation.[2]

To trace the path from "Man vs. Machine" to "Man + Machine", we decided to build our own AI model for 12-month stock returns predictions (inferred from 12-month target prices), to be compared to analyst forecasts made at the same time on the same stock. Such a process provides a consistent and time-adapted benchmark for AI performance that we understand and are able to explain. Earnings and target prices (from which one could infer stock returns) are the two most important subjects of analyst forecasts. We choose the latter as our primary target variable because earnings are more subject to managerial discretion, often made in the direction to meet (or slightly beat) consensus analyst forecast,[3] making it an unfair comparison for the AI model for which such a feedback effect is absent.

Our "AI analyst" is built on training a combination of current machine learning (ML) tool kits using timely, publicly available data and information. More specifically, we collect firm-level, industry-level, and macroeconomic variables, as well as textual information from firms' disclosures, news, and social media (updated to right before the time of an analyst forecast), as inputs or predictors. We deliberately exclude information from analyst forecasts (past and current) themselves so that the AI model does not benefit from analyst insights. Machine learning models have been shown to outpace traditional economics models (such as regressions) in such a setting thanks to their advantages in managing high-dimensional and unstructured data, and in their flexibility in optimizing and fitting unspecified functional forms. More recent development in the area has made significant progress in mitigating overfitting and thus improving out-of-sample performance.

We select a set of state-of-the-art machine learning models and build our AI analyst based on an ensemble model. Our AI analyst is able to beat human analysts as a whole: the AI analyst outperforms in 54.5% of the stock return predictions made by all I/B/E/S analysts during the sample period of 2001 to 2018. The machine's advantage could arise from its superior ability to process information, or its immunity from predictable human biases due to incentives or psychological traits (e.g., Abarbanell, 1991; Stickel, 1990). To separate the two, we compare AI predictions with "debiased" analyst forecasts where biases are predicted and then removed using machine learning (henceforth, "Machine–debiased Man" or "MDM" forecasts). Such an improved version of human analyst still trails the machine (MDM only outperforms AI in 46.5% of forecasts), suggesting that "correctable" biases explain around 22% of the Man–Machine gap.[4]

The power of AI aside, we are more interested in knowing the circumstances under which human analysts retain their advantage, in that a forecast made by an analyst beats the concurrent AI forecast in terms of lower squared forecast error relative to the ex post realization (i.e., the actual 12-month stock price). We find that human analysts perform better for smaller and more illiquid firms and those with asset-light business models (i.e., higher intangible assets), consistent with the notion that such firms are subject to higher information asymmetry and require better institutional knowledge or industry experience to decipher. Analysts affiliated with large brokerage houses also stand a higher chance of beating AI thanks to a combination of their abilities and the research resources available to them. Analysts are more likely to have the upper hand when the firm is in a dynamically competitive environment or is subject to higher distress risk, again revealing AI's limitation in analyzing unfamiliar and rapidly evolving situations.[5] As expected, AI enjoys a clear advantage in its capacity to process information and is more likely to outsmart analysts when the volume of public information is larger.

Just like the centaur chess player which Kasparov pioneered, the superior performance of an AI analyst does not rule out the value of human inputs. If humans and machines have relative advantages in information processing and decision making, then human analysts can still contribute critically to a "centaur analyst". After we add analyst forecasts to the information set of the machine learning models

[1] Source of information: *The Inevitable*, by Kevin Kelly, Penguin Publishing Group, 2016. See also "Defeated chess champ Garry Kasparov has made peace with AI", *Wired*, February 2020.

[2] Sources: "What machine learning will mean for asset managers", Robert C. Pozen and Jonathan Ruane, *Harvard Business Review*, December 3, 2019. "How startup investors can utilize AI to make smarter investments", Jia Wertz, *Forbes*, January 18, 2019.

[3] A large accounting literature documents such an effect (e.g., Abarbanell and Lehavy, 2003; Doyle et al., 2013). The cash flow (as opposed to the accrual) component of the earnings is less discretionary; but unfortunately only 3.9% of the I/B/E/S earnings observations contain separate forecasts on the cash flow component, making a large sample comparison infeasible.

[4] Since analysts outperform AI in 45.5% of the cases, the percentage attributable to bias correction is (46.5 - 45.5)/(50 - 45.5) = 22%.

[5] This is consistent with the limitation of current machine learning and AI models which still lack reasoning functions to handle unfamiliar situations well. Source: "What AI still can't do", Brian Bergstein, *MIT Technology Review*, February 19, 2020.

underlying our AI analyst, the resulting "Man + Machine" model outperforms 54.8% of the forecasts of the AI-only model. Furthermore, we find inputs from analysts are more valuable when covering firms that are more illiquid and those with more intangible assets. In addition, analyst input has more incremental value when a firm faces a higher risk of distress. Importantly, the incremental value of humans does not decrease as the volume of information (hence the demand for processing capacity) increases, though this constitutes a disadvantage for humans working alone.

The synergy between humans and machines is correlated with, but goes beyond, the incremental information value of the analyst forecast to that made by AI. An alternative measure for the synergy could be uncovered from the residuals in forecast errors by Man + Machine that are not explained by forecast errors by either side alone. Synergies turn out to be correlated with characteristics that provide human advantage (trading illiquidity and close to default) as well as those that favor machine advantage (frequent corporate events and large market cap). Moreover, man–machine synergies are higher when data is relatively sparse and situations are fast evolving. Perhaps most importantly, the Man + Machine model avoids about ninety percent of extreme errors made by analysts and forty percent of those by the AI (while with minimal creation of its own large errors). To the extent that large errors are calamitous in many skilled professions, there is substantial benefit in combining human and AI capabilities.

Finally, we resort to an event study to sharpen the inference of the impact of integrating humans and machines in stock analyses. In recent years, the infrastructure of "big data" has created a new class of information about companies that is collected and published outside of the firms, and this information provides unique and timely clues about investment opportunities. An important and popular type of alternative data captures "consumer footprints", often in the literal sense, such as satellite images of retail parking lots. Such data, which must be processed by machine learning models, have been shown to contain incremental information on stock prices (Zhu, 2019; Katona et al., 2024). We build on data from (Katona et al., 2024) on the staggered introduction of several important alternative databases and conduct a difference-in-differences test of analysts' performance versus our own AI model before and after the availability of the alternative data. The underlying premise is that analysts who cover firms while using this alternative data could be in the situation of Man + Machine, as they have the opportunity to use the additional AI-processed information. Indeed, we find that post alternative data, analysts covering affected firms improve their performance relative to the AI model we build, which serves as a benchmark. Furthermore, such improvement concentrates on the subset of analysts who are affiliated with brokerage firms with strong AI capabilities, measured by AI-related hiring using the Burning Glass U.S. job posting data.

Overall, this study supports the hypothesis that analyst capabilities could be augmented by AI, and more importantly, that analysts' work possesses incremental value to and synergies with AI modeling, especially in unusual and fast-evolving situations. In high-stake situations, "centaur players" are particularly helpful in subduing severe prediction errors. This unique finding has important implications for the safety and robustness of AI assistants while retaining the critical roles of humans in decision making. If there is some external validity from chess and stock analysis to skilled workers in general, the inference from our study provides guidance on how humans can leverage their own strength and be better adapted in the age of AI.

2. Literature, data construction, and machine learning models

2.1. Relation to the literature

Our work is related to the rapidly growing literature on the competition and threat to human workers posed by new technology including robots and AI.[6] This literature overall finds that when low- or intermediate-skill jobs are replaced by machines, humans tend to move to high-skill jobs that are more difficult to replace (Autor et al., 2003). While most recent studies highlight how AI innovations disrupt many high-skill jobs, our study focuses on humans' relative advantage over machines and, more importantly, the potential synergies between humans and machines. We envision a future in which AI and machines can assist humans with the more tedious and quantitative tasks and democratize access to information, while allowing humans to be more creative and productive.[7]

A few recent and contemporaneous papers also study the impact of big data and AI in the financial industry. Abis (2022) studies how quantitative investment strategies influence mutual fund performance. Abis and Veldkamp (2024) examine the change in labor shares in the financial industry driven by the new data management and AI jobs. Coleman et al. (2022) compare the performance of robot analysts from fintech companies with that of human analysts. Grennan and Michaely (2020, 2021) study how analysts perform and adjust in response to the advent of AI-processed recommendations in the markets. Rossi and Utkus (2021) compare human asset managers with robot advisors. Agrawal et al. (2019b) discuss the ambiguous impact of AI on labor given the elements of AI that tend toward automating decisions versus enhancing human decisions. Jansen et al. (2023) analyze human and machine decisions in loan underwriting. Cao et al. (2023b) study the impact of AI readership on corporate disclosure policies. Finally, Pagliaro et al. (2023) consider human interactions with algorithmic wealth management advisors. Our paper differs from the existing literature in that we explore the internal mechanism of the AI process we constructed ourselves instead of market-level proxies,[8] and aim to identify their relative advantages to, as well as synergies with, humans using model inputs and outputs in our own hands.

We also contribute to the literature of building and assessing the performance of machine learning models in financial applications, such as in predicting asset prices (Gu et al., 2020, Brogaard and Zareei, 2022), robo-advising (D'Acunto and Ross, 2019), managing portfolios (Chen et al., 2023; Cong et al., 2022), estimating values of artwork (Aubry et al., 2023), forecasting earnings (van Binsbergen et al., 2023; Cao and You, 2024, de Silva and Thesmar, 2024), making lending decisions (Liu, 2022), classifying and evaluating innovations (Chen et al., 2019; Zheng, 2022) and estimating bank risk (Hanley and Hoberg, 2019).[9] While the structure of our analysis shares similarity with some of the papers, notably de Silva and Thesmar (2024) and van Binsbergen et al. (2023) which calibrate biases in analyst expectations regarding earnings using author-developed AI model, the primary research questions of our paper are different from theirs. de Silva and Thesmar (2024) and van Binsbergen et al. (2023) focus on the term structure of analyst biases, and link them to corporate actions such as security issuance, while our ultimate goal is to explore the complementary value humans can offer in the age of AI once we have a good understanding of their relative advantage.

[6] An incomplete list of recent papers includes Aghion et al. (2019), Acemoglu and Restrepo (2018, 2019), Brynjolfsson et al. (2018), Webb (2020), Ray and Mookherjee (2022), Cao et al. (2023a), Acemoglu et al. (2022), and Jiang et al. (2023).

[7] Due to the complementary nature of AI and humans, the advent of AI technologies can potentially create more jobs than they destroy. See "Artificial intelligence to create 58 million new jobs by 2022, says report", Amit Chowdry, *Forbes*, September 18, 2018.

[8] For example, Grennan and Michaely (2020) resort to the amount of social media information as a proxy for the AI research intensity for a stock, and focus on analysts' response to the AI shock. This study, in contrast, aims at decipher the nature of the AI shock.

[9] See also Cong et al. (2020), Martin and Nagel (2022) and Goldstein et al. (2021) for surveys and discussions of methodologies.

In summary, our study contributes to the emerging literature studying the implications of combining humans and machines in the financial markets.[10] Given the increasing presence of machines and AI beyond finance, we also hope that this case study contributes to a better understanding of how technology can complement and improve humans, bringing to fruition the original mission of AI development.[11]

2.2. Sample of forecasts

Stock analysts routinely make forecasts on corporate earnings, price targets, and much less often, sales and operating profits. We consider both earnings and stock returns (implied by price target forecasts) in our analyses but choose the latter as the main forecast objects for two reasons. First, earnings are more subject to managerial discretion compared to stock returns. If decomposing earnings into the cash and accrual components, then managerial discretion is even larger over the later part. Moreover, management exercises discretion in earnings formation often in the direction of meeting or slightly beating analyst forecasts (Abarbanell and Lehavy, 2003; Doyle et al., 2013), creating a "feedback effect" that draws analyst forecasts closer to the realized earnings as opposed to the true fundamentals. On the other hand, AI-issued earnings forecasts (used in our study) are not managerial targets. For this reason, it will not be a fair comparison between analyst and AI forecasts in terms of matching announced earnings. In contrast, such a feedback effect is largely absent for long-term target prices (from which we could impute returns).

Second, valuation targets have grown in importance over cash flow projections for investors. Early study by Asquith et al. (2005) finds that target price forecasts provide important information, and the market reacts strongly to target price revisions. This trend has been growing with the "new economy". For example, after Nvidia announced its financial results in May 2023, dozens of analysts promptly revised their 12-month price targets and made recommendations based on price, but not nearly as much on earnings revisions. The Nvidia story is not a lone event. In fact we find that analysts revise price target more frequently than earnings forecast, and with the gap increasing from 2001 to 2018, consistent with the argument that real-time information about price targets receive more attention with investors, and their importance is growing over time.[12]

For these reasons, we use analysts' 12-month price forecasts as the main target in our analyses. We also present results based earnings forecasts for all main analyses, and explain the differences accordingly. Our sample of analyst forecasts builds on the Thomson Reuters I/B/E/S analyst database using data from 1996 to 2018.[13] For target price forecasts, we choose the 12-month horizon because the target prices for other horizons are negligible (less than 1%). We consider earnings predictions from one quarter up to four quarters as those are the most common horizons for earnings forecasts. After merging I/B/E/S with CRSP and Compustat, the final sample consists of 1,153,565 12-month target price forecasts on 6315 firms issued by 11,890 analysts from 861 brokerage firms, and 5,885,063 1-quarter to 4-quarter earnings predictions on 8062 firms issued by 14,363 analysts from 926 brokerage firms.

2.3. Building the information set for the AI analyst

Given our goal to build an AI analyst to compete with professional analysts, we need to define the information set available to such a professional whenever a price forecast is made. We illustrate the process to forecast the 12-month stock price for firm i by human analyst k on date t (in year u), while that for earnings follows analogously. The information set, $\mathcal{I}_t$, would, in an ideal setting, include all publicly available data and information up to $t-$. We assume that professional analysts do not have access to material nonpublic information, which is essentially the requirement of Regulation FD.[14] We approximate $\mathcal{I}_t$ with firm and industry information from CRSP and Compustat; textual information from firms' SEC filings, including annual reports (10-K), quarterly reports (10-Q), material corporate news and developments (8-K), and other reports, news sentiment from Ravenpack, and social media coverage from Twitter and macroeconomic data from the Federal Reserve Economic Data at the Federal Reserve Bank of St. Louis and recent research papers.

To operationalize time adaptation, we adopt the following rolling-window approach. For a given forecast made by a human analyst on date t in year u, all forecasts in the previous three years $u-3, u-2, u-1$ form the training sample. That is, data up to the dates of those forecasts (but excluding the forecasts themselves) and the corresponding realized prices were used to train our machine learning models. Moreover, if the past three years include a "distress" year (2001, 2008, and 2009), we expand the training window to the first year after the preceding distress period or the start year of the dataset. The benefit of this approach is that human analysts in a recession likely predict future prices based on information over a full business cycle. Including the years before distress can mimic the information used by human analysts. Moving to the estimation sample, we then feed data available up to date $t-1$ in year u into the trained model to make the 12-month-price prediction at time t. Our AI analyst makes its first prediction in 2001. Though we allow (public) information to be updated till $t-1$, most of the information inputs came from disclosed quarterly data from the previous eight quarters.

We considered longer training windows that include all past years so that the training period covers one or more full business cycles. Perhaps surprisingly, the longer training period results in slightly worse performance (e.g., the ratio at which the AI analyst could beat human analysts in forecasting accuracy). This highlights the trade-off between more data in training for better fitting and robustness in out-of-sample prediction, and the importance of a stable information structure from the training to prediction. A short-rolling window (three-year) ensures that the relation among firms' fundamental conditions and external states remains relatively stable. Perhaps for this reason, the three-year rolling window has been standard in the asset pricing literature (e.g., Narasimhan et al., 2019 for stock returns; Elton et al., 1996 for mutual funds). Therefore, we keep the three-year rolling window as the default setting, but report results from the long-window specification in Table IA.1 of the Internet Appendix.

[10] In different settings, Armour et al. (2022) study the impact of AI and the associated digital technologies on the law profession. They find that AI-enabled services will augment the capabilities of human lawyers and also generate new roles for legal experts to produce such services. Brogaard et al. (2024) find that human floor traders can complement algorithmic traders in providing information to the market in complicated environments.

[11] This echoes the mission of the Stanford Human-Centered AI Institute, "to advance AI research, education, policy and practice to improve the human condition". See https://hai.stanford.edu/about.

[12] See "Nvidia Stock Hits New Closing High as Chip-Maker's Valuation Approaches $1 Trillion", May 25, 2023, Will Feuer, *Wall Street Journal*, available at https://www.wsj.com/articles/nvidia-shares-jump-as-chip-maker-approaches-1-trillion-valuation-7f8ccd68. See Table IA.7 in the Internet Appendix for the time series of frequencies in earnings and price targets revisions.

[13] The I/B/E/S coverage prior to 1996 was limited, with fewer than 2,000 target price forecasts in total.

[14] Regulation FD ("fair disclosure"), implemented in 2000, generally prohibits public companies from disclosing previously nonpublic, material information to certain parties unless the information is distributed to the public first or simultaneously.

2.4. Information and variables as inputs to machine learning

Firm characteristics. The firm characteristics fed into machine learning models are retrieved or processed based on information from standard databases accessed via WRDS, especially CRSP/Compustat and Thomson Reuters Ownership databases. The first set of predictors include stock prices at the end of the previous month as well as the stock prices one to four years before the end of the previous month. The 12-month returns over the past 5 years are also included, together with the realized earnings within the past 3, 6, 9, 12, 24, and 36 months. We also include a number of firm characteristics known to predict cross-sectional differences of the stock prices.[15] Variables in this group are constructed quarterly using information available at the previous quarter-end.

Industry variables. We compose a set of industry-level variables that capture competition, industry dynamics, and other factors relevant for firm valuation based on the existent literature. These variables include (i) The competition measure from 10-Ks following (Li et al., 2013), which captures the degree of competition resulting from rivalry within and across industries as perceived by the management; (ii) the product market fluidity measure following (Hoberg et al., 2014), which quantifies the product market poaching threat posed by the movement of competitors toward the focal firm; (iii) industry affiliation with the Fama–French 12 industries (12 indicator variables); (iv) industry size, measured by the number of firms in the Fama–French 12 industry within the past 3, 6, 9, 12, 24, and 36 months; and (v) equally weighted industry average earnings per share realized within the past 3, 6, 9, 12, 24, and 36 months.

Macro variables. Macroeconomic and stock market development are common factors to all firms' valuation and returns (e.g., Fama and French, 1989; Chen et al., 1986). We first include the following variables: (i) Industrial Production Index; (ii) Consumer Price Index; (iii) Crude oil price (WTI); (iv) three-month treasury bill rate; (v) ten-year treasury constant-maturity rate; and (vi) The BAA–AAA yield spread, obtained from the Federal Reserve Economic Data at the Federal Reserve Bank of St. Louis on a monthly frequency. We also consider macro state variables commonly used in the business cycle literature, including (1) dividend yield (Fama and French, 1989), (2) stock market illiquidity (Chen et al., 2018), (3) new orders (Jones and Tuzel, 2013); (4) technical indicators (Neely et al., 2014); (5) average correlation of largest stocks in the SP500 index (Pollet and Wilson, 2010). These variables are constructed in Goyal et al. (2024) (and have been shown to be related to asset prices by Wayne et al. (2024)) at the monthly frequency.[16] While macroeconomic state variables serve as predictors/inputs, our machine-learning model is able to incorporate their interactions with other variables, effectively making them conditioning variables as well.

Corporate disclosure, news, social media, patents. One leading strength of AI over human beings is the former's ability to digest large volume of information. One new edge that machine learning models boast over traditional statistical methods is the capacity to process unstructured textual data based on firms' SEC filings (including annual reports (10-K), quarterly reports (10-Q), corporate news (8-K), and other reports), news articles, and social media. The new developments allow researchers to quantify information which was considered qualitative or "soft", commonly termed "sentiments".

Several different sets of sentiment variables from textual data serve as inputs to our AI analyst. The first is based on the (Loughran and McDonald, 2011) sentiment, which has been widely used in the academic literature. We calculate the frequency of positive and negative sentiment words from the firm-issued SEC filings following (Loughran and McDonald, 2011). The second set of machine-learning-based sentiment variables follow (Cao et al., 2021b), who trained a deep-learning neural network model to incorporate contextual information and syntactic relations between performance-related words. The second approach aims to isolate managerial sentiment related to the firm's future performance from sentiment regarding other issues (such as location and weather). The third set of variables are macro and firm-specific news sentiment variables from Ravenpack, which covers more than 40,000 news media sources globally. Finally, we include Twitter sentiment variables following (Cao et al., 2021a). Patents filed by firms contain important information regarding firms' innovation capacity and future growth prospects. We thus provide the annual firm-level patent value following (Kogan et al., 2017) as an additional input for the AI analyst.

2.5. Potential factors for the relative performance of AI and human analysts

A main objective of the study is to assess the factors that contribute to the relative performance of AI vs. humans as well as the synergy between the two. We hypothesize that these factors are related to the information environment of the firm, industry, and analysts. Needless to say, equity analysts are often evaluated along dimensions other than forecast accuracy, such as promoting investment banking or trading businesses, and intermediating between firms and large investors. We focus on forecast accuracy, not only because it is objective and quantifiable, but also because it represents a primary quality in analyst evaluation (Stickel, 1992; Desai et al., 2019).

We first consider the following firm-level variables: the *Amihud Illiquidity* measure (Amihud, 2002), which is the ratio of absolute daily stock return to the daily trading volume (in dollars); *Log Market Cap*, the logarithm of market capitalization; *Number of Information Events*, which is the number of firm-specific information events within 90 days in Capital IQ Key Development data and represents the volume of available information about the firm; *Text Complexity*, given by the Flesch Kincaid Grade Level of corporate filings, that measures the complexity and difficulty of texts based on U.S. school grade levels; and *Intangible Assets*, defined as the first principal component of three proxies: one minus the ratio of PP&E to total assets, organization capital scaled by assets, and knowledge capital scaled by assets. The other two measures, derived from the accumulation of SG&A and R&D expenditures, are constructed following (Ewens et al., 2024) (see e.g., Eisfeldt and Papanikolaou, 2013, 2014; Peters and Taylor, 2017; Falato et al., 2022 for the modeling and development of these and related measures of intangible capital).

We further include a number of variables that characterize the information environment and resources for analysts: *Star Analyst* is an analyst with an average forecast accuracy in the top quintile among all analysts for the past six months; *# Analysts in Brokerage Firm* is the number of analysts and proxies for the size and resources of the brokerage firm; *% Institutional Holdings* is the 13F institutional holdings as a percentage of shares outstanding, which can reflect the prevalence of informative investors; *Distance to Default* is the distance to default calculated following (Merton, 1974) and proxies for default risk for firms;[17] *Industry Recession* is an indicator variable that equals one if the FF-48 industry return in the previous year is negative and in the bottom quintile of Fama–French 48 industry returns and zero otherwise; *Fluidity* represents the competition firms face in the product markets by tracing changes in rival firms' products relative to the firm's products (Hoberg et al., 2014); and *Time Trend* equals the number of years elapsed from the beginning of the sample (2001). The final set of variables are related to analysts' access to alternative data and AI resources, which will be introduced in Section 5.3. Table 1 presents the summary statistics of variables.

[15] We list all variables serving as inputs into the machine learning models, their definitions, and sources in Table A.1 in Appendix A.

[16] We are grateful to Amit Goyal for sharing the data with us.

[17] We thank the National University of Singapore's Credit Research Initiative for providing the distance to default data.

Table 1
Summary statistics.
This table reports the summary statistics of key variables. The firm-level, industry-level, and macroeconomic variables are defined in Section 2.5. The AI and alternative data variables *AI Hiring*, *Alt Data Covered*, and *Post* are defined as follows. textitAI Hiring is the ratio of the number of AI jobs to the total number of job postings. *Alt Data Covered* is an indicator variable equal to one if alternative data are available for the firm by the end of the sample, and if the firm is in an industry with retail footprints. *Post* is an indicator variable equal to one if a "treated" firm has been covered by alternative data. For "untreated" firms, *Post* is coded one if the year is after 2014. The mean, median, standard deviation, 25 percentile, 75 percentile, and number of observations for the panel data are reported in the table.

Variables	Mean	Median	Std	P25	P75	N
Panel A. Firm-level, industry-level, and macroeconomic variables						
Amihud Illiquidity	0.40	0.00	56.10	0.00	0.02	352 358
Intangible Assets	0.01	−0.13	1.11	−0.34	−0.04	352 358
# Analysts in Brokerage Firm	0.03	0.04	0.01	0.03	0.04	352 358
Star Analysts	0.03	0.00	0.16	0.00	0.00	352 358
Number of Information Events	2.64	2.64	0.69	2.30	3.00	352 358
Distance to Default	5.44	5.20	2.64	3.52	6.80	352 358
Market Cap	8.07	7.98	1.68	6.87	9.21	352 358
% Institutional Holdings	0.66	0.77	1.85	0.53	0.90	352 358
Fluidity	7.19	6.48	3.66	4.50	9.14	352 358
Text Complexity	16.04	15.96	1.22	15.40	16.55	352 358
Industry Recession	0.11	0.00	0.31	0.00	0.00	352 358
Panel B. AI and alternative data variables						
AI job	0.01	0.00	0.07	0.00	0.00	56 697
Alt data cover	0.02	0.00	0.40	0.00	0.00	56 390
Post	0.66	1.00	0.47	0.00	1.00	56 697

2.6. Machine learning models

There are a number of candidate machine learning models developed in recent decades to build our AI analyst, including lasso, elastic-net, support vector machines, random forest, gradient boosting, and long short-term memory neural networks. Because the machine learning models are essential tools but not the ultimate objectives of this study, we provide an overview in Appendix B of the models referenced where we outline the economic intuition of each methodology's mechanism and strength without going into in-depth technical details. For further details, we refer the reader to representative references in this field; for example, Goodfellow et al. (2016) and James et al. (2023). Of the models considered, random forest, gradient boosting, and long short-term memory neural networks are the state-of-art nonlinear models that have been increasingly used and proved of their advantage over the other methods in the existent literature of computer sciences, finance, and other disciplines. We note that the long short-term memory model is a time-series machine learning model which specializes in learning various time-series patterns, such as momentum and reversals, from past data. With complementarity, the random forest and gradient boosting models are adept at learning complex cross-sectional relations among variables. Our main AI model thus is built as an ensemble of these three models, i.e., adopts the median prediction of the three models.

Our candidate machine learning models strive to be at the leading edge of AI practice in investment management. They are similar to those covered in two prominent industry reports: The JP Morgan Big Dada and AI Strategies report and the report on Artificial Intelligence in Asset Management by the CFA institute, and are also favored in the current industry practice.[18]

3. Construction and performance of the AI analyst

While constructing an AI analyst per se is not the ultimate objective of this study, it remains a crucial milestone for our AI analyst to reach a level of refinement where it can compete with, or even surpass, human analysts. This achievement will allow us to investigate both the relative strengths and the synergistic potential between machines and humans. This section outlines the methodology employed in developing an AI analyst of such caliber.

3.1. The predictive models

For each stock i at time t, where t is the day when an analyst makes a forecast, $F^{Man}_{i,j,t}$ (wherein i, j, t are indices for the stock, the analyst, and the date, and the superscript Man indicates human as opposed to AI), of the 12-month target price, we convert the forecast to the corresponding 12-month return $R^{Man}_{i,j,t}$ for stationarity and comparability in the cross section. Because analysts make predictions about future prices, and do not systematically provide dividend forecasts, the stock returns we infer from the target prices exclude dividends (Brav and Lehavy, 2003). Included in the predictive information set is all public information (as described in Section 2 and Appendix A) up to $t-$. We summarize the prediction model as

$$R_{i,t,T} = R^{AI}_{i,t} + \epsilon_{i,t},\ \ R^{AI}_{i,t} = f_{t-}(X_{i,t-}). \tag{1}$$

Here, f_{t-} is the prediction function for all stocks at time $t-$. This is consistent with asset pricing models with conditioning information; that is, we assume there is a uniform prediction model for every stock at a given time while allowing the model to be time-varying.

Next, we compare the AI forecast $R^{AI}_{i,t}$ and its human counterpart $R^{Man}_{i,j,t}$ in terms of their accuracy based on the ex-post realized return $R_{i,t,T}$. AI beats human if $|R^{AI}_{i,t} - R_{i,t,T}| < |R^{Man}_{i,j,t} - R_{i,t,T}|$, and vice versa. We define $Beat$ to be an indicator variable for human analyst winning against AI. Fig. 1 shows the relative performance of AI vs. human analyst forecasts over time. Out of 922,157 forecasts made from 2001 to 2018, human analyst beats AI in 45.5% of the time. The p-value for the percentage to be drawn from a distribution with the neutral probability of half for this sample size is less than 0.01%. However, the human analyst disadvantage is volatile from year to year, with a weak time trend of improvement.

3.2. Contribution of variables to the AI prediction

The process of machine learning is often opaque. We mitigate the opacity by examining the contribution of different groups of input variables to the predictions of the AI model. Specifically, we divide

[18] These reports can be found at the following links: https://www.cognitivefinance.ai/single-post/big-data-and-ai-strategiesand https://zonavalue.com/wp-content/uploads/2020/09/CFA-Institute-artificial-intelligence-in-asset-management.pdf. We have presented and discussed our paper and models with about half a dozen teams who are leaders in AI-directed investments. Most importantly, we confirmed with these teams that the rates at which AI models beat human analysts are on par with the current state-of-art.

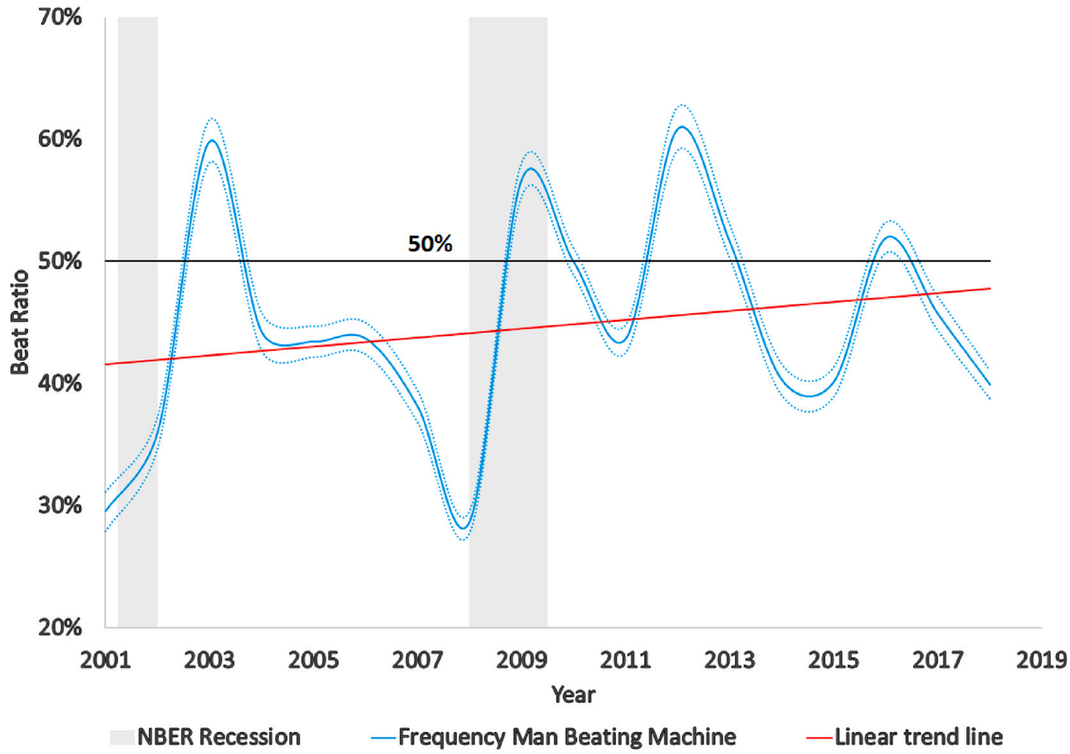

Fig. 1. Man vs. machine: the performance of analysts vs. AI.
This figure plots the beat ratio, or the proportion of analysts' price forecasts that are more accurate than the corresponding AI price forecasts in each year. The blue line in the middle plots the annual beat ratios, and the surrounding blue-dotted lines indicate the 95% confidence interval of the beat ratio. The red line gives the best linear approximation of the time-series trend in beat ratios. The shaded gray bars represent the NBER recessions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the features into six groups: return-based variables, firm characteristics, earnings (firm and industry), industry information, macroeconomic variables, and variables extracted from textual information. We compute the contribution of the groups by taking one set of variables out at a time. Specifically, we consider a drop-one-set AI model by setting the values of a given group of variables to their past average values in the trained AI model. The contribution of each variable group is defined as the difference between the squared forecast error of the full AI model and the drop-one-set AI model, scaled by the sum. By construction, the contributions of all groups of variables sum up to unit. Fig. 2 presents the composition.

Each group of features contributes substantially to AI prowess. Macro Variables and firm returns contribute the most (27.6% and 24.4%, respectively), followed by firm characteristic variables (22.0%). The 9.3% contribution from textual information highlights the importance of qualitative information. It is perhaps not surprising that information from earnings claims the lowest share (2.0%), as such information is likely already impounded in past returns and other firm characteristic variables.

3.3. Debiased analysts vs. AI

It has been well documented that analysts exhibit biases in their forecasts (e.g., Abarbanell, 1991; Stickel, 1990). There are a multitude of explanations of such biases, including the incentive to issue more favorable forecasts for corporate clients of the analysts' affiliated brokerage firms (Michaely and Womack, 1999), the need to obtain access to information from the management (Lim, 2002), and human psychological traits (e.g., De Bondt and Thaler, 1990; Hirshleifer et al., 2019). A natural question thus arises: Could human underperformance relative to AI be remedied simply by "debiasing" analyst forecasts with a machine learning model (henceforth, "Machine–debiased Man" forecasts or "MDM"), or will the human shortfall remain after such a procedure in which case it would be due to the limitation in human ability to acquire and process information? A comparison of MDM forecasts with the AI analyst would reveal the nuance regarding the innate predictive ability of analysts after filtering out their predictable biases.

In constructing the MDM forecasts, we first predict the analyst forecast errors in the next period with all current information, analogous to Eq. (1).

$$\log(1 + R^{Man}_{i,j,t}) - \log(1 + R_{i,t,T(t)}) = g(X_{i,t-}, Z_{i,j,t-}) + \epsilon_{i,t}, \tag{2}$$

where we include all variables $X_{i,t-}$ that we have employed to predict target returns,[19] and a set of analyst and brokerage-firm characteristics $Z_{i,j,t-}$, including the mean and standard error of analysts' past prediction biases, analysts' experiences (number of years covering the firm, the industry, or any public firm), analysts' efforts (whether the analysts provide forecasts of additional information such as sales or cash flows), and brokerage firm size proxied by the number of analysts. We use the same procedure as in Section 3.1 to train the same machine learning model and then estimate (2). The MDM prediction is then the analyst prediction $R^{Man}_{i,j,t}$ minus the corresponding bias as predicted by the machine learning model. To compare the MDM with AI, we plot the MDM beat ratio, or the frequency of MDM forecasts beating AI forecasts, in each year from 2001 to 2018 in Fig. 3. As expected, MDM exhibits better performance than the raw forecasts and beats the AI more frequently than human analysts alone in most years.

Over the entire sample, MDM beats AI analysts in 46.5% of the cases, a one percentage-point improvement over humans without debiasing. Since pre-debiased analysts outperform AI in 45.5% of the cases, the enhancement amounts to a 22.2% (= (46.5%–45.5%)/(50%–45.5%))

[19] We estimate forecast errors in the logarithm of returns to alleviate the influence of outliers in forecast errors.

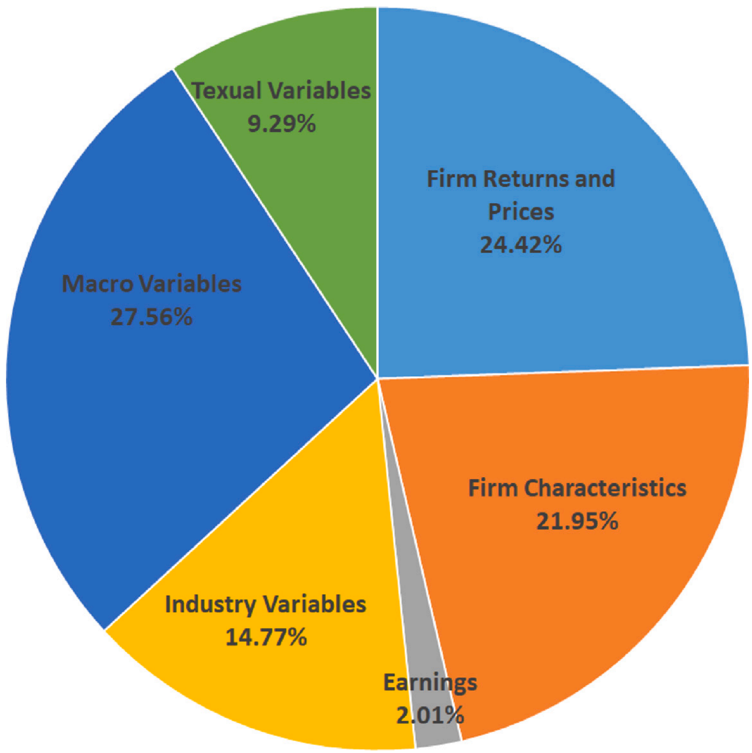

Fig. 2. Contribution of groups of variables to the AI prediction: take-one-out approach.
This figure plots the contribution of each group of features to the AI model's price prediction. The features are divided into six groups: Firm Returns and Prices (past returns and prices of stocks), Firm Characteristics, Earnings (past firm and industry earnings), Industry Variables, Macro Variables, and Textual Variables. We employ all variables to train the model but adjust the predictors for the exclusion group to the past average value over the entire time in our sample. Afterwards, we calculate the squared errors for the predicted values. We then compute the average squared errors for each year and aggregate them across all years. The contribution of each variable group is expressed by the increase in the average squared errors of each group relative to the sum of increments across all groups.

closing-up of the Man–Machine gap. Given that analyst characteristics are likely to be correlated with mis-incentives and agency issues and hence their biases, we also use only analyst-specific variables $Z_{i,t-}$ to predict analyst biases to produce a "lower-bound MDM". The lower-bound MDM beats Machine 46.2%, implying an improvement of (46.2%–45.5%)/(50%–45.5%) = 15.6% over the analysts.[20] We believe both the upper and lower bounds are meaningful to calibrate how much analyst forecasts could be improved upon if we can predict the biases therein, since De Bondt and Thaler (1990), Michaely and Womack (1999), Lim (2002), and Hirshleifer et al. (2019) suggest that both incentive- and cognitive-driven bias are associated with both firm and analyst characteristics.

3.4. AI vs. Analysts with persistent performance

Analysts are a large group with heterogeneous skill levels such that forecast performance would be persistent if skills were innate. Moreover, the market recognizes, at least partially, the relatively more skilled analysts by responding more strongly to their forecasts or recommendations (Chen et al., 2005; Li, 2005; Mikhail et al., 2007). Thus, a higher hurdle is for our AI analyst to beat the subset of skilled analysts. We assess the relative performance with respect to the higher hurdle with two tests. First, we sort all analysts into the top and bottom halves based on their average prediction error (normalized by stock prices) over a past period with length ranging one, two, three, four, and five years. We then track the percentage of their future forecasts that beat our AI analyst during each time period. In the second test, we repeat the same procedure except selecting the analysts that are among top and bottom quantiles each year during the past one, two, three, four and five years. The second specification is more demanding on persistent skills as only about 7.3% of the analysts are able to stay at the top half in each year for a five-year period. Table 2 reports the results.

Results in Table 2 show that the AI comfortably beats the analysts in the low-skill quantiles. It is basically neck and neck to the more successful analysts and is almost even with analysts (analyst beat ratio of 49.3%–50.3%) who demonstrated superior performance in each of the past five years, an excellence only achieved by less than one tenth of all analysts.

3.5. Performance of portfolio following AI recommendations

Analysts make forecasts as a way to advise portfolio formation or turnover. The performance of a portfolio following the analyst's advice is thus a natural metric for analyst skill. For the same reason, we can form portfolios based on the different opinions between the AI and human analysts. The performance of the resulting portfolio is a testament of their relative proficiency. Our approach is different from the usual one that follows analyst directional recommendations as our model requires a clear investment horizon that is lined up with the horizon of the signal (i.e., 12-month price target).

In each month, we gather all predictions made by all analysts and the corresponding AI forecasts in the past 30, 60, 90, 180, and 360 days. For each pair of predictions, if the Machine's prediction is greater than the median of Machine predictions in the prior month and the human's prediction is less than the median analyst prediction in the prior month, we define it as a buy signal. When both conditions are negated, we define it as a sell signal. During the given time horizon, the portfolio will long the stock if there are more buy than sell signals and short the stock otherwise. The portfolio is value-weighted. In a semi-annually rebalanced portfolio, we hold the position for at least six months or

[20] We thank one referee for suggesting this lower bound of the improvement brought by de-biasing.

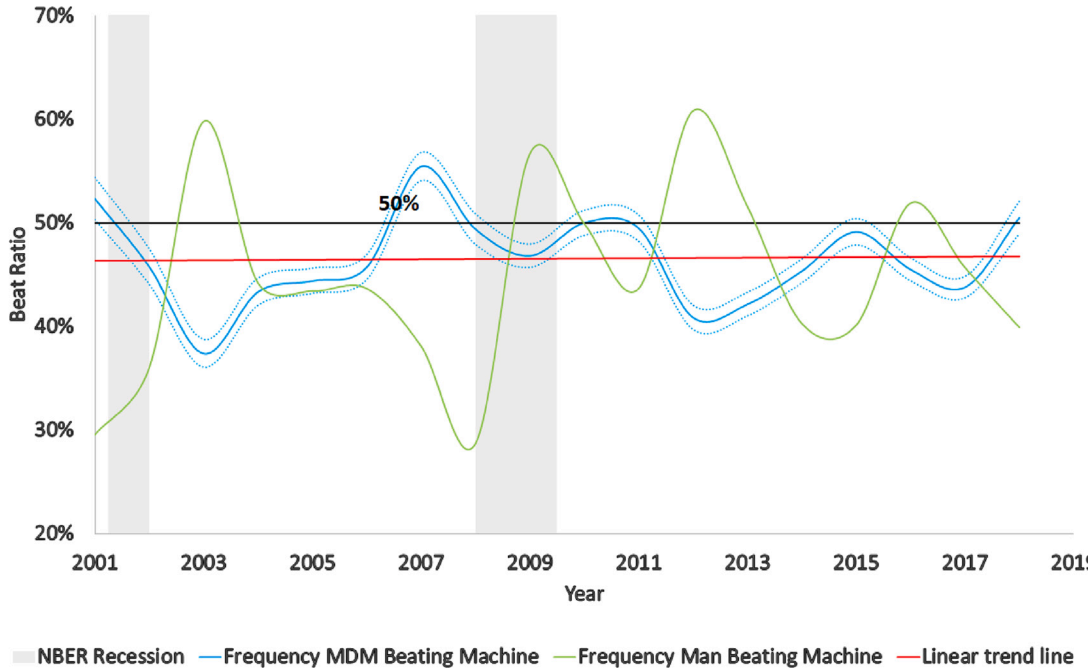

Fig. 3. Man + Machine: the performance of machine–debiased analyst vs. AI.
This figure plots the proportion of machine–debiased analyst (MDM) price forecasts that are more accurate than the machine recommendations alone on an annual basis. The blue line in the middle gives the annual machine–debiased analyst beat ratios, the blue-dotted lines above and below are the 95% confidence interval of the beat ratio, the green line represents the analyst beat ratios, and the red line gives the best linear approximation of the trend in beat ratios. The shaded gray bars represent the NBER recessions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Persistence of performance of AI analyst.
Each year, analysts are sorted by mean squared prediction errors of log prices based on the past one, past two, and up to five years. If the mean squared error over the last year is below (above) the median during the specified past period, the analyst is in the top (bottom) in the current year. In Panel A, the sorting is based on the full period of the past one, two, …, five years. In Panel B, the sorting requires that an analyst be in the top half in each of the past one, two, …, five years to be placed in the "top" group. Both panels report the analyst beat ratio, i.e., the number of times analysts beat AI, as a proportion of total number of predictions.

Panel A: Analyst beat ratio sorted by analysts who are above/below median

	1 year	2 years	3 years	4 years	5 years
Analyst top	49.25%	49.26%	49.11%	49.10%	49.08%
Analyst bottom	43.01%	42.95%	43.04%	43.04%	43.05%

Panel B: Analyst beat ratio sorted by analysts who are above median each of the past years

	1 years	2 years	3 years	4 years	5 years
Analyst Persistent top	49.25%	50.33%	49.87%	49.63%	49.40%
Analyst Persistent bottom	43.01%	42.34%	41.39%	40.95%	40.79%

until the signals reverse. The results are robust to equal weighting or a different re-balancing frequency such as one month. The portfolio contains 420 to 785 stocks with signals from past 30 days to 360 days.[21] Table 3 reports the performance of the semi-annually rebalanced long-short portfolio in terms of average return and alpha estimated using Fama–French three-factor, Carhart four-factor, Fama–French five-factor and Fama–French six-factor models.

Results in Table 3 are encouraging in that the AI model is able to generate superior returns/alpha, relative to analysts, on the order of 50 to 72 basis points monthly, which are also statistically significant at the 1% level in almost all cases. To the extent that our portfolio approach compares the AI with the median of all human analysts, our result implies that the AI forecasts outperform analyst consensus. When we separately examine the long and short portions of the portfolio, we discover that the superior returns, while significant on both sides, are larger in magnitude and more significant for the long side (for which transaction costs are lower). Such an asymmetry could be driven by the well-documented positive bias in analyst forecasts (Lim, 2002), that is, analyst signals tend to be less informative when they are more optimistic than AI as the latter does not share the positive bias.

To further examine the differences between AI and human predictions, we separately examine the performance of portfolios based on AI and human forecasts. In Table IA.10 in the Internet Appendix, we present the quintile-sorted portfolio returns based on the machine-only and human-only signals, respectively. The results show that portfolio returns increase monotonically across the quintiles based on machine-only signals, suggesting a strong performance of our AI model. The analyst-only signal is not informative, but its underperformance is not pivotal to the performance of Man vs. Machine as shown in the 5 × 5 double-sorted portfolio returns reported in Table IA.11. The results show that machine and human predictions are relatively independent: In each quintile sorted by human signals, the top minus bottom machine signal-based portfolio exhibits similar abnormal returns. The same is true if we condition on machine signals first and examine

[21] The average monthly turnover rate of the semi-annually rebalanced portfolios ranges is around 12%.

Table 3
Portfolio performance following machine vs. man recommendations.
At the beginning of each month, we gather all predictions made by all analysts and the corresponding AI forecasts in the past 30, 60, 90, 180, and 360 days. For each pair of predictions, if the Machine's prediction is greater than the median of Machine predictions in the month and the human's prediction is less than the median Machine prediction in the month, we define it as a buy signal. When both conditions are negated, we define it as a sell signal. During the given time horizon, the portfolio will long the stock if there are more buy than sell signals and short the stock otherwise. The portfolios are value-weighted and rebalanced every six months, i.e., a position is held for six months or till the signals reverse. The monthly percentage returns of the long-short, long-leg (stocks only with a buy sign) and short-leg portfolios (stocks only with a short sign), as well as the alphas generated from the FF3, FFC4, FF5, and FF6 models, are presented. The OLS standard error is used to construct t-stats. The t-stats are reported in parentheses. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tailed), respectively.

Portfolio returns — AI vs. Analyst						
Machine vs Human						
Long-Short		30 day inform	60 day inform	90 day inform	180 day inform	360 day inform
Monthly returns	Ret	0.64***	0.65***	0.58***	0.60***	0.53***
		(3.06)	(3.67)	(3.33)	(3.79)	(3.66)
	FF3	0.69***	0.72***	0.64***	0.65***	0.61***
		(3.40)	(4.19)	(3.80)	(4.24)	(4.45)
	FFC4	0.62***	0.65***	0.58***	0.58***	0.52***
		(3.16)	(3.97)	(3.57)	(4.03)	(4.21)
	FF5	0.51**	0.62***	0.55***	0.55***	0.51***
		(2.54)	(3.73)	(3.37)	(3.72)	(3.74)
	FF6	0.51***	0.63***	0.56***	0.56***	0.50***
		(2.62)	(3.92)	(3.54)	(4.02)	(4.09)
Long-Leg		30 day inform	60 day inform	90 day inform	180 day inform	360 day inform
Monthly returns	Ret	0.92***	0.97***	0.93***	0.97***	0.98***
		(2.69)	(2.96)	(2.83)	(3.07)	(3.32)
	FF3	0.42***	0.48***	0.44***	0.44***	0.42***
		(3.12)	(4.12)	(3.99)	(4.53)	(5.09)
	FFC4	0.39***	0.47***	0.42***	0.41***	0.38***
		(2.93)	(3.98)	(3.84)	(4.34)	(4.84)
	FF5	0.37***	0.46***	0.42***	0.40***	0.37***
		(2.76)	(3.90)	(3.91)	(4.28)	(4.43)
	FF6	0.37***	0.46***	0.42***	0.40***	0.36***
		(2.81)	(3.92)	(3.94)	(4.40)	(4.59)
Short-Leg		30 day inform	60 day inform	90 day inform	180 day inform	360 day inform
Monthly returns	Ret	−0.28	−0.32	−0.35	−0.37	−0.44
		(−0.78)	(−0.89)	(−0.99)	(−1.09)	(−1.33)
	FF3	0.27**	0.23***	0.20	0.21***	0.19***
		(2.49)	(2.82)	(2.44)	(2.83)	(2.63)
	FFC4	0.23**	0.18**	0.16**	0.17**	0.13**
		(2.22)	(2.51)	(2.10)	(2.52)	(2.18)
	FF5	0.14	0.16**	0.13	0.15**	0.14**
		(1.27)	(2.02)	(1.58)	(2.01)	(1.98)
	FF6	0.14	0.16**	0.13*	0.15**	0.14**
		(1.30)	(2.29)	(1.74)	(2.27)	(2.17)

human signals' performance. Further, Figure IA.1 shows a roughly balanced composition of the doubly sorted portfolios, confirming a low correlation between machine and human selection of stocks.

3.6. Combined wisdom of man + machine

Results from the previous sections suggest that the analyst profession could be seriously disrupted by AI technology given that analysts trail AI for a majority of the time. However, the superior performance of the AI analyst does not rule out the possibility that analyst forecasts contain valuable information that is incremental to AI-produced forecasts. In other words, if analysts possess information that is not picked up by the AI, then the AI forecast is not sufficient to replace the analyst forecast, even though analysts lose to AI in forecasting accuracy. An investor who combines the wisdom of both should attain even better performance.

To assess the performance of the combined analytical power, we consider adding the analyst forecasts to the information set for our machine learning model. That is, the information set $\mathcal{I}_t$ now includes the analyst forecasts, $F_{i,j,t}$, made on the same firm i during the 90-day window ending on date t. In particular, we obtain analyst and brokerage-firm characteristics (including analysts' experiences, analysts' efforts, and the number of analysts in the brokerage firm), the mean forecast and forecast accuracy (mean square error) of each analyst for the past five years, the mean consensus and forecast accuracy (mean square errors) of the forecasts by all analysts in the previous 90 days, current analysts' predictions, Machine–debiased Man predictions, and stand-alone machine predictions and build a "Man + Machine" hybrid analyst using an ensemble model. We find that the hybrid analyst outperforms human analysts 57.8% of the time and outperforms AI-alone forecasts 54.8% of the time.

Fig. 4 illustrates the comparative performance of the hybrid analyst (Man + Machine) versus the plain AI (Machine). The fact that Man + Machine outperforms Machine-alone in 15 out of 18 years, and with the three lagging years having beat ratio close to being neutral (49.98%, 48.28%, and 47.98%), is encouraging evidence for the added value of analysts. While the future of AI remains uncertain, the parts of human skills that are incremental to AI, as we document, allow for promising Man + Machine collaboration and augmentation. We note from the Figure that the Man + Machine advantage over Machine remains stable over the 18-year period. Such a result projects a hopeful outlook for skilled analysts who can be augmented by AI, instead of being replaced by it.

3.7. Alternative forecast target: earnings

Though stock returns are our primary target variable of forecast, we nevertheless present the main results with earnings forecasts. Using the same procedure as outlined in Section 2, we find that human

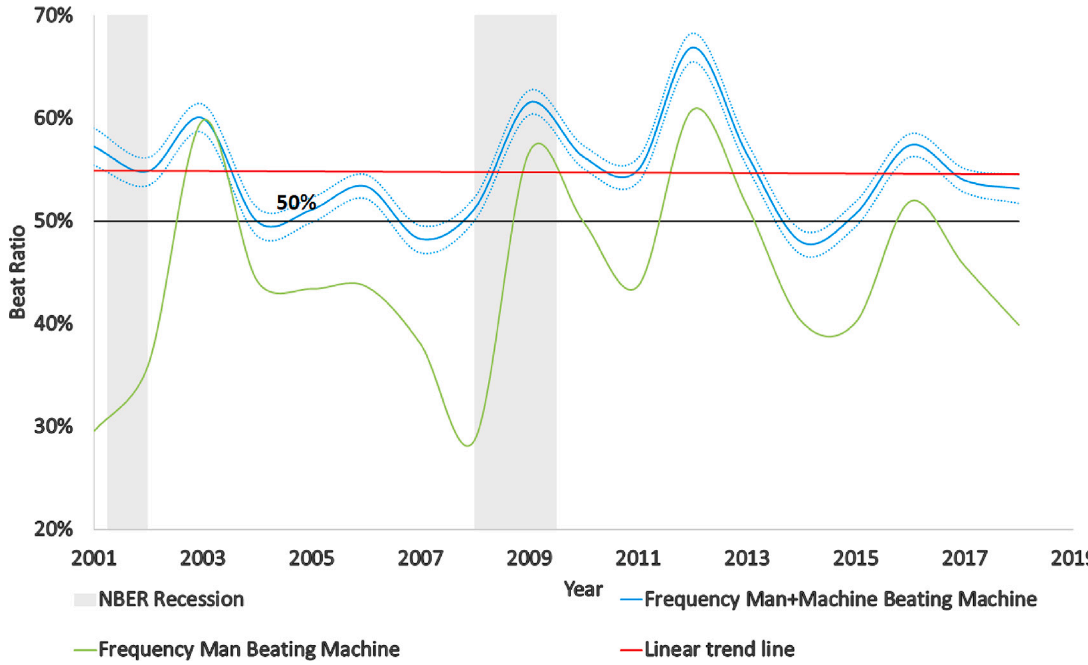

Fig. 4. Man + Machine: the performance of AI-assisted analysts vs. AI.
This figure plots the proportion of AI-assisted analyst price forecasts that are more accurate than the AI recommendations alone on an annual basis, or the "beat ratio". The blue line in the middle gives the annual AI-assisted analyst beat ratios, the blue-dotted lines above and below are the 95% confidence interval of the beat ratio, and the red line gives the best linear approximation of the trend in beat ratios. The shaded gray bars represent the NBER recessions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

analysts beat machine with a probability 69.2%, significantly higher than the beat ratio corresponding to return forecasts, consistent with the hypothesis (discussed in Section 2) that analysts benefit from firms' desire and ability to produce earnings that match market expectation for which analyst forecasts serve as the most important proxy. However, the M+M model still comes above both analysts (55.1%) and stand-alone machine (71.8%), confirming high synergies in information production between Man and Machine. Therefore, the key thesis of our study about combining the wisdom of man and machine remains robust with earnings predictions. We present the results in Table IA.2 in the Internet Appendix.

Extensive accounting literature suggests that the information content of earnings varies significantly when they are broken down into cash flow and accrual components. Since accruals are subject to management discretion, often utilized to meet earnings targets (Dechow et al., 1995), we anticipate that AI would enjoy greater predictive power in forecasting the cash flow component. To this end, we focus on a subsample of I/B/E/S firm-quarter observations where analysts provided separate forecasts for the cash flow components. We compare these forecasts to the realized values using the methodology pioneered (Sloan, 1996). It turns out that the human-beat-machine ratio is 45.6% when it comes to predicting the cash flow component of earnings, mirroring the performance observed in 12-month return prediction. This contrast between total earnings and cash-component forecasts supports our hypothesis that managerial discretion in earnings management contributes to analysts' outperformance over AI in earnings prediction. Because of the limited sample size (only 3.9% of the I/B/E/S data), we present this analysis, as a robustness check, in Table IA.3 in the Internet Appendix.

4. Comparative advantages of man vs. machine

4.1. Determinants of relative performance

In this section, we strive to understand when human analysts perform better than the AI and when otherwise. Such understanding will help "unbox" the black box associated with AI or machine learning and provide intuition and guidance on the applicability of AI for researchers and investors.

We consider a number of variables at the analyst, firm, and industry levels that are potentially relevant for the performance of human analysts and AI. These are defined in Section 2.5. We group these variables into several classes. First, we consider a number of proxies for information asymmetry or opacity, including *Amihud Illiquidity*, *Log Market Cap*, and *% Institutional Holdings*. Second, we include variables representing the volume of information (*# Information Events*), readability of corporate filings (*Text Complexity*), and the tangibility of information (*Intangible Assets* and *Fluidity*). Third, we examine several variables that affect the information and resources available to the analyst, such as *# Analysts in Brokerage Firm* and *Star Analyst*. Finally, we consider *Distance to Default* and *Industry Recession*, highlighting the financial exposure of firms to shocks, and *Time Trend*, which can help capture temporal patterns.

For each target price forecast, we define two variables that measure the outcome of relative performance of humans vs AI. First, the indicator variable *Analyst Beats AI* equals one if the absolute value of forecast error of the analyst is smaller than that of the AI, and zero otherwise. Second, the continuous measure *Forecast Error Difference* is the difference between the absolute prediction error (of return as defined in Eq. (1)) of the AI and that of the analyst, scaled by the maximum of these two prediction errors. A positive and large value of *Forecast Error Difference* is in favor of analyst accuracy.

We estimate the following regression on the panel data of firm i, analyst j, and date t to understand the determinants of the relative strengths of humans and AI,

$$Relative\ Performance_{i,j,t} = X'_{i,j,t}\beta + \alpha_i + \alpha_j + \alpha_{year} + \epsilon_{i,j,t}, \qquad (3)$$

wherein the dependent variable *Relative Performance* is either *Analyst Beats AI* or *Forecast Error Difference*. The vector of independent variables, $X_{i,j,t}$, includes those discussed in Section 2.5, and α_i/α_j and α_{year} represent firm/analyst and year fixed effects, respectively. The results are reported in Table 4.

Table 4
Man vs. Machine: The relative advantage of analyst vs AI.
This table presents the coefficients and *t*-stats of regressing the *Analyst Beats AI* indicator (Panel A) and the *Forecast Error Difference: Analyst vs. AI* (Panel B) on the firm-level, industry-level, and macroeconomic variables presented in Table 1. *Analyst Beats AI* is an indicator variable equal to one if the analyst beats the AI. *Forecast Error Difference: Analyst vs. AI* is defined as the difference between absolute prediction errors between the AI and the analysts, divided by the maximum value of these two prediction errors. The number is positive if the analyst has a smaller absolute forecast error, i.e., the analyst beats AI. The *t*-statistics are based on standard errors clustered at the firm level. The *t*-statistics are based on standard errors clustered at the firm level. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tailed), respectively.

Panel A: Analyst Beats AI

Variables				
Amihud Illiquidity	0.252***	0.218***	0.200***	0.119**
	(11.51)	(6.06)	(6.39)	(2.24)
Intangible Assets	0.027***	0.026***	0.021**	0.023***
	(4.09)	(4.57)	(2.49)	(2.88)
# Analysts in Brokerage Firm	0.700***	0.931***	0.109	0.501
	(5.28)	(7.14)	(0.29)	(1.34)
Star Analyst	−0.106	−0.290	0.373	0.207
	(−0.17)	(−0.48)	(0.51)	(0.29)
# Information Events	−0.026***	−0.012***	−0.025***	−0.010**
	(−5.81)	(−2.68)	(−5.47)	(−2.16)
Distance to Default	−0.009***	−0.005***	−0.014***	−0.011***
	(−8.16)	(−3.75)	(−12.55)	(−8.41)
Market Cap	−0.071***	−0.049***	−0.085***	−0.055***
	(−15.18)	(−10.84)	(−16.79)	(−10.79)
% Institutional Holdings	−0.033**	−0.011	−0.012	−0.004
	(−1.99)	(−0.58)	(−0.70)	(−0.43)
Fluidity	0.243*	−0.273**	0.529***	0.032
	(1.89)	(−2.02)	(4.13)	(0.24)
Text Complexity	−0.001	0.001	−0.003	0.000
	(−0.67)	(0.65)	(−1.25)	(0.09)
Industry Recession	0.012**	0.030***	0.027***	0.044***
	(2.07)	(5.06)	(4.28)	(6.84)
Time Trend	0.010***		0.009***	
	(11.90)		(8.60)	
Year Fixed Effect	No	Yes	No	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes
Analyst Fixed Effect	No	No	Yes	Yes
Observations	352,358	352,358	352,358	352,358
Adjusted R-squared	0.07	0.09	0.13	0.15

Panel B: Forecast Error Difference: Analyst vs. AI

Variables				
Amihud Illiquidity	0.292***	0.263***	0.205***	0.129***
	(8.60)	(11.03)	(5.52)	(3.71)
Intangible Assets	0.027***	0.026***	0.024***	0.025***
	(3.92)	(4.04)	(3.03)	(3.31)
# Analysts in Brokerage Firm	0.762***	0.990***	0.435	0.787*
	(5.47)	(7.14)	(1.07)	(1.95)
Star Analyst	0.105	−0.082	0.436	0.262
	(0.16)	(−0.13)	(0.57)	(0.35)
# Information Events	−0.022***	−0.013**	−0.023***	−0.011**
	(−4.30)	(−2.35)	(−4.37)	(−2.04)
Distance to Default	−0.008***	−0.004***	−0.013***	−0.010***
	(−6.41)	(−2.82)	(−10.09)	(−6.91)
Market Cap	−0.068***	−0.047***	−0.077***	−0.049***
	(−13.76)	(−9.23)	(−14.62)	(−8.71)
% Institutional Holdings	0.031***	0.050*	0.056***	0.061***
	(2.77)	(1.83)	(5.71)	(3.81)
Fluidity	0.117	−0.383**	0.418***	−0.080
	(0.76)	(−2.32)	(2.81)	(−0.50)
Text Complexity	−0.000	0.002	−0.001	0.001
	(−0.16)	(0.67)	(−0.45)	(0.54)
Industry Recession	0.014**	0.030***	0.027***	0.042***
	(2.00)	(4.33)	(3.53)	(5.56)
Time Trend	0.009***		0.008***	
	(9.35)		(7.02)	
Year Fixed Effect	No	Yes	No	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes
Analyst Fixed Effect	No	No	Yes	Yes
Observations	352,358	352,358	352,358	352,358
Adjusted R-squared	0.10	0.12	0.16	0.17

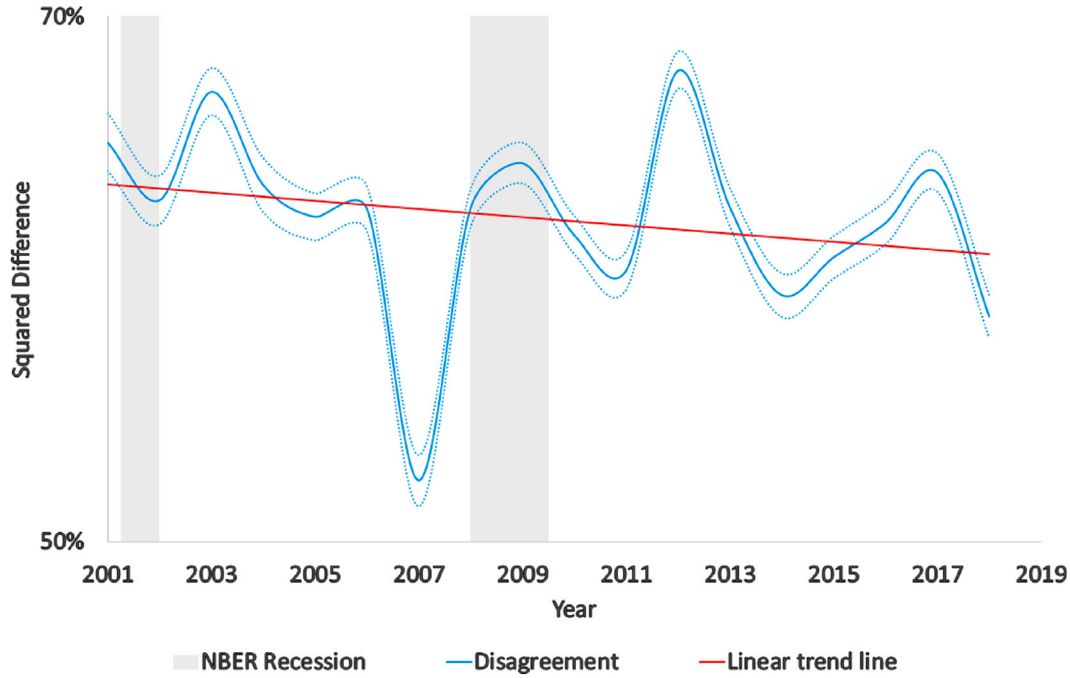

Fig. 5. Man vs. machine: disagreement.
This figure plots the disagreement between man and machine. The disagreement is defined as the squared difference between the returns predicted by the analysts and the AI. Each year, the average value of the disagreement is calculated. The blue line in the middle gives this average disagreement, the blue-dotted lines above and below are the 95% confidence interval of the disagreement, and the red line gives the best linear approximation of the trend. The shaded gray bars represent the NBER recessions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4 shows that, controlling for year and firm fixed effects, humans are more likely to outperform when covering illiquid and small firms and those with higher intangible assets, consistent with the notion that such firms are subject to higher information asymmetry and require deeper institutional knowledge to understand. A one-standard deviation increase in *Intangible Assets* is associated with a 3.0% increment on beat ratio. On the other hand, equipped with vast processing power, AI performs better for firms with a larger volume of disclosed information, as proxied by *# Information Events* each year. A one-standard deviation increase in *# Information Events* is associated with a 1.8% decrement on beat ratio. Analysts working for larger brokerage firms perform better, potentially because of the more abundant resources and research capacity at such places as well as a positive match between analyst skill and brokerage house prestige.

Humans perform better when the focal firm is subject to higher financial distress risk, captured by distance to default and industry recession, suggesting that the AI has more difficulty handling more uncertain scenarios. Analysts also perform better for firms with higher institutional holdings, possibly because analysts are immersed with information produced and processed by institutional investors, including brokerage houses. Finally, when year fixed effects are not included, we are able to uncover the time trend of the comparative performance, showing that human advantage increases with time. This is probably due to the fact that human analysts are increasingly assisted by AI and big data technologies. Perhaps surprisingly, *Star Analysts* do not demonstrate significantly superior performance over AI, suggesting that the superior ability star analysts possess over their peers could be replicated by the machine.

We also extend the portfolio return analysis in Table IA.8 of the Internet Appendix to subsamples sorted on the four firm characteristic variables that are expected to be associated with the relative performance of man vs. machine: Liquidity, intangibility, number of information events (voluminous information) and distance to default. Results show that portfolio performance is more favorable to the AI model when firms have fewer intangible assets, more voluminous information, and are far from default. The results for firm illiquidity are mixed. With 60 and 90-day information, portfolio performance is more favorable to the machine when firms are more liquid, but not for 30, 180, and 360 days. Overall, most results are consistent with regression results presented in Table 4.

4.2. Disagreement between man and machine

An equally important question is whom we should trust more if and when humans and machines disagree to a large extent. To start with, Fig. 5 plots the annual time series of the average squared differences in predicted returns between analysts and AI. Interestingly, we find that the man–machine disagreement has been on a downward trend, possibly because analyst forecasts increasingly incorporate insights from big data and AI tools. Further, the disagreement tends to be high before recessions, when high investor sentiment may exert a disproportional influence on analysts.

We next examine the relative performance of Man vs. Machine precisely when they disagree to a large degree. Gaining an understanding into such situations has significant implications for AI-guided decision making including investment. For each pair of forecasts, we define an indicator variable, *Disagreement*, to be one if the magnitude of the disagreement between the analyst and our AI model, normalized by the maximum value of these two prediction errors, is above the 90th percentile among all forecasts on the same firm within three years. Such benchmarking ensures that the disagreement could be measured on a similar scale. Conditional on the existence of a *Disagreement*, we further define one sub-indicator, *Human Wins*, equal to one when human has a lower absolute prediction error than machine. We then relate these outcome variables to the set of regressors, with results reported in Table 5. Because the regressions involve high-dimensional fixed effects, we apply the linear probability model.

The first two columns of Table 5 examine the relation between the occurrence of *Disagreement* and the underlying firm and analyst attributes and economic conditions in the full sample, with firm fixed effects (column (1)) or double firm/year fixed effects (column (2)).

Table 5
Disagreement between man and machine.
This table presents the coefficients and *t*-stats of regressing the *Disagreement* indicator on the firm-level, industry-level, and macroeconomic variables presented in Table 1. For each pair of forecasts, we define the indicator variable *Disagreement* to be one if the magnitude of the difference between the absolute prediction errors of the analyst and our AI model, normalized by the maximum value of these two prediction errors, is above the 90th percentile among all forecasts on the same firm within three years. Conditional on *Disagreement* being positive, we further a sub-indicator, *Human wins*, which equals 1 if human analyst has a lower absolute prediction error. We report regression results with *Human wins* in the subsample where *Disagreement* being positive. To calculate *t*-statistics, standard errors are clustered at the firm level. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tailed), respectively.

Dependent variable	Full sample		Disagreement sample	
	Disagreement		*Human wins*	
	(1)	(2)	(3)	(4)
Amihud Illiquidity	0.034***	0.034***	0.445***	0.383***
	(2.78)	(2.74)	(12.08)	(8.02)
Intangible Assets	0.005***	0.005***	0.029**	0.028**
	(3.07)	(2.87)	(2.05)	(2.18)
# Analysts in Brokerage Firm	−0.605***	−0.626***	1.055***	1.209***
	(−9.15)	(−9.39)	(3.32)	(3.86)
Star Analyst	−0.332	−0.284	4.129	5.603*
	(−0.93)	(−0.79)	(1.36)	(1.87)
# Information Events	−0.007***	−0.003	−0.020**	−0.011
	(−3.94)	(−1.51)	(−2.07)	(−1.09)
Distance to Default	−0.005***	−0.004***	−0.002	−0.002
	(−12.96)	(−8.55)	(−0.94)	(−0.90)
Market Cap	−0.003**	−0.005***	−0.062***	−0.041***
	(−2.04)	(−3.74)	(−8.26)	(−5.16)
% Institutional Holdings	−0.083***	−0.080***	−0.618	2.335
	(−5.50)	(−5.50)	(−0.30)	(0.80)
Fluidity	0.138***	0.062	0.251	−0.217
	(3.56)	(1.59)	(1.10)	(−0.93)
Text Complexity	0.000	0.001	−0.003	−0.000
	(0.43)	(1.01)	(−0.56)	(−0.09)
Industry Recession	0.018***	0.018***	−0.006	0.014
	(7.03)	(6.75)	(−0.50)	(1.09)
Time Trend	−0.000		0.007***	
	(−0.18)		(5.34)	
Year Fixed Effect	No	Yes	No	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes
Observations	352,358	352,358	42,023	42,023
Adjusted R-squared	0.01	0.01	0.10	0.13

Columns (3) and (4) focus on the subsample with *Disagreement* and examine when human wins. Results from the first two columns indicate that large disagreement tends to occur for illiquid, intangible firms with less abundant information, which are characteristics of firms that are associated with human comparative advantage. Indeed, the last two columns confirm that the same set of characteristics are also associated with human winning conditional on that the two sides disagree.

One exception to the pattern stated above is *# Analysts in Brokerage Firm*. Analysts from large brokerage firms are less likely to disagree with AI. However, conditional on large disagreement, these analysts are also more likely to beat the machine.

5. Man + Machine: Incremental contributions and synergies

5.1. Incremental value of analysts to man + machine & man–machine synergies

Acknowledging that Man + Machine is superior to either the human or machine alone, it is still instructive to understand the respective incremental values of the human and the machine in the combination. Analogous to the previous section, we define relative performance measures of the hybrid analyst vs the AI to capture the incremental value of humans. We then reestimate Eq. (3) with these relative performance measures as dependent variables. Table 6 presents the results.

Similar to the previous findings, we find inputs from analysts are more valuable when covering firms that are less liquid and firms with more intangible assets and earnings volatility. Moreover, analyst inputs have more incremental value when firms have higher distress risk. The institutional holding percentage also helps the hybrid model beat AI analyst. Finally, the incremental value of human does not decrease significantly with the volume of information, whereas humans alone are capacity-constrained as shown in Table 4. This finding supports AI augmentation of skilled professionals and highlights synergies between the two sides.

The synergy between humans and machines is correlated with, but goes beyond, the incremental information value of analyst forecast to that made by AI. An alternative measure for the synergy could be uncovered from regressing the squared (or absolute) forecast error of the Man + Machine model on the Man- and Machine-alone errors.[22] The residual term then proxies for the incremental value of Man + Machine above and beyond Man and Machine alone. We then take the residual and regress it on various firm and analyst characteristics to understand what drive this synergy. In this regression, we negate the sign of the residual so that a positive coefficient conveys a positive outcome, i.e., being associated with higher Man–Machine synergies.

Table 7 reports the results, showing that synergies could be correlated with characteristics that afford human advantage (trading illiquidity and close to default) as well as those favoring machine advantage (frequent corporate events and large market cap). Moreover, man–machine synergies are higher during recessionary times, where data is relatively sparse and situations are fast evolving. Such a mix suggests that drivers for synergies are distinct from comparative advantages between Man and Machine.

[22] To mitigate the influence of extreme forecast errors, we take the logarithm of one plus the squared (or absolute) forecast errors and truncate the top 1% largest values.

Table 6
Man + Machine: The incremental value of analyst.
This table presents the coefficients and *t*-stats of regressing the *Analyst + AI Beats AI* indicator (Panel A) and *Forecast Error Difference: Analyst + AI vs. AI* (Panel B) on the firm-level, industry-level, and macroeconomic variables presented in Table 1. *Analyst + AI Beats AI* is an indicator variable equal to one if Analyst + AI beats AI. *Forecast Error Difference: Analyst + AI vs. AI* is defined as the difference between absolute prediction errors between AI and Analyst + AI, divided by the maximum value of these two prediction errors. The number is positive if the analyst has smaller absolute forecast error, i.e., Analyst + AI beats AI.

Panel A. Analyst + AI Beats AI

Variables				
Amihud Illiquidity	0.274***	0.256***	0.181***	0.142**
	(6.30)	(4.31)	(4.37)	(2.53)
Intangible Assets	0.021***	0.018***	0.020***	0.018**
	(3.39)	(3.15)	(2.65)	(2.54)
# Analysts in Brokerage Firm	0.246*	0.321**	0.471	0.417
	(1.93)	(2.54)	(1.32)	(1.17)
Star Analyst	−0.001	−0.105	−0.113	−0.216
	(−0.00)	(−0.17)	(−0.15)	(−0.30)
# Information Events	−0.012***	−0.004	−0.013***	−0.003
	(−2.80)	(−1.02)	(−2.89)	(−0.62)
Distance to Default	−0.010***	−0.003**	−0.013***	−0.006***
	(−10.53)	(−2.34)	(−11.76)	(−4.35)
Market Cap	−0.025***	−0.017***	−0.036***	−0.023***
	(−6.40)	(−4.37)	(−7.77)	(−4.86)
% Institutional Holdings	0.041	0.061***	0.075***	0.089***
	(1.25)	(3.77)	(3.19)	(8.06)
Fluidity	0.344***	−0.076	0.618***	0.167
	(3.16)	(−0.67)	(5.27)	(1.32)
Text Complexity	0.002	0.005**	0.001	0.004*
	(1.23)	(2.41)	(0.41)	(1.69)
Industry Recession	0.024***	0.030***	0.031***	0.036***
	(4.33)	(5.20)	(5.03)	(5.71)
Time Trend	0.004***		0.004***	
	(6.54)		(4.01)	
Year Fixed Effect	No	Yes	No	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes
Analyst Fixed Effect	No	No	Yes	Yes
Observations	352,358	352,358	352,358	352,358
Adjusted R-squared	0.04	0.05	0.09	0.10

Panel B: Forecast Error Difference: Analyst + AI vs. AI

Variables				
Amihud Illiquidity	0.233***	0.224***	0.223***	0.207***
	(14.38)	(14.67)	(9.63)	(9.29)
Intangible Assets	0.012***	0.010***	0.012***	0.011**
	(2.95)	(2.67)	(2.71)	(2.54)
# Analysts in Brokerage Firm	0.168*	0.229***	0.456*	0.563**
	(1.92)	(2.61)	(1.80)	(2.20)
Star Analyst	−0.088	−0.150	−0.018	−0.083
	(−0.20)	(−0.35)	(−0.04)	(−0.16)
# Information Events	−0.001	−0.002	−0.003	−0.001
	(−0.41)	(−0.60)	(−0.92)	(−0.38)
Distance to Default	−0.005***	−0.002***	−0.007***	−0.004***
	(−7.19)	(−2.73)	(−7.74)	(−4.08)
Market Cap	−0.012***	−0.007**	−0.016***	−0.008**
	(−4.52)	(−2.50)	(−5.05)	(−2.34)
% Institutional Holdings	−0.007	−0.001	0.002	0.005
	(−0.44)	(−0.11)	(0.24)	(0.87)
Fluidity	0.042	−0.090	0.147*	0.023
	(0.52)	(−1.03)	(1.81)	(0.26)
Text Complexity	0.002	0.003*	0.002	0.003
	(1.24)	(1.67)	(1.01)	(1.55)
Industry Recession	0.010**	0.013***	0.014***	0.016***
	(2.54)	(3.01)	(3.01)	(3.36)
Time Trend	0.002***		0.002***	
	(3.89)		(2.66)	
Year Fixed Effect	No	Yes	No	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes
Analyst Fixed Effect	No	No	Yes	Yes
Observations	352,358	352,358	352,358	352,358
Adjusted R-squared	0.07	0.07	0.10	0.11

5.2. Can man + machine avoid extreme error?

As in many other skilled professions, extreme forecast errors could be calamitous to the reputation of the forecasters and to the welfare of the recipients of investment advice. However, as the common saying “to err is human; to forgive is divine” goes, machine errors are far less tolerated than human mistakes (Prahl and Swol, 2017). We are thus interested in the resilience of Man + Machine against extreme errors,

Table 7
Incremental effect of M+M over both man and machine.
This table reports the synergy of Man + Machine over both Man and Machine. The M+M Square Error Residual is defined as follows: We regress the squared forecast error of M+M on 2 variables: the machine-alone squared forecast error and man-alone squared forecast error. By flipping the sign of residual of the regression, we obtain M+M Square Error Residual. The M+M Forecast Error Residual is similarly defined using absolute forecast errors instead of squared forecast errors. To calculate t-statistics, standard errors are clustered at the firm level. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tailed), respectively.

	M+M squared error residual				M+M forecast error residual			
Amihud Illiquidity	0.375***	0.371***	0.398***	0.398***	0.218***	0.215***	0.248***	0.250***
	(5.11)	(5.22)	(3.96)	(4.24)	(3.34)	(3.70)	(2.80)	(3.29)
Intangible Assets	0.001*	0.001	0.000	−0.000	−0.000	−0.001	0.000	−0.000
	(1.71)	(0.96)	(0.02)	(−0.36)	(−0.16)	(−0.88)	(0.23)	(−0.23)
# Analysts in Brokerage Firm	−0.060***	−0.072***	0.011	−0.004	−0.061***	−0.074***	0.036	0.026
	(−3.69)	(−4.48)	(0.24)	(−0.08)	(−3.28)	(−4.08)	(0.74)	(0.57)
Star Analyst	−0.054	−0.048	0.005	0.009	−0.076	−0.071	−0.036	−0.034
	(−0.86)	(−0.78)	(0.07)	(0.12)	(−1.05)	(−1.00)	(−0.40)	(−0.39)
# Information Events	0.001**	0.001**	0.001**	0.001**	0.002***	0.002**	0.002***	0.002**
	(2.25)	(2.57)	(2.07)	(2.41)	(3.14)	(2.54)	(2.80)	(2.46)
Distance to Default	−0.001***	0.000	−0.001***	0.000	−0.001***	−0.000**	−0.001***	−0.000
	(−7.77)	(0.26)	(−5.90)	(1.42)	(−8.22)	(−2.01)	(−5.90)	(−0.21)
Market Cap	0.006***	0.004***	0.005***	0.004***	0.005***	0.003***	0.005***	0.003***
	(7.81)	(5.55)	(6.48)	(4.65)	(7.83)	(4.79)	(6.78)	(4.19)
% Institutional Holdings	−0.010***	−0.010**	−0.015***	−0.014***	−0.009**	−0.009*	−0.011***	−0.011***
	(−2.75)	(−2.16)	(−6.82)	(−4.77)	(−2.12)	(−1.67)	(−5.31)	(−3.30)
Fluidity	0.010	0.026	0.008	0.020	0.007	0.040**	−0.007	0.025
	(0.67)	(1.52)	(0.50)	(1.01)	(0.46)	(2.24)	(−0.38)	(1.20)
Text Complexity	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.03)	(0.37)	(0.06)	(0.44)	(0.53)	(0.90)	(0.84)	(1.12)
Industry Recession	0.004***	0.003***	0.004***	0.002***	0.004***	0.002***	0.003***	0.001
	(4.68)	(3.23)	(3.95)	(2.59)	(4.43)	(2.60)	(3.16)	(1.54)
Time Trend	−0.001***		−0.000***		−0.001***		−0.001***	
	(−6.32)		(−3.85)		(−6.23)		(−4.47)	
Year Fixed Effect	No	Yes	No	Yes	No	Yes	No	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Analyst Fixed Effect	No	No	Yes	Yes	No	No	Yes	Yes
Observations	346,301	346,301	346,301	346,301	346,301	346,301	346,301	346,301
Adjusted R-squared	0.13	0.14	0.23	0.24	0.08	0.09	0.15	0.16

a quality which would be crucial for the future of the combination, in addition to an superior average forecast accuracy.

To set the stage, we benchmark the forecast error of each forecast to the 90th (or 75th, as a sensitivity check) percentile of squared prediction errors from all analysts on the same firm in the same year. Such a setup leads to four outcomes with regard to who commit(s) an extreme error: (1) both the analyst and the AI model ("Both"); (2) Analyst; (3) AI; and (4) neither commits an extreme error ("Neither"). We examine these four scenarios and compute their empirical frequencies.[23] We then compute the unconditional and conditional probabilities that the Man + Machine model can avoid the extreme error committed in the first three scenarios and, equally importantly, the probability that Man + Machine creates an extreme error in the fourth scenario. All probabilities are reported in Table 8.

We discover that the analyst and the AI are about equally likely to make extreme errors (9.3% and 7.8% using the 90th percentile threshold).[24] There is a further probability of 3.5% that both make lousy forecasts. It turns out that the Man + Machine model can help avoid 90.7% of extreme errors made by human and 43.6% of those by AI. Even when both analysts and AI seem to be out of the ballpark, their combination still manages to bring 4.6% of such cases back to a reasonable range. Furthermore, Man + Machine only creates its own extreme error in 0.1% of the "Neither" scenario. The overall results present a significant complementary benefit of combining human and AI capabilities.

This collaborative model emphasizes a crucial aspect of human–machine synergy, particularly in high-stakes scenarios, encompassing not only financial markets but also potential applications in healthcare diagnostics, climate modeling, and emergency response systems. Given that human and machine errors often stem from different factors, their combined efforts have the potential to subdue the likelihood of severe lapses. To the best of our knowledge, this synergy between Man and Machine has not been empirically documented in the existing literature.

5.3. Impact of man + machine: an event study

In this section, we resort to an event study to sharpen the inference of the impact of integrating man and machine in stock analyses. In recent years, the infrastructure of "big data" has created a new class of information about companies that is collected and published outside of the firms and which can provide unique and timely clues into market demand, profit prospects, and investment opportunities. An important and popular type of such alternative data captures "consumer footprints", oftentimes in the literal sense such as satellite images of retail parking lots. Such data, which have to be processed by machine learning models, have been shown to contain incremental information for earnings and stock prices conditional on corporate disclosure and news coverage (Zhu, 2019; Katona et al., 2024). Chi et al. (2024) show that analysts who use alternative data more frequently have more precise forecasts.

We build on data from (Katona et al., 2024) on the staggered introduction of several important alternative databases, and conduct a difference-in-differences test of analysts' performance versus our AI model before and after the availability of the alternative data on specific firms. The underlying premise is that analysts who cover firms that are served by the alternative data are potentially in the situation of

[23] These four cases are not mutually disjoint, as the "Analyst" (scenario 2) and "AI" (scenario 3) cases both include the "Both" cases (scenario 1). We adopt this convention to evaluate how the Man + Machine model performs in terms of avoiding extreme errors relative to Man/Machine, independent of the counterparty's performance. Untabulated, we also conduct the same analysis for four disjoint scenarios, i.e., "Both", "Analyst Only", "AI Only", and "Neither", and find qualitatively similar results; in fact, the Man + Machine model corrects an even greater fraction of extreme errors committed by analysts alone.

[24] A sensitivity analysis using the 75th percentile yields qualitatively similar results.

Table 8
Probabilities of extreme errors: Man vs. Machine and Man + Machine.
This table presents the probabilities of extreme errors by analysts and the AI and how the Man + Machine model helps to correct such errors. We benchmark the forecast error of each forecast to the 90th (or 75th, as a sensitivity check) percentile of squared prediction errors from all analysts on the same firm in the same year. Such a setup leads to four outcomes with regard to who commit(s) an extreme error: (1) both the analyst and the AI model ("Both"); (2) Analyst; (3) AI; and (4) neither commits an extreme error ("Neither"). We examine these four scenarios and compute their empirical frequencies. We then compute the unconditional and conditional probabilities that the Man + Machine model can avoid the extreme error committed in the first three scenarios, and equally importantly, the probability that Man + Machine creates an extreme error in the fourth scenario. Panel A and B show results for extreme errors defined by the 90th percentile and 75th percentile of forecast errors, respectively.

Panel A: Probabilities of Extreme Errors (90th percentile)				
	Both	Analyst	AI	Neither
Uncond. Prob.	3.46%	9.34%	7.84%	79.36%
	M+M Avoids Both	M+M Avoids Analyst	M+M Avoids AI	M+M Creates EE
Uncond. Prob.	0.16%	8.47%	3.41%	0.11%
	M+M Avoids Both/ Both EE	M+M Avoids Analyst/ Analyst EE	M+M Avoids AI/ AI Only	M+M Creates EE/ Neither EE
Conditional Prob.	4.57%	90.72%	43.56%	0.13%
Panel B: Probabilities of Extreme Errors (75th percentile)				
	Both	Analyst	AI	Neither
Uncond. Prob.	10.60%	16.82%	12.27%	60.32%
	M+M Avoids Both	M+M Avoids Analyst	M+M Avoids AI	M+M Creates EE
Uncond. Prob.	0.47%	14.35%	5.01%	0.16%
	M+M Avoids Both/ Both EE	M+M Avoids Analyst/ Analyst EE	M+M Avoids AI/ AI Only	M+M Creates EE/ Neither EE
Conditional Prob.	4.39%	85.31%	40.86%	0.26%

Man + Machine, as they have the opportunity to use the additional, AI-processed information. We define two variables based on the staggered introduction of alternative data coverage. The first is *Alt Data Covered*, which is one if satellite imaging data are available for the firm at any point in our sample period (based on the list of covered firms and coverage start dates in Table A1 in Katona et al., 2024), and if the firm is in an industry with a retail footprint,[25] and zero otherwise. The second variable is *Post*, which is an indicator variable that is one if satellite data are currently available (based on coverage start dates in Table A1 in Katona et al., 2024), or if the firm is not listed in that table but the date is after 2014,[26] and zero otherwise. In our analysis, a firm is "treated" by the alternative data if it is an $Alt\ Data\ Covered$ firm and the time is $Post$ the coverage. Moreover, we only include an observation if the brokerage house with which the analyst is affiliated is covered by the Burning Glass job posting data any time during the past five years.[27]

Alternative data tend to be large in volume and unstructured. Such data are hard to process with traditional tool kits. Commercial data vendors may preprocess the alternative data; for example, by converting satellite imaging data into car counts for each business location. However, substantial additional analysis is still needed to render such data useful for stock analysis. Whether analysts covering the alternative data "treated" firms could capitalize on the novel information source depends on the AI resources in their workplace. We measure AI resources that analysts have access to by the variable *AI Hiring*, which is the ratio of the number of AI jobs to the total number of job postings using the Burning Glass U.S. job posting data and following the classification algorithm developed in Babina et al. (2024).

By its nature, the satellite data covers a segment of the economy, mostly firms in the business-to-consumer (B2C) sectors. Unlikely earlier research based on the data that focused on whether the satellite data help predicting earnings, this research assesses whether the opportunity to work with AI by some analysts affiliated with AI-capable brokerages (i.e., the Man + Machine in reality) are able to close the gap with, or even outperform, AI models. More specifically, we estimate the following difference-in-differences model,

$$\begin{aligned} Analyst\ Beats\ AI_{i,j,t} = {} & \beta_1 Treat_{i,t} \times AI\ \ Hiring_{j,t} \\ & + \beta_2 AI\ \ Hiring_{j,t} + \beta_3 Alt\ Data\ Covered_i \\ & + \beta_4 Treat + Controls_{i,j,t} + \alpha_l + \alpha_{year} + \epsilon_{i,j,t}. \end{aligned} \tag{4}$$

Here $Treat_{i,t} = Alt\ Data\ Covered_i \times Post_{i,t}$ and α_l represents either firm or analyst fixed effects. Note that *Alt Data Covered* and *Post* are indexed by firm i and date t while *AI Hiring* is indexed by the analyst j (or the brokerage firm associated with the analyst) and date t. Table 9 reports the results. The sample here is smaller than those in Tables 4 and 6 due to the requirement that the *AI Hiring* be observable.

Columns (1) to (3) of Table 9 show that post alternative data, analysts covering affected firms improve their performance relative to the AI model, but only significantly so when interacting with *AI Hiring*.[28] In other words, the improvement of predictive performance post alternative data concentrates in the subset of analysts who are affiliated with brokerage firms with strong AI capabilities. Overall results suggest that augmenting humans with new technologies constitutes a promising direction for the analyst profession.

6. Concluding remarks

In this paper, we built an AI analyst to digest corporate disclosure and other information (qualitative and quantitative) and perform forecast tasks similar to those of stock analysts. Our AI analyst is able to beat the majority of human analysts in stock-return forecasts. In the

[25] We define industries with retail footprints to be those that rely mainly on retail traffic, such as the entertainment, healthcare, personal services, retail, restaurant, and hotel industries. Specifically, these include industries 7, 11, 33, 40, 42, 43, and 46 in the Fama–French 48-industry classification.

[26] Based on anecdotal evidence from news and discussion with industry experts, 2014 is the year most alternative data became widely available.

[27] The reason for this restriction is to ensure that the information about AI hiring is reasonably accurate, as we cannot infer AI hiring in case of missing data.

[28] In these specifications, we do not simultaneously control for firm and analyst fixed effects due to insufficient variation in the pairing during the few years around the event.

Table 9
Man + Machine event study: Alternative Data Coverage.
This table presents the coefficients and *t*-stats of regressing the *Analyst Beats AI* indicator on brokerage *AI Hiring*, *Alt Data Covered*, *Post*, and the interactions among these variables. *Analyst Beats AI* is an indicator variable equal to one if the analyst beats the AI. The AI and alternative data variables *AI Hiring*, *Alt Data Covered*, and *Post* are defined as follows. *AI Hiring* is the ratio of the number of AI jobs to the total number of job postings. *Alt Data Covered* is an indicator variable equal to one if alternative data are available for the firm by the end of the sample, and if the firm is in an industry with retail footprints. *Post* is an indicator variable equal to one if a "treated" firm has been covered by alternative data. For "untreated" firms, *Post* is coded one if the year is after 2014. The control variables are the firm-level, industry-level, and macroeconomic variables presented in Table 1. Standard errors are clustered at the firm level. ***, **, * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tailed), respectively.

Variables	Analyst Beats AI		
Treat × AI Hiring	0.533**	0.503**	0.627***
	(2.16)	(2.04)	(2.71)
Treat: Alt Data Cover × Post	0.002	0.028	0.029
	(0.05)	(0.61)	(0.64)
AI Hiring	0.032	0.069	0.018
	(0.90)	(1.52)	(0.52)
Alt Data Cover	0.029	−0.007	
	(0.69)	(−0.17)	
Year Fixed Effect	Yes	Yes	Yes
Firm Fixed Effect	No	No	Yes
Analyst Fixed Effect	No	Yes	No
Observations	56,697	56,697	56,697
Adjusted R-squared	0.03	0.07	0.14

contest of "Man vs. Machine", we find that the relative advantage of such an AI analyst is stronger when information is more transparent and voluminous. Human analysts remain competitive when critical information requires institutional knowledge (such as the nature of intangible assets and conditions associated with financial distress). Combining AI and the art of human experts produces the highest potential in generating accurate forecasts in settings wherein the two skills are complementary, with the benefit particularly appealing in dramatically avoiding extreme mistakes that would have been committed by either human or machine alone. Synergies between humans and machines documented in this study provide guidance on how humans can leverage their advantages in better adaptation for the future of growing AI prowess.

Although we have constructed our AI analyst with a rather comprehensive set of data inputs and adopted state-of-the-art machine learning techniques, it is inherently impossible for a model to be inclusive of all publicly available data and all advanced learning algorithms. We see ample room for future research in terms of machine capability and, more importantly, integration of human and machine intelligence. The following directions might be particularly promising. First, further studies on how best to exploit the Man + Machine's potential in reducing negative tail outcomes can be crucial for risk management in AI Adoption. Second, better training of machine learning models to understand business cycles and deal with evolving environments would critically expand model capability and applicability. Finally, while AI models become increasingly complex, the interpretability of the models would be important both for model robustness and human–machine collaboration.

CRediT authorship contribution statement

Sean Cao: Writing – review & editing, Methodology, Investigation, Conceptualization. **Wei Jiang:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Junbo Wang:** Writing – review & editing, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Baozhong Yang:** Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data_Submit_2.7z (Original data) (Mendeley Data)

Appendix A. List of variables

See Table A.1.

Table A.1
List of all variables used in AI Algorithms.
All Variables (and the definition/source) used in the machine learning algorithms are provided.

Firm characteristics	Definition and/or source
Gross Profits-to-Assets	Novy-Marx (2013)
Return on Assets	Fama and French (2006) and Chen et al. (2014)
Book-to-Market Equity	Rosenberg et al. (1985)
Debt-to-Market	Bhandari (1988)
Earnings-to-Price	Basu (1983)
Cash Flow-to-Price	Lakonishok et al. (1994)
Payout Yield	Boudoukh et al. (2007)
Five-year Sales Growth Rank	Lakonishok et al. (1994)
Enterprise Multiple	Loughran and Wellman (2011)
Sales-to-Price	Barbee et al. (1996)
Abnormal Corporate Investment	Titman et al. (2004)
Investment-to-Assets	Cooper et al. (2008)
Changes in PPE and Inventory/Assets	Lyandres et al. (2008)
Investment Growth	Xing (2008)
Inventory Changes	Thomas and Zhang (2002)
Operating Accruals	Sloan (1996)
Total Accruals	Richardson et al. (2005)
Net External Finance	Bradshaw et al. (2006)
Return on Net Operating Assets	Soliman (2008)
Profit Margin	Soliman (2008)
Asset Turnover	Soliman (2008)
Operating Profits-to-Equity	Fama and French (2015)
Book Leverage	Fama and French (1992)
Advertising Expense-to-Market	Chan et al. (2001)

(continued on next page)

Table A.1 (*continued*).

Firm characteristics	Definition and/or source
R&D-to-Market	Chan et al. (2001)
Operating Leverage	Novy-Marx (2011)
Financial Constraints	Kaplan–Zingales index, Lamont et al. (2001)
Asset Liquidity	Scaled by book assets, Ortiz-Molina and Phillips (2014)
Asset Liquidity	Scaled by market assets, Ortiz-Molina and Phillips (2014)
I/B/E/S Actual Earning	I/B/E/S actual earning 4 quarter before scaled by adjusted price
Number of Institutional Owners	Number of 13F institutional investors that own the stock
Ownership Concentration	Herfindahl–Hirschman Index
Total Institutional Ownership	Percent of shares outstanding owned by 13F investors
Industry Variables	Definition and/or Source
Competition Measure from 10-K	Li et al. (2013)
Fluidity	Product market Fluidity, Hoberg et al. (2014)
Industry Dummies	Dummy variables that indicate Fama–French 12 industries
Industry Size	Industry Size within past 3, 6, 9, 12, 24 and 36 months
Industry Earning	Industry earning within past 3, 6, 9, 12, 24 and 36 months
Macro Variables	Definition and/or Source
IP	Industrial Production Index
CPI	Consumer Price Index
Oil price	Crude Oil Price
Tbill3	3-month Treasury Bill
TBond10	10-Year Treasury Constant Maturity Rate
Credit Spread	Baa-AAA yield spread
Dividend Yield	Fama and French (1989)
Stock Market Illiquidity	Chen et al. (2018)
New Orders	Jones and Tuzel (2013)
Technical Indicators	Neely et al. (2014)
Average Correlation	Average correlation of largest stocks in the SP500 index Pollet and Wilson (2010)
Filing and SEC Variables	Definition and/or Source
Neg 10KQ	Percentage of negative words from 10K/10Q
NegPos 10KQ	Percentage of negative minus positive words from 10K/10Q
Neg 8k	Percentage of negative words from 8K
NegPos 8K	Percentage of negative minus positive from 8K
Neg Other	Percentage of negative words from other reports
NegPos Other	Percentage of negative minus positive from other reports
ML-based Sentiment	ML-based negative tones minus ML-based positive tones scaled by the length of SEC filings, Cao et al. (2021b)
ML-based Neg Sentiment	ML-based negative tones scaled by the length of SEC filings
File size	Size of the filings
GF Index	Gunning Fog Readability Index
Itemized 8K	Average percentage of negative minus positive word of items 1-5, 7, 8, and 12 in 8K over past three months
Ravenpack Variables	Definition and/or Source
ESS	Event Sentiment Score (Firm, US, World)
AES	Aggregate Event Sentiment (Firm, US, World)
AEV	Aggregate Event Volume (Firm, US, World)
CSS	Composite Sentiment Score (Firm)
PEQ	Global Equities Sentiment Score (Firm)
BEE	Earning Evaluations Sentiment Score (Firm)
BMQ	Editorials & Commentary Sentiment Score (Firm)
BAM	Venture Company Merge and Acquisitions Sentiment Score (Firm)
BCA	Report on Corporate Actions Sentiment Score (Firm)
BER	Earnings Releases Sentiment Score (Firm)
Patent Variables	Definition and/or Source
Xi_real	Value of innovation deflated to 1982 (million) dollars from Kogan et al. (2017)
Twitter Variables	Definition and/or Source
Sentiment and Uncertainty	Derived from peer disclosure tweets following (Cao et al., 2021a)
Sentiments GI	Sentiment: Harvard General Inquirer IV-4 dictionary
Sentiment HE	Sentiment: (Henry, 2008) financial dictionary
Sentiment LM	Sentiment: (Loughran and McDonald, 2011) dictionary
Sentiment QDAP	Sentiment: (Hu and Bin, 2004) QDAP dictionary
Ratio Uncertainty LM	Uncertainty words ratio: (Loughran and McDonald, 2011) dictionary

Appendix B. Details of the machine learning models

In this section, we briefly describe the basic structure and strengths of machine learning models considered in our paper. Interested readers are referred to representative references for more details, such as Goodfellow et al. (2016) and James et al. (2023).

B.1. Decision-tree based models

The linear models considered above may not work well if there are nonlinear relationships among the predictive variables. In this section, we discuss a class of versatile nonlinear models – decision trees and derived models.

B.1.1. Decision trees

Decision trees are modeled after human decisions. A decision tree is a series of binary decisions based on cutoffs of independent variables at each branching point. The tree thus will divide the rectangular feature space into smaller rectangular blocks. The decision tree regression then use the sample mean of the dependent variable in each block as the prediction for any point in the block.

Decision trees have the benefit of being easily interpretable because it is modeled after human decisions (similar to a step-by-step instructions) and can also be displayed graphically (as binary trees). Trees are also a flexible non-linear model that can model a variety of nonlinear patterns given the large degree of freedom in specifying the sequences of branching rules.

However, trees do not have a high level of accuracy by themselves because of the restrictive form of the binary branching process, which forces the sample to be split into rectangular regions and may not approximate the real underlying patterns (whether linear or nonlinear) well. Trees are also non-robust. In addition, a small change in the data can lead to large changes in the structure of the estimated tree because the tree structure is discrete, not continuous. Several methods, including random forest and gradient boosting, use trees as basic building blocks to form ensemble predictors and achieve superior performance.

B.1.2. Random forest

A random forest (introduced by Breiman, 2001) proceeds in the following way. First, it involves drawing a bootstrapped sample (drawing with repetition) from the original sample. Second, on the bootstrapped sample, one builds a decision tree, selecting a splitting predictor among only a random m features of the total p predictors. Third, one repeats the above two steps to build a number of decision trees, and form the ensemble predictor by taking the mean predictor of all the trees.

Random forests perform better than simple trees for several reasons. First, through aggregating predictions over bootstrapped samples, it reduces the variance and non-robustness of single trees. Second, the random feature selection in the second step above ensures that the estimated trees are not too correlated, avoiding relying only on a few prominent features and further reducing the variance of the model.

B.1.3. Gradient boost

Boosting also combines a number of weak models to generate a stronger model. In boosting of trees, a number of trees are constructed sequentially, i.e., each tree is constructed using information based on the previously constructed trees. In gradient boosting, each decision tree is fit to the residuals of the model, not to the outcome. Once a new tree is obtained, it is added to the predictive function to update it, usually with a learning weight multiplied to the tree predictor to adjust the rate of learning new information. Then new residuals are obtained from the updated predictive function and the process is repeated for a number of times to obtain the final ensemble predictor. Because boosting models aggregate results of decision trees sequentially, each component tree does not need to be very precise and can be simple, i.e., having a low depth.

In a sense, gradient boosting is similar to the Newton's gradient algorithm in optimization. It approximates the true underlying function sequentially by improving on the predicted residuals/errors gradually. This allows the final predictive function to have a much richer and more flexible structure and thus much better performance than single decision trees. It also reduces the non-robustness of single trees through using an ensemble of trees. For these reasons, gradient boosting is one of the best off-the-shelf machine learning methods.

B.2. Deep learning model: Long short-term memory neural networks

The neural networks models, initial motivated by the neuron structures in the brains of humans and animals, blossomed after breakthroughs in algorithms and computing power (LeCun et al., 2015). Neural networks models, also called deep learning models, have become some of the most powerful models and achieved near- or super-human capabilities in a wide variety of applications, such as natural language processing, speech recognition, computer vision, game playing, and autonomous driving.

There are many different architectures of neural networks, such as the simplest Feedforward Neural Networks for straightforward classification tasks, the Convolutional Neural Networks for image and pattern recognition, and Recurrent Neural Networks (RNN) that can process sequential data such as speech and text. Long Short-Term Memory (LSTM) Neural Networks are a special type of RNN that is the key to the many successes of RNN, including speech recognition, language modeling, and translation.

In a neural network, there are nodes (neurons) that are connected to each other. There are three types of nodes: input nodes that are used to receive data; output nodes that produce desired outcomes or predictions; and intermediate nodes that process the data from input nodes and convert them to outputs. The connections of the nodes determine the structure of the neural network and its features. RNNs are neural networks with loops, or nodes that are connected to themselves.

LSTM networks are introduced by Hochreiter and Schmidhuber (1997) to solve the problem that standard RNNs have trouble retaining "memory" of the much earlier parts of sequential input data, when processing the later parts of the data. Since sequential data may have long-term dependencies, i.e., parts far away in the sequence may be related, it is important to have "long-term memory" to handle them. LSTM networks have a sequence of nodes that are specifically designed to retain long-term information and update it continuously with new information in a flexible way. As a result, LSTM can capture both short-term and long-term relations in sequential or time-series data very well, such as momentum and reversal patterns, supporting its applications in financial economics, given the abundance of time-series financial data.

B.3. Implementation of models

Standalone machine model. In the training sample for year s, when making predictions in year $s-2$ or earlier, the target return spans the cumulative return from the onset of the current month until the end of 12 months from the present date. When making predictions in year $s-1$ in the training sample, the target return consists of the cumulative return from the start of the current month until the end of the year. Consequently, the training data for each year only includes data prior to that specific year, which prevents any look-ahead bias. We apply the same principle in constructing training samples for earnings predictions.

We construct our AI model as the ensemble (using the median value) of three well-established machine learning models: random forest, extreme gradient boosting, and LSTM models. The data fed to the random forest and extreme gradient boosting models do not contain the time-series history and consist of predictors in the current quarter. These two models allow us to extract the most important predictors and their nonlinear interactions from a large number of characteristics that have the potential to predict stock returns. On the other hand, the long short-term memory neural network model (LSTM) incorporates the time series of all predicting variables over eight past quarters. It contains five layers, with the first layer being a standard LSTM model and the remaining layers a dense feed-forward neural network model. This model ensures that time-series patterns that can predict future stock returns, such as seasonality and momentum, are not ignored. It also offers insight into economic conditions such as recessions or expansions. While the eight window could potentially extend further, it's worth noting that a commonly accepted definition of a recession involves two consecutive quarters of real GDP decline, a pattern that can be captured by the model. The final AI model we used to make predictions is an ensemble of the three trained models above. Specifically, we take the median value of the predictions from the three models.

We apply a rolling window approach to train the model. The typical training window spans three years. However, in instances where the preceding three years encompass a "distress" year (e.g., 2001, 2008, and 2009), we extend the training window to the first year after the preceding distress period or commence from the dataset's starting year. For example, for the years 2002 to 2004, the training window encompasses data from the sample's inception. For the years 2009 to 2012, the training window starts from 2002, reflecting the preceding recession in 2001.

The hyperparameters used by our machine learning methods are reported in Table IA.9 in the Internet Appendix. We also conduct robustness tests for a wide range of hyperparameters (results are reported in the same table). The Human vs. AI beat ratios are quantitatively similar under these scenarios.

In our main results, the predictors are not normalized. However, we also explore robustness by normalizing the variables. Specifically, at each analyst announcement date, we normalize all independent variables (excluding returns and prices) utilized in the ML model training process. This normalization involves transforming each variable into a uniformly distributed variable at the unit percentile level. Therefore, all the independent variables fall in the range between zero and one. The Human vs. AI beat ratios remain consistent regardless of normalization.

Man + machine (M+M) model. To construct the M+M model, we proceed in two steps. First, we construct a machine learning model utilizing analyst forecasts and analyst characteristics as inputs, in addition to all predictors in the standalone machine. Specifically, we include as predictors the mean forecast and forecast accuracy (mean square error) of each analyst for the past five years, and the mean consensus and forecast accuracy (mean square error) of all analysts in the previous 90 days. With this comprehensive set of information, we train the model and make the prediction. Second, we ensemble this prediction with three others: the current analyst prediction, the Machine–debiased Man (MDM) prediction, and the standalone machine prediction, to form the M+M prediction (as before, the ensemble value is the median value of these four predictions).

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jfineco.2024.103910.

References

Abarbanell, Jeffrey S., 1991. Do analysts' earnings forecasts incorporate in-formation in prior stock price changes? J. Account. Econ. 14, 147–165.

Abarbanell, Jeffery, Lehavy, Reuven, 2003. Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts earnings forecasts. J. Account. Econ. 36, 105–146.

Abis, Simona, 2022. Man vs. machine: quantitative and discretionary equity management. Working paper.

Abis, Simona, Veldkamp, Laura, 2024. The changing economics of knowledge production. Rev. Financ. Stud. 37, 89–118.

Acemoglu, Daron, Autor, David, Hazell, Jonathon, Restrepo, Pascual, 2022. Artificial intelligence and jobs: Evidence from online vacancies. J. Labor Econ. 40, 293–340.

Acemoglu, Daron, Restrepo, Pascual, 2018. The race between man and machine: Implications of technology for growth, factor shares and employment. Amer. Econ. Rev. 108, 1488–1542.

Acemoglu, Daron, Restrepo, Pascual, 2019. Automation and new tasks: How technology displaces and reinstates labor. J. Econ. Perspect. 33, 3–30.

Aghion, Philippe, Jones, Benjamin F., Jones, Charles I., 2019. Artificial intelligence and economic growth, The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, Chicago.

Agrawal, Ajay, Gans, Joshua, Goldfarb, Avi, 2019b. Artificial intelligence: The ambiguous labor market impact of automating prediction. J. Econ. Perspect. 33, 31–50.

Agrawal, Ajay Joshua Gans, Goldfarb, Avi, 2018. Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Press, Boston.

Amihud, Yakov, 2002. Illiquidity and stock returns: Cross-section and time series effects. J. Financial Mark. 5, 31–56.

Armour, John, Parnham, Richard, Sato, Mari, 2022. Augmented Lawyering. University of Illinois Law Review.

Asquith, Paul, Mikhail, Michael B., Au, Andrea S., 2005. Information content of equity analyst reports. J. Financ. Econ. 75, 245–282.

Aubry, Mathieu, Kraeussl, Roman, Manso, Gustavo, Spaenjers, Christophe, 2023. Biased auctioneers. J. Finance 78, 795–833.

Autor, David H., Levy, Frank, Murnane, Richard J., 2003. The skill content of recent technological change: An empirical exploration. Q. J. Econ. 118, 1279–1333.

Babina, Tania, Fedyk, Anastassia, He, Alex Xi, Hodson, James, 2024. Artificial intelligence, firm growth, and industry concentration. J. Financ. Econ. forthcoming.

Barbee, William C., Mukherji, Sandip, Raines, Gary A., 1996. Do sales–price and debt–equity explain stock returns better than book–market and firm size? Financ. Anal. J. 52, 56–60.

Basu, Sanjoy, 1983. The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. J. Financ. Econ. 12, 129–156.

Bhandari, Laxmi Chand, 1988. Debt/equity ratio and expected common stock returns: Empirical evidence. J. Finance 43, 507–528.

Boudoukh, Jacob, Michaely, Roni, Richardson, Matthew, Roberts, Michael, 2007. On the importance of measuring payout yield: Implications for empirical asset pricing. J. Finance 62, 877–915.

Bradshaw, Mark T., Richardson, Scott A., Sloan, Richard G., 2006. The relation between corporate financing activities, analysts' forecasts and stock returns. J. Account. Econ. 42, 53–85.

Brav, Alon, Lehavy, Reuven, 2003. An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. J. Finance 58, 1933–1967.

Breiman, Leo, 2001. Random forests. Mach. Learn. 45, 5–32.

Brogaard, Jonathan, Ringgenberg, Matthew C., Rösch, Dominik, 2024. Does floor trading matter? J. Finance forthcoming.

Brogaard, Jonathan, Zareei, Abalfazl, 2022. Machine learning and the stock market. J. Financ. Quant. Anal. 58, 1431–1472.

Brynjolfsson, Erik, Mitchell, Tom, Rock, Daniel, 2018. What can machines learn, and what does it mean for occupations and the economy? AEA Pap. Proc. 108, 43–47.

Cao, Sean, Cong, Lin William, Yang, Baozhong, 2023a. Financial reporting and blockchains: audit pricing, misstatements, and regulation. Working paper.

Cao, Sean, Fang, Vivian, Lei, Lijun, 2021a. Negative peer disclosure. J. Financ. Econ. 140, 815–837.

Cao, Sean, Jiang, Wei, Yang, Baozhong, Zhang, Alan L., 2023b. How to talk when a machine is listening: corporate disclosure in the age of AI. Rev. Financial Stud. 36, 3603–3642.

Cao, Sean, Kim, Yongtae, Wang, Angie, Xiao, Houping, 2021b. Power of deep learning: quantifying language to explain cross-sectional returns. Working paper.

Cao, Kai, You, Haifeng, 2024. Fundamental analysis via machine learning. Financial Anal. J. 80, 74–98.

Chan, Louis K.C., Lakonishok, Josef, Sougiannis, Theodore, 2001. The stock market valuation of research and development expenditures. J. Finance 56, 2431–2456.

Chen, Yong, Eaton, Gregory W., Paye, Bradley S., 2018. Micro(structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity. J. Financ. Econ. 130, 48–73.

Chen, Qi, Francis, Jennifer, Jiang, Wei, 2005. Investor learning about analyst predictive ability. J. Account. Econ. 39, 3–24.

Chen, Long, Novy-Marx, Robert, Zhang, Lu, 2014. An alternative three-factor model. working paper.

Chen, Luyang, Pelger, Markus, Zhu, Jason, 2023. Deep learning in asset pricing. Manage. Sci. 70, 714–750.

Chen, Nai-Fu, Roll, Richard, Ross, Stephen, 1986. Economic forces and the stock market. J. Bus. 59, 383–403.

Chen, Mark A., Wu, Qinxi, Yang, Baozhong, 2019. How valuable is FinTech innovation? Rev. Financ. Stud. 32, 2062–2106.

Chi, Feng, Hwang, Byoung-Hyoun, Zheng, Yaping, 2024. The use and usefulness of big data in finance: Evidence from financial analysts. Manage. Sci. forthcoming.

Coleman, Braiden, Merkley, Kenneth J., Pacelli, Joseph, 2022. Human versus machine: A comparison of robo-analyst and traditional research analyst investment recommendations. Account. Rev. 97, 221–244.

Cong, Lin William, Liang, Tengyuan, Yang, Baozhong, Zhang, Xiao, 2020. Analyzing textual information at scale. In: Balachandran, Kashi R. (Ed.), Information to Facilitate Efficient Decision Making: Big Data, Blockchain and Relevance. World Scientific Publishers, Singapore.

Cong, Lin William, Tang, Ke, Wang, Jingyuan, Zhang, Yang, 2022. Alphaportfolio: direct construction through reinforcement learning and interpretable AI. Working paper.

Cooper, Michael J., Gulen, Huseyin, Schill, Michael J., 2008. Asset growth and the cross-section of stock returns. J. Finance 63, 1609–1651.

Crane, Alan, Crotty, Kevin, 2020. How skilled are security analysts. J. Finance 75, 1629–1675.

D'Acunto, Francesco Nagpurnanand Prabhala, Ross, Alberto G., 2019. The promises and pitfalls of robo-advising. Rev. Financ. Stud. 32, 1983–2020.

De Bondt, Werner, Thaler, Richard, 1990. Do security analysts overreact? Amer. Econ. Rev. 80, 52–57.
de Silva, Tim, Thesmar, David, 2024. Noise in expectations: Evidence from analyst forecasts. Rev. Financ. Stud. forthcoming.
Dechow, Patricia M., Sloan, Richard G., Sweeney, Amy P., 1995. Detecting earnings management. Account. Rev. 70, 193–225.
Desai, Hemang, Liang, Bing, Singh, Ajai K., 2019. Do all-stars shine? Evaluation of analyst recommendations. Financ. Anal. J. 56, 20–29.
Doyle, Jeffrey T., Jennings, Jared N., Soliman, Mark T., 2013. Do managers define non-GAAP earnings to meet or beat analyst forecasts? J. Account. Econ. 56, 40–56.
Eisfeldt, Andrea L., Papanikolaou, Dimitris, 2013. Organization capital and the cross-section of expected returns. J. Finance 68, 1365–1406.
Eisfeldt, Andrea L., Papanikolaou, Dimitris, 2014. The value and ownership of intangible capital. Amer. Econ. Rev. 104, 189–194.
Elton, Edwin J., Gruber, Martin J., Blake, Christopher R., 1996. The the persistence of risk-adjusted mutual fund performance. J. Bus. 69, 133–157.
Ewens, Michael, Peters, Ryan H., Wang, Sean, 2024. Measuring intangible capital with market prices. Manage. Sci. forthcoming.
Falato, Antonio, Kadyrzhanova, Dalida, Sim, Jae, 2022. Rising intangible capital, shrinking debt capacity, and the US corporate savings glut. J. Finance 77, 2799–2852.
Fama, Eugene F., French, Kenneth R., 1989. Business conditions and expected returns on stocks and bonds. J. Financ. Econ. 25, 23–49.
Fama, Eugene F., French, Kenneth R., 1992. The cross-section of expected stock returns. J. Finance 47, 427–465.
Fama, Eugene F., French, Kenneth R., 2006. Profitability, investment and average returns. J. Financ. Econ. 82, 491–518.
Fama, Eugene F., French, Kenneth R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116, 1–22.
Goldstein, Itay, Spatt, Chester S., Ye, Mao, 2021. Big data in finance. Rev. Financ. Stud. 34, 3213–3225.
Goodfellow, Ian, Bengio, Yoshua, Courville, Aaron, 2016. Deep Learning. MIT Press, Cambridge, MA.
Goyal, Amit, Welch, Ivo, Zafirov, Athanasse, 2024. A comprehensive 2021 look at the empirical performance of equity premium prediction II. Rev. Financ. Stud. forthcoming.
Grennan, Jillian, Michaely, Roni, 2020. Artificial intelligence and high-skilled work: evidence from analysts. Working paper.
Grennan, Jillian, Michaely, Roni, 2021. FinTechs and the market for financial analysis. J. Financ. Quant. Anal. 56, 1877–1907.
Gu, Shihao, Kelly, Bryan, Xiu, Dacheng, 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33, 2223–2273.
Hanley, Kathleen Weiss, Hoberg, Gerard, 2019. Dynamic interpretation of emerging risks in the financial sector. Rev. Financ. Stud. 32, 4543–4603.
Henry, Elain, 2008. Are investors influenced by how earnings press releases are written? J. Bus. Commun. 45, 363–407.
Hirshleifer, David, Yaron, Levi, Ben, Lourie, Siew, Hong Teoh, 2019. Decision fatigue and heuristic analyst forecasts. J. Financ. Econ. 133, 83–98.
Hoberg, Gerard, Phillips, Gordon, Prabhala, Nagpurnanand, 2014. Product market threats, payouts, and financial flexibility. J. Finance 69, 293–324.
Hochreiter, Sepp, Schmidhuber, Jürgen, 1997. Long short-term memory. Neural Comput. 9, 1735–1780.
Hu, Mingqiu, Bin, Liu, 2004. Mining opinion features in customer reviews. In: AAAI04 Proceedings of the 19th National Conference on Artificial Intelligence, vol. 4, pp. 755–760.
James, Gareth, Witten, Daniela, Hastie, Trevor, Tibshirani, Robert, Taylor, Jonathan, 2023. Introduction to Statistical Learning: with Applications in Python. Springer, New York.
Jansen, Mark, Nguyen, Hieu, Sham, Amin, 2023. Rise of the machines: The impact of automated underwriting. Manage. Sci. forthcoming.
Jegadeesh, Narasimhan, Kim, Joonghyuk, Krische, Susan D., Lee, Charles M.C., 2004. Analyzing the analysts: When do recommendations add value? J. Finance 59, 1083–1124.
Jiang, Wei, Tang, Yuehua, Xiao, Rachael Jiqiu, Yao, Vincent, 2023. Surviving the fintech disruption. NBER Working paper 28668.
Jones, Christopher S., Tuzel, Selale, 2013. Inventory investment and the cost of capital. J. Financial Econ. 107, 557–579.
Katona, Zsolt, Painter, Marcus, Patatoukas, Panos N., Zeng, Jean, 2024. On the capital market consequences of alternative data: Evidence from outer space. J. Financ. Quant. Anal. forthcoming.
Kogan, Leonid, Papanikolaou, Dimitris, Seru, Amit, Stoffman, Noah, 2017. Technological innovation, resource allocation, and growth. Q. J. Econ. 132, 665–712.
Lakonishok, Josef, Shleifer, Andrei, Vishny, Robert W., 1994. Contrarian investment, extrapolation, and risk. J. Finance 49, 1541–1578.
Lamont, Owen, Polk, Christopher, Saá-Requejo, Jesús, 2001. Financial constraints and stock returns. Rev. Financ. Stud. 14, 529–554.
LeCun, Yann, Bengio, Yoshua, Hinton, Geoffrey, 2015. Deep learning. Nature 521, 436–444.
Li, Xi, 2005. The persistence of relative performance in stock recommendations of sell-side financial analysts. J. Account. Econ. 40, 129–152.
Li, Feng, Lundholm, Russell, Minnis, Michael, 2013. A measure of competition based on 10-K filings. J. Account. Res. 51, 399–436.
Lim, Terence, 2002. Rationality and analyst forecast bias. J. Finance 56, 369–385.
Liu, Miao, 2022. Assessing human information processing in lending decisions: A machine learning approach. J. Account. Res. 60, 607–651.
Loughran, Tim, McDonald, Bill, 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J. Finance 66, 35–65.
Loughran, Tim, Wellman, Jay W., 2011. New evidence on the relation between the enterprise multiple and average stock returns. J. Financ. Quant. Anal. 46, 1629–1650.
Lyandres, Evgeny, Sun, Le, Zhang, Lu, 2008. The new issues puzzle: Testing the investment-based explanation. Rev. Financ. Stud. 21, 2825–2855.
Martin, Ian W.R., Nagel, Stefan, 2022. Market efficiency in the age of big data. J. Financ. Econ. 145, 154–177.
Merton, Robert C., 1974. On the pricing of corporate debt: The risk structure of interest rates. J. Finance 29, 449–470.
Michaely, Roni, Womack, Kent L., 1999. Conflict of interest and the credibility of underwriter analyst recommendations. Rev. Financ. Stud. 12, 653–686.
Mikhail, Michael, Walther, Beverly, Willis, Richard, 2007. When security analysts talk, who listens? Account. Rev. 82, 1227–1253.
Muro, Mark, Maxim, Robert, Whiton, Jacob, 2019. Automation and Artificial Intelligence: How Machines are Affecting People and Places. Brookings.
Narasimhan, Jegadeesh, Noh, Joonki, Pukthuanthong, Kuntara, Roll, Richard, Wang, Junbo, 2019. Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. J. Financ. Econ. 133, 273–298.
Neely, Christopher, Rapach, David E., Tu, Jun, Zhou, Guofu, 2014. Forecasting the equity risk premium: The role of technical indicators. Manage. Sci. 60, 1772–1791.
Novy-Marx, Robert, 2011. Operating leverage. Rev. Finance 15, 103–134.
Novy-Marx, Robert, 2013. The other side of value: The gross profitability premium. J. Financ. Econ. 108, 1–28.
Ortiz-Molina, Hernan, Phillips, Gordon, 2014. Real asset illiquidity and the cost of capital. J. Financ. Quant. Anal. 49, 1–32.
Pagliaro, Cynthia, Ramadorai, Tarun, Rossi, Alberto G., Utkus, Stephen, Walther, Ansgar, 2023. Algorithm aversion: theory and evidence from robo-advice. Working Paper.
Peters, Ryan H., Taylor, Lucian A., 2017. Intangible capital and the investment-q relation. J. Financ. Econ. 123, 251–272.
Pollet, Joshua M., Wilson, Mungo, 2010. Average correlation and stock market returns. J. Financ. Econ. 96, 364–380.
Prahl, Andrew, Swol, Lyn Van, 2017. Understanding algorithm aversion: when is advice from automation discounted? J. Forecast. 36, 691–702.
Ray, Debraj, Mookherjee, Dilip, 2022. Growth, automation and the long run share of labor. Rev. Econ. Dyn. 46, 1–26.
Richardson, Scott A., Sloan, Richard G., Soliman, Mark T., Tuna, Irem, 2005. Accrual reliability, earnings persistence and stock prices. J. Account. Econ. 39, 437–485.
Rosenberg, Barr, Reid, Kenneth, Lanstein, Ronald, 1985. Persuasive evidence of market inefficiency. J. Portf. Manag. 11, 9–16.
Rossi, Alberto G., Utkus, Stephen P., 2021. Who benefits from robo-advising? evidence from machine learning. Working paper.
Sloan, Richard G., 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? Account. Rev. 71, 289–315.
Soliman, Mark T., 2008. The use of DuPont analysis by market participants. Account. Rev. 83, 823–853.
Stickel, Scott, 1990. Predicting individual analyst earnings forecasts. J. Account. Res. 28, 409–417.
Stickel, Scott, 1992. Reputation and performance among security analysts. J. Finance 47, 1811–1836.
Thomas, Jacob K., Zhang, Huai, 2002. Inventory changes and future returns. Rev. Account. Stud. 7, 163–187.
Titman, Sheridan, Wei, K.C. John, Xie, Feixue, 2004. Capital investments and stock returns. J. Financial Quant. Anal. 39, 677–700.
van Binsbergen, Jules H., Han, Xiao, Lopez-Lira, Alejandro, 2023. Man vs. machine learning: The term structure of earnings expectations and conditional biases. Rev. Financ. Stud. 36, 2361–2396.
Wayne, Ferson, Siegel, Andrew F., Wang, Junbo L., 2024. Factor model comparisons with conditioning information. J. Financ. Quant. Anal. forthcoming.
Webb, Michael, 2020. The impact of artificial intelligence on the labor market. Working paper.
Xing, Yuhang, 2008. Interpreting the value effect through the Q-theory: An empirical investigation. Rev. Financ. Stud. 21, 1767–1795.
Zheng, Xiang, 2022. How can innovation screening be improved? A machine learning analysis with economic consequences for firm performance. Working paper.
Zhu, Christina, 2019. Big data as a governance mechanism. Rev. Financ. Stud. 32, 2021–2061.