Machine Readership and Financial Reporting Decisions[*]

Sean Cao[†], Ying Liang[‡], and Jason (Youngseok) Moon[§]

This Draft: November 19, 2023

Abstract

Machine learning and AI technologies can identify data patterns related to financial misreporting that traditional methods cannot detect. With rising machine readership of corporate financial statements, managers may have less incentive to engage in financial misreporting. This study empirically investigates this possibility and finds a reduction in financial misreporting when machine readership is higher. These results hold after addressing potential identification issues. The impact of machine readership is more pronounced in cases where machine learning offers greater advantages, such as complex financial statements and the availability of alternative data. Notably, for misreporting patterns detectable by traditional linear models, machine readership offers no incremental disciplining, indicating that the strength of machines, instead, lies in recognizing non-linear and high-dimensional patterns. Furthermore, we observe an overall decrease in misstatements, suggesting that machine readership enhances overall financial reporting quality rather than prompting managers to shift misreporting to areas beyond machine detection. This paper highlights the disciplining effect of machine adoption in the capital market on financial reporting.

Keywords: Artificial Intelligence, Financial Reporting, Machine Learning, Earnings Manipulation.
JEL Classification: G34, M41, M48

[*] We thank Blair Marquardt (discussant), Christina Zhu (discussant), Alyssa Moore (discussant) and conference and workshop participants at Georgia State University, Haskayne and Fox Accounting Conference 2022, International Corporate Governance Society 8th Annual Meeting, FARS Midyear Meeting 2023, Chinese University of Hong Kong, Lingnan University, University of Maryland, Erasmus University, Johannes Gutenberg University of Mainz, University of Hong Kong, University of Turku, Hanken School of Economics, AAAI'23 Bridge on AI for Financial Services, and AAA Annual meeting 2023.

[†] Sean Cao is at Robert H. Smith School of Business, University of Maryland, College Park. Email: scao824@umd.edu.

[‡] Ying Liang is at J. Mack Robinson College of Business, Georgia State University. Address: 35 Broad St NW, Atlanta, GA 30303. Email: yliang15@gsu.edu.

[§] Jason (Youngseok) Moon is at J. Mack Robinson College of Business, Georgia State University. Address: 35 Broad St NW, Atlanta, GA 30303. Email: ymoon6@gsu.edu.

1 Introduction

Recent years have witnessed a significant increase in the adoption of artificial intelligence (AI) and machine learning technologies in the capital market. As highlighted by the 2018 Barclay Hedge Fund Manager/Investor Survey, over half of hedge funds utilize machine learning to analyze vast amounts of data to generate trading insights. Over time, financial analysts, auditors, and even regulators have likewise integrated AI and machine learning technologies into their work. One notable example is the Division of Economic and Risk Analysis of the Securities and Exchange Commission (SEC), which has leveraged machine learning algorithms to predict misconduct among investment advisers utilizing data from regulatory filings (Bauguess, 2017). In sum, we find ourselves in an era where corporate financial statements are predominantly read and analyzed by machines.

The increasing reliance on machines for processing financial statement data represents a significant shift in financial information analysis. Therefore, it is essential for researchers to gain a deeper understanding of its implications. While recent studies have started to explore the adoption of novel technologies by corporations (Charoenwong, Kowaleski, Kwan, and Sutherland, 2022; Azarmsa, Liu, and Noh, 2022), investors (Cao, Jiang, Yang, and Zhang, 2023; Abis and Veldkamp, 2022), and auditors (Ham, Hann, Rabier, and Wang, 2022; Law and Shen, 2020), our understanding of how machine readership affects the managers' financial reporting decision remains limited.Our study aims to bridge this gap and investigate whether the rising machine readership of financial statements influences the quality of financial reporting.[1]

Ex ante, it is not certain that the rise of machine readers will lead to enhanced quality in financial reporting. For instance, these readers may incentivize managers to engage in opportunistic behaviors to cater to machine trading algorithms. Cao, Jiang, Yang, and Zhang (2023) reveals that machine readership can trigger distortions in qualitative disclosure, leading managers to "positify" language by avoiding words perceived negative by machines. However, we argue that for financial statements, which are easier to verify than language styles, machine readership could reduce the managerial discretion in reporting. This could lead to outcomes differ from the catering effect documented by Cao, Jiang, Yang, and Zhang (2023).

[1] The compositions of machine readers differ from algorithmic traders. Several studies sindicate that hedge funds and asset managers are the primary users of machine downloads of EDGAR filings for their trading decisions (Crane, Crotty, and Umar, 2022; Cao, Du, Yang, and Zhang, 2021). Cao, Jiang, Yang, and Zhang (2023) identities of the top 20 machine downloaders and their types, with half of the top ten being notable quantitative hedge funds: Renaissance Technologies, Two Sigma, Point 72, Citadel, and D.E. Shaw (See Appendix D).

Machine readership has two main advantages that allow it to better identify irregularities compared to traditional readership: the ability to process high-dimensional data and the flexibility to formulate intricate nonlinear model structures. To better understand these advantages, consider a model with 20 input variables. While this number may not strike readers as significant, when exploring potential non-linear relationships, one would need to include a substantial 190 interaction terms. Taking it a step further and incorporating triple interactions results in a staggering increase of additional more terms. Indeed, Yan and Zheng (2017) use 240 accounting variables from financial statements to construct over 18,000 fundamental signals for identifying significant predictors of cross-sectional stock returns. Traditional models already struggle to handle such a volume of variables, not to mention considering the interaction among 18,000 signals. Machines, on the other hand, inherently possess the capability to discern significant combinations within this intricate landscape. Therefore, it is reasonable to posit that machine learning algorithms can detect abnormal data patterns beyond the scope of traditional methods. As a result, we suspect that managers will respond to this enhanced detection capability and alter their misreporting decisions.

From the managers' perspective, prior to the rise of machine readership, concealing misreporting and evading detection by traditional models was feasible, as these models mainly relied on linear or logistic regression techniques to identify signs of financial misreporting (Dechow, Ge, Larson, and Sloan, 2011; Kothari, Leone, and Wasley, 2005). However, with the presence of machine readership, previously overlooked misreporting becomes considerably more visible to machines. So, what types of suspicious transactions can be detected by machines but often elude traditional models? For instance, when a manager aims to boost earnings and reports inflated sales, they can achieve this as long as they adjust other input variables in traditional models to maintain predicted values within the acceptable range. In contrast, machines take into consideration a wide range of potential interactions and can uncover irregularities, such as unusual sales-to-employee-growth ratio or sales-to-leverage ratio, leading to a red flag for the firm (See Section 2 for a concrete example). Therefore, the emergence of machine readership has significantly shifted the equilibrium in managerial misreporting decision-making, compelling managers to reduce the incidence of misreporting. Consequently, we anticipate that machine readership plays a pivotal role in guiding and disciplining managers' decisions regarding financial reporting.

We access two types of misreporting patterns to document managers' reaction: *traditional* *methods (TM)-sensitive misreporting* and *machine-sensitive misreporting.* We define *TM-sensitive misreporting* as irregularities detectable by traditional methodologies such as logistic and linear regression models, whereas *machine-sensitive misreporting* as irregularities that are detectable by AI and machine learning models but not traditional methodologies. We anticipate in response to rising machine readership - managers will adjust their reporting practices particularly on patterns that are more vulnerable to detection by machines. Thus, we expect a negative correlation between machine readership and *machine-sensitive misreporting.* In the case of *TM-sensitive misreporting*, if machine learning demonstrates an additional advantage within the linear structures over traditional methods, it is plausible that managers also reduce *TM-sensitive misreporting* in response to machine readership. However, if traditional methods already possess a robust ability to detect misreporting patterns within linear models, the impact of machine readership on the detection rates for such patterns might not be substantial. Thus, it is ex-ante unclear whether managers would adjust *TM-sensitive misreporting* along with the increase in machine readership.

We measure machine readership using the percentage of *Machine Downloads* of 10-K filings in the SEC Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system, following Cao, Jiang, Yang, and Zhang (2023). A *Machine Download* is defined as a download request from an IP address that conducts daily downloads of more than 50 unique firms' filings from the SEC EDGAR system. By aggregating the number of download requests made by machines and scaling it against the total number of downloads, we obtain our measure of *Machine Downloads %.* This measure can be used as a proxy for machine readership of financial reports for several reasons. Firstly, machine requests are a prerequisite and a necessary condition for machine reading activity. Secondly, given the substantial volume of machine downloads, it is highly improbable for humans alone to process all the files downloaded by machines. Moreover, Cao, Jiang, Yang, and Zhang (2023) demonstrate that machine downloads are associated with quicker and more profitable high speed trades once a filing becomes publicly available. Given the brief window of time available to process copious amounts of information, it is unlikely that a human subsequently reads the filings downloaded by machine.[2]

[2] Cao, Jiang, Yang, and Zhang (2023) provides evidence to support the assumption that managers are aware of machine downloads. First, managers and other market participants can obtain near real-time information about downloading activities via Freedom of Information Act (FOIA) requests. Secondly, managers can learn through other ways about the interest of AI-equipped investors in real time, including interactions with public relations departments and top management teams of the firms. Additionally, Cao, Jiang, Yang, and Zhang (2023) show that machine download is correlated with AI-equipped investors. They provide a list of the Top 20 institutional machine downloaders (Appendix C), which prominently features hedge funds and banking conglomerates known

Our measure of *machine-sensitive misreporting* is based on the *Restatement Risk* developed by Bertomeu, Cheynel, Floyd, and Pan (2021), which utilizes gradient boosted regression tree (GBRT) methods and data from public records to detect misstatements. Given the widespread usage of the machine learning methods and the availability of its input data to the public, we can reasonably expect a strong correlation between the *machine-sensitive misreporting* measure and the actual outcomes generated by market participants who employ machine learning techniques to analyze financial statements. In fact, institutional investors, particularly hedge funds, often have access to proprietary information sources and possess advanced computing capabilities. This suggests that our *machine-sensitive misreporting* measure likely captures the minimum level of irregularities detectable by these entities.[3]

Our findings reveal negative correlations between *Machine Downloads %* and all specifications of *machine-sensitive misreporting.* Specifically, a one-standard-deviation increase in machine readership is associated with a 10% decrease in restatement risk and a substantial 31% reduction in the probability of being identified as high risk. We also use an alternative measure of machine readership *AI ownership*, which captures a firm's cumulative ownership by investment companies that are equipped with artificial intelligence (AI) capabilities. *AI ownership* is a more direct measure for machine readership, and captures managers' awareness of machine readership more accurately. Results using *AI ownership* also show a significant and negative correlation with *machine-sensitive misreporting.* Overall, our findings suggest that machine readership has a disciplining effect on managers and reflects an improvement in financial reporting quality in areas that are sensitive to machine learning.

It is possible that firms with higher reporting quality attract more machine downloads because their filings are easier for machines to analyze. To address this potential reverse-causality issue, we employ an instrumental variable approach using the ownership-weighted *AI Talent Supply* accessible to institutional investors. We calculate *AI Talent Supply* based on local AI talent supplies at each institutional investor's headquarter, which is then scaled at the firm level according to institutional ownership (Jiang, Tang, Xiao, and Yao, 2021). *AI Talent Supply* and machine readership are positively correlated given that a significant proportion of machine downloads from the SEC EDGAR system are carried out by institutional investors (Crane,

for their utilization of AI. Relatedly, we utilize an alternative measure of *AI ownership* to directly test its effect on financial reporting quality, and the results align with our assumption.

[3] We also use the *fraud risk* measure developed Bao, Ke, Li, Yu, and Zhang (2020) using ensemble learning method, and the results are consistent and presented in Section 6.

Crotty, and Umar, 2022; Cao, Jiang, Yang, and Zhang, 2023). Moreover, the exclusion restriction also holds as the *AI Talent Supply* to institutional investors is unlikely to be associated with companies' financial reporting decisions. Through a two-stage least squared regression analysis, we find consistent results indicating a negative correlation between machine readership and *machine-sensitive misreporting.*

We expect the disciplining effect of machine readership to be more pronounced in areas where AI and machine learning techniques offer substantial advantages. We thus focus on two of AI and machine learning's key attributes: the ability to handle more complex information structures and the capacity to incorporate a greater variety of information sources. To explore the former, we use the setting of financial statement complexity. Financial statement complexity reflects both business complexity and reporting complexity (Guay, Samuels, and Taylor, 2016). More complex financial statements require more time and effort to extract relevant information, which leads to higher information processing costs for financial statement readers (Bloomfield, 2002). We investigate the interaction between financial statement complexity and machine readership on *machine-sensitive misreporting.* Our results reveal that, consistent with our hypothesis, the negative association between machine readership and *machine-sensitive misreporting* is more pronounced when financial statements are more complex. Next, to explore AI and machine learning's capacity to incorporate more sources of information, we use the setting of the introduction of satellite coverage of major retailers. Prior research shows investors incorporate alternative data sets into their investment decisions (Katona, Painter, Patatoukas, and Zeng, 2022; Kang, Stice-Lawrence, and Wong, 2021; Zhu, 2019). Consistent with our expectations, we find that firms that are covered by satellite data experience a reduction in *machine-sensitive misreporting* after the initiation of satellite data coverage.

For the *TM-sensitive misreporting* measures, we utilize traditional metrics for the likelihood of financial misreporting: F−scores from Dechow, Ge, Larson, and Sloan (2011) and discretionary accruals measures as proposed by Dechow, Sloan, and Sweeney (1995) and Kothari, Leone, and Wasley (2005). In contrast to its observed effect on *machine-sensitive misreporting*, we do not find significant correlations between *Machine Downloads %* and any of the *TM-sensitive misreporting* measures. Overall, the result implies that machine readership does not strongly affect financial misreporting trends that traditional methods can already spot. This suggests that the forte of machines lies in identifying complex, non-linear, and high-dimensional patterns. Additionally, this finding echoes the perspectives outlined by Kelly and Xiu (2023), confirming that the power of machine learning and AI lies in their capacity to incorporate larger amounts of predictors and rich nonlinear models.

The results above indicate that the disciplining effect of machine readership has on managers' misreporting decisions is primarily concentrated on misreporting patterns susceptible to detection through machine learning techniques. However, it remains uncertain whether this effect translates into an overall improvement in investors' welfare. To address this concern, we examine the relationship between machine readership and actual instances of misstatements. Our result reveals a negative correlation between *Machine Downloads %* and the likelihood of misstatements. This finding suggests that managers are unlikely to successfully shift the *machine-sensitive misreporting* to areas that eludes machine readership, but rather, there is an overall reduction in misreporting behavior. In conclusion, the adoption of machine effectively disciplines managers by reducing overall misreporting, especially *machine-sensitive misreporting*, and such reduction enhances the overall welfare of financial reporting users.

We further conduct several cross-sectional analyses to corroborate these findings. Specifically, we anticipate that firms will be more sensitive to the emergence of machine readership when the costs associated with misreporting are higher. Prior literature shows that, when a firm in a given industry issues a restatement, this can prompt investors to scrutinize other firms in the same industry (Gleason, Jenkins, and Johnson, 2008). Building on this observed spillover effect, we find a stronger negative correlation between machine readership and *machine-sensitive misreporting* for firms exposed to restating peers. Studies has also demonstrated that technology industries are generally subject to greater litigation risks (Kasznik and Lev, 1995; Ajinkya, Bhojraj, and Sengupta, 2005). Working from this premise, we find that firms in the technology sector also exhibit a more pronounced negative relationship between machine readership and *machine-sensitive misreporting.*

Our study makes three contributions to the existing literature. Firstly, it adds to the growing body of research exploring the impact of new technology adoption on firm behaviors. This research includes works by Charoenwong, Kowaleski, Kwan, and Sutherland (2022); Abis and Veldkamp (2022); Azarmsa, Liu, and Noh (2022); Cao, Jiang, Yang, and Zhang (2023); Zhu (2019). Specifically, Cao, Jiang, Yang, and Zhang (2023) demonstrates that firms manage sentiment and tone perception to cater to machine readers by avoiding words perceived as negative by algorithms, suggesting that new technology induces opportunistic behaviors in managers. Our paper contrasts with Cao, Jiang, Yang, and Zhang (2023) and shows that, since financial statement information is easier to verify than qualitative disclosure, machine readership actually exerts a disciplining effect on managers' opportunistic behavior. Additionally, our paper relates to Zhu (2019), which examines how the availability of alternative data affects corporate governance, particularly in terms of managers' opportunistic trading and investment efficiencies. While Zhu's findings focus on the availability of alternative data, our study centers on the technology itself and its application to a crucial information source for companies: their financial statements.

Secondly, prior research has documented that the information acquisition by fundamental investors increases scrutiny on managers' financial reporting choices. For example, Ahmed, Li, and Xu (2020) demonstrates that an increase in non-robotic information acquisition from EDGAR filings reduces managers' incentives to engage in misreporting. We complement these findings by showing that machine adopters, too, exert a disciplining effect on the quality of financial reporting.

Lastly, our findings reveal that machine readership significantly reduces machine-detectable misreporting and also lowers the likelihood of misstatement occurrences. This suggests that in response to increasing machine readership, managers do not simply shift their opportunistic behaviors to areas undetectable by machines. Rather, there is a general decrease in manipulation. This decline in misreporting signifies a notable improvement in the welfare of all investors, whether they are machine adopters or traditional investors.

2 Literature Review

2.1 Information acquisition from SEC-filings

Over the past decade, researchers have examined retrievals of EDGAR data to better understand how capital market participants use corporate financial statements. Drake, Roulstone, and Thornock (2015, 2016); Drake, Quinn, and Thornock (2017) and Loughran and McDonald (2017) explore the determinants and map out usage of the EDGAR database. Building on these studies, researchers figured out how to identify EDGAR users by unmasking the Internet Protocol (IP) addresses that accessed filings. Bozanic, Hoopes, Thornock, and Williams (2017) identifies acquisitions of financial reports by the Internal Revenue Service (IRS), Li, Lind, Ramesh, and Shen (2023) attends to the Federal Reserve's usage of accounting reports, Gibbons, Iliev, and Kalodimos (2021) studies analysts' information acquisition through EDGAR, and Bernard, Blackburne, and Thornock (2020) and Cao, Du, Yang, and Zhang (2021) identify accessions of corporate financial reports by industry peers. Still more studies have focused on how investors use EDGAR information, for example, via the association between EDGAR usage and institutional investors' profitability (Chen, Cohen, Gurun, Lou, and Malloy, 2020; Crane, Crotty, and Umar, 2022; Drake, Johnson, Roulstone, and Thornock, 2020) and geographical location (Dyer, 2021; Chen, 2022). Other researchers have examined specific filings, including Form 8-Ks (Ben-Rephael, Da, Easton, and Israelsen, 2022; Iliev, Kalodimos, and Lowry, 2021).

Earlier this year, Cao, Jiang, Yang, and Zhang (2023) became the first paper to study the feedback effect of machine readership. That is, how companies adjust the way they talk knowing that machines are listening. In the paper, Cao, Jiang, Yang, and Zhang (2023) identify machine downloads of EDGAR filings and use them to proxy machine readership. They show that firms with more machine readership prepare filings that are better suited to machine processing and avoid linguistic tones that tend to be perceived negatively by algorithms.

2.2 Financial Reporting Quality

Although measuring the quality of firms' financial reporting is difficult, a number of studies have used measures based on firms' accounting information. For example, accrual-based earnings management proxies have been widely used in the literature (see, e.g., Dechow, Ge, and Schrand (2010) for a review). Many of these studies are built on Jones (1991), which proposed that the nondiscretionary portion of accruals is correlated with changes in revenue, representing the change in economic circumstances and gross property, plant, and equipment (PPE); hence, the residual part of total accruals can be a proxy for earnings management. Dechow, Sloan, and Sweeney (1995) introduced a modified version of the Jones model, confining revenue to cash revenue only. Later, Kothari, Leone, and Wasley (2005) adopted a performance matching procedure to mitigate concerns regarding misspecification. Dechow, Ge, Larson, and Sloan (2011) then conducted a comprehensive analysis of the firm characteristics of material misstating firms and develop a prediction measure (F-score).

Recent studies have applied machine learning methods to further improve prediction power. Using Ensemble learning to predict Accounting and Auditing Enforcement Releases (AAERs), Bao, Ke, Li, Yu, and Zhang (2020) introduced a model that outperforms the F-score in Dechow, Ge, Larson, and Sloan (2011) and Cecchini, Aytug, Koehler, and Pathak (2010). Bertomeu, Cheynel, Floyd, and Pan (2021) use the gradient boosted regression tree (GBRT) to predict misstatements, also outperform the logit models in Dechow, Ge, Larson, and Sloan (2011).

2.3 An example of machine detection of financial misreporting

To gain a clearer understanding of how machine learning methods can detect misreporting patterns more effectively than traditional approaches, let's explore a derived example based on variables from prior research (Dechow, Ge, Larson, and Sloan, 2011; Bao, Ke, Li, Yu, and Zhang, 2020; Bertomeu, Cheynel, Floyd, and Pan, 2021). Imagine a firm trying to boost its Return on Assets (ROA) by one unit without triggering suspicion. In the pre-machine learning and AI era, the manager could do so by simultaneously increasing reported inventory by 0.78 unit (0.932/1.191), as calculated through the ratios of their coefficients in the model of Dechow, Ge, Larson, and Sloan (2011), (refer to Panel A in Figure 1). Importantly, this manipulation would not alter the predicted value of F-score in the model.

However, when investors adopt AI and machine learning methods, these approaches also consider the relationships between variables, including how changes in ROA and inventory interact. In this context, the previous manipulation becomes highly likely to trigger alerts from machine learning algorithms. This is because the ratio of *Change in ROA*/*Change in inventory* no longer remains unchanged; it shifts from *Original change in ROA*/*Original change in inventory* to (*Original change in ROA* +1)/(*Original change in inventory* +0.78). In Panel B of Figure 1, the last line illustrates some potential interactions based on existing input variables. When machine learning models incorporate these interactions, the predicted values no longer align with those before the manipulation.

This is just one example among numerous possibilities, as machine learning explores an extensive range of potential interactions. The precise interactions that influence the final model remain known only to the machine algorithm. In light of this intricate complexity, machine learning methods can discourage managers from attempting to manipulate earnings.

Panel C in Figure 1 illustrates the relation between misreporting patterns identifiable through traditional methods (*TM-sensitive misreporting*) and those detectable by machines (*Machine-sensitive misreporting*). We suggest that machine-detected misreporting (outer oval) can encompass misreporting caught by traditional methods (inner oval). The shaded blue area between the two ovals represents patterns exclusively discernible by machine learning.

Figure 1: A simplified example on TM-sensitive misreporting and machine-sensitive misreporting

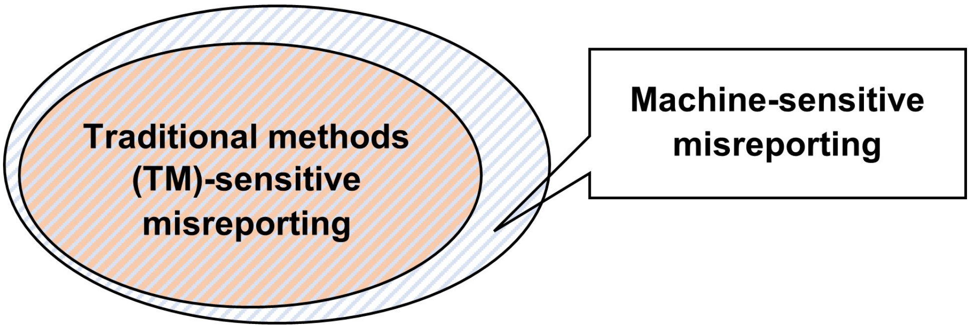
Panel A. Misreporting undetected by traditional models (Dechow et al., 2011)

$$\begin{aligned} F\text{-score} = &-7.893 + 0.790 \times (rsst\_acc) + 2.518 \times (ch\_rec) \\ &+ 1.191 \times (ch\_inv + 0.78(\frac{0.932}{1.191})) + 1.979 \times (soft\_assets) + 0.171 \times (ch\_cs) \\ &+ (-0.932) \times (ch\_roa + 1) + 1.029 \times (issue) \end{aligned}$$

Panel B. Misreporting now detected by machine learning methods

$$\begin{aligned} \text{Predicted Value} = &-7.893 + 0.790 \times (rsst\_acc) + 2.518 \times (ch\_rec) \\ &+ 1.191 \times (ch\_inv + 0.78(\frac{0.932}{1.191})) + 1.979 \times (soft\_assets) + 0.171 \times (ch\_cs) \\ &+ (-0.932) \times (ch\_roa + 1) + 1.029 \times (issue) \\ &\underbrace{+\beta_1 \times \frac{ch\_roa + 1}{ch\_inv + 0.78} + \beta_2 \times \frac{rsst\_acc}{ch\_roa + 1} + \beta_3 \times ch\_rec \times (ch\_inv + 0.78) + \ldots}_{\text{Possible interaction terms that trigger machine alerts}} \end{aligned}$$

Panel C. Relation between Machine-sensitive misreporting and TM-sensitive misreporting



3 Data and Variable Construction

Our sample comprises all publicly listed companies in the United States. The sample period spans from 2004 to 2016, as this timeframe aligns with the availability of SEC filings download records. We collected the companies' SEC filings from the SEC Edgar system, their financial information from Compustat, stock performance data from CRSP, analyst following data from I/B/E/S, and institutional investors' information from Thomson Reuters. The primary sample contains a total of $44,528$ firm-year observations. For a detailed explanation of the sample selection process, please refer to Appendix A.

In Appendix B, we present the definitions of all variables used in our study. Below are the definitions of the main variables: machine readership and financial misreporting.

3.1 Machine Readership

Our proxy for machine readership is the percentage of downloads of the companies' SEC filings that are conducted by machines, following Cao, Jiang, Yang, and Zhang (2023). The primary data source is the SEC EDGAR system. The EDGAR system provides information on all records of requests and downloads of SEC filings made through SEC.gov from 2003 to 2017, known as EDGAR Log File Data Sets. Each Log File contains IP address, date, time, CIK, and accession number associated with a given document request. Following Cao, Jiang, Yang, and Zhang (2023), we process the raw Log File data by narrowing it down to visits to Form 10-K filings using the accession number. We also exclude requests on index pages since they are not related to actual downloads. After matching these visits to CRSP and Compustat, our sample of filings consists of 60,495 10-K filings.

To construct *Machine Download*, we first identify an IP address as a machine reader if the IP address downloads more than 50 unique firms' filings from the SEC EDGAR system, following the definition of Lee, Ma, and Wang (2015). We then use this data to construct our main measure *Machine Downloads %.* We also include requests attributed to web crawlers in the SEC Log File Data as machine-initiated requests following Cao, Jiang, Yang, and Zhang (2023). Since the majority of download requests occur within seven days of a filing becoming available on EDGAR, we then aggregate the number of download requests made by machines (as defined in the previous step) for each 10-K filing over the seven-day period after it appears on EDGAR. We define *Machine Downloads* as the natural logarithm of the average number of machine downloads of a firm's filings at t that were filed during the $[t-4, t-1]$ quarters. Similarly, *Total Downloads* is the natural logarithm of the average number of all downloads of a firm's filing at time t that were filed during the $[t-4, t-1]$ quarters. Finally, we combine this figure with *Total Downloads* and define *Machine Downloads %* as the ratio of *Machine Downloads* to *Total Downloads* before taking the natural logarithm for both variables.

Cao, Jiang, Yang, and Zhang (2023) conduct two validity tests to establish the effectiveness of *Machine Downloads* as an effective proxy for the presence of machine readership. In their first test, they match the IP addresses with the highest volumes of machine downloads to the universe of investors who submit 13F filings. Their analysis reveals that half of the top twenty machine downloaders are prominent quantitative hedge funds, while brokers and investment banks with significant asset management divisions also featuring prominently (refer to Appendix D for more details on the top 20 entities). Additionally, they manually identify hedge funds that have adopted AI strategies and find a significant association between firms' AI hedge fund ownership and their *Machine Downloads %.* Both of these tests provide compelling evidence supporting the validity of *Machine Downloads %* as a proxy for machine readership.

Appendix C shows an increasing trend of *Machine Downloads* as a proportion of *Total Downloads*. Specifically, *Machine Downloads* for 10-K filings increased from 37.7% in 2004 to 83% in 2016 (37.5% in 2004 and 77.1% in 2016 for 10-Q filings[4]). In addition, the annual changes in *Machine Downloads* from 2004 to 2016 range from 32% to 125%, indicating high variations over the period.

The average *Machine Downloads %* in firm-level data increased from 48% in 2004 to 92% in 2016. At the same time, the standard deviation of *Machine Downloads %* decreased from 0.16 to 0.06, indicating that the proportion of *Machine Downloads* has increased throughout the capital market. The actual volume of *Machine Downloads* also increased from 2.76 to 7.46 during the same period. Based on Fama-French 12 Industry Classification, *Finance* and *Healthcare, Medical Equipment, and Drugs* have the highest average *Machine Downloads %*, at 77% and 76%, respectively, while *Consumer Nondurables* and *Telephone and Television Transmission* industries have the lowest average *Machine Downloads %*, 67% and 68% respectively. The *Telephone and television transmission* industry has the highest standard deviation of *Machine Downloads %*, with a value of 0.063, while *Utilities* has the lowest standard deviation of *Machine Downloads %*, with a value of 0.051.

We also use an alternative measure of machine readership by directly measuring the proportion of firm shares that are held by investment firms with AI capabilities, where investment firms include all “alternative asset managers" in Preqin, and all filers of form 13F, following Abis and Veldkamp (2022).[5] This *AI Ownership* measure is calculated by identifying an investment company as having AI capabilities if it has posted AI-related job openings over the past five years. We then aggregate ownership data at the firm level from the quarter before the firm's most recent filing. Since Burning Glass data is only accessible after 2010, the *AI Ownership* variable is available from 2011 to 2016.

3.2 Financial Misreporting

Our main proxy for *machine-sensitive misreporting* is *Restatement_ Risk*, a measure from Bertomeu, Cheynel, Floyd, and Pan (2021), which applies the gradient boosted regression tree

[4] We only tabulate from 2003 to 2016 because the SEC log information is partially available for the first half the year 2017. The decrease in 2016 is temporary, the upward trend still remains during the first half of 2017 (Cao, Jiang, Yang, and Zhang, 2023).

[5] We thank the authors of Cao, Jiang, Yang, and Zhang (2023) for providing the dataset of *AI Ownership* at the institutional investor level.

(GBRT) method to predict firm misstatements.[6] Bertomeu, Cheynel, Floyd, and Pan (2021) uses over a hundred variables organized into five categories: financial, audit, credit ratings, opinion divergence, and corporate governance. Based on this data, they provide a probability of misreporting measure that ranges from zero to one. Figure 2 shows the distributions of *Restatement_ Risk*. Since *Restatement_ Risk* is a probability measure that has a heavy right tail, most observations are concentrated in the area where *Restatement_ Risk* less than 0.1. To capture the significance of *Restatement_ Risk*, we follow the spirit of Dechow, Ge, Larson, and Sloan (2011) and create two binary measures, *I_ Rrisk1* and *I_ Rrisk2. I_ Rrisk1* is an indicator that equals one when the misstatement probability is above an annual cutoff with a Type I error of five percent, and zero otherwise.[7] This classification indicates that the non-misstating observations are correctly identified at 95 percent. Similarly, *I_ Rrisk2* equals one when the misstatement probability is above a certain threshold, that is, a yearly cutoff above which the Type I error is at ten percent, and zero otherwise.[8] Figure 3 shows a comparison between the indicator *I_ Rrisk1* and data on actual misstatement cases. The Type-I error in which $I\_Rrisk1 = 0$ and *actual restatement*= 1 is controlled at 5%. Among the whole sample, 21% of the observations where $I\_Rrisk1 = 1$ are actual misstatements.

We use a variety of measures for *TM-sensitive misreporting*, including *F-score, MJones*, and *Performance-matched Jones.* The *F-score* is a measure of the likelihood of financial misconduct developed by Dechow, Ge, Larson, and Sloan (2011) to predict material misstatements. *MJones* is an absolute value of the discretionary accruals in the modified Jones model (Dechow, Sloan, and Sweeney, 1995). We also use *Performance-matched Jones* based on the performance-matched discretionary accruals measure developed by Kothari, Leone, and Wasley (2005) and Kothari, Mizik, and Roychowdhury (2016).

Lastly, we also include the actual restatement, an indicator variable equal to one if the firm issues a restatement through SEC Form $8-K$ Item 4.02, and zero otherwise. We focus on non-reliance restatements that undermine previous or current financial statements or both, due to material accounting misstatements, and exclude nonmaterial errors, including out-of-period adjustments, and revision restatements, such as voluntary or mandatory changes in accounting

[6] We thank the authors of Bertomeu, Cheynel, Floyd, and Pan (2021) for sharing their measure online.

[7] Type I error is calculated as #(predicted misstate = 1 and observed misstate =0)#(observed misstate = 0). Since the incidence of restatement has decreased over the year, the R-Risk measure is also decreasing by year. We create the 5% cutoff by year to overcome the year trend. The average cutoff for 5% is 0.15, while the average cutoff for 10% is 0.10.

[8] In the untabulated analysis, we use one total cutoff to define *I_ Rrisk1* and *I_ Rrisk2* when the misstatement probability is above 0.157 and 0.099, respectively. The results remain unchanged under this alternative definition.

standards. We obtain the data from Audit Analytics.

3.3 Other Variables

We include several control variables related to firms' financial information. Specifically, we control for the following variables: Return on Assets (*ROA*), *Market-to-Book* ratio, *Size*, *Leverage*, *Sales growth*, research and development (*RD*), *Loss*, *Analysts following*, *Institutional ownership*, and *Big4*. *ROA* is defined as net income over the beginning-of-year total assets. *Market-to-Book* is the ratio of the market value of equity to the book value of equity. *Size* is the natural logarithm of market capitalization. *Leverage* is the ratio of total liabilities to assets at book values. *Sales growth* is the change in sales scaled by total sales. *RD* is defined as total research and development expenses scaled by sales. *Loss* is an indicator variable equal to one if net income is negative, and zero otherwise. *Analysts following* is the natural logarithm of the number of analysts that have issued forecasts for the firm. *Institutional ownership* is the ratio of the total shares of institutional ownership scaled by total shares outstanding. *Big4* is an indicator variable equal to one if the firm's auditor is a Big 4 auditing firm, and zero otherwise. All continuous variables are winsorized at the top and bottom 1%.

Table 1 presents descriptive statistics for the main variables. The average of *Machine Downloads %* included in the sample is 71.66%, with a standard deviation of 0.18. *Restatement_Risk* has a mean of 0.06 and a 75th percentile of 0.064, suggesting a long right tail. The mean of *I_Rrisk1* is 0.06, *I_Rrisk2* is 0.12, with a standard deviation of 0.24 and 0.32, respectively. The mean of the market-to-book measure (*MTB*) is 2.15, comparable to 2.28 of the Compustat universe. The mean of *Sales Growth* is 11%, and the mean of *Leverage* is approximately 20%. In addition, 29% of the firm-year observations experience negative net income (*Loss*). The average *Sales Growth*, *Leverage*, and *Loss* of the Compustat universe are 9%, 25%, and 38%, respectively. Overall, our sample is comparable to the Compustat universe.

Table 2 shows the Pearson and Spearman correlations for all variables in the sample. Both the Spearman and Pearson correlations between *Machine Downloads %* and *I_Rrisk1* are negative and significant at the 1% level, providing preliminary evidence that the increase in machine readership is negatively associated with the probability of misstating.

3.4 Model Specification

To test the effect of machine readership on financial reporting decisions, , we consider the following model specification:

$$\begin{aligned} \textit{Financial Misreporting}_{i,t} = & f(\textit{Machine Downloads \%}_{i,t},\ \textit{Total Downloads}_{i,t}, \\ & \textit{Controls}_{i,t},\ \textit{Fixed Effects}) \end{aligned}$$

where i is firm index, and t is year index. The dependent variable *Financial Misreporting*$_{i,t}$ contains two types of misreporting: *machine-sensitive misreporting* and *TM-sensitive misreporting. Machine-sensitive misreporting* includes *Restatement_Risk*, *I_Rrisk1*, and *I_Rrisk2.* Our measures for *TM-sensitive misreporting* are *MJones*, *PMJones*, and *F-score.* The control variables include *ROA*, *Market-to-Book*, *Size*, *Leverage*, *Sales growth*, *RD*, *Loss*, *Analysts following*, *Institutional ownership*, and *Big4*. As for model specifications, we apply an ordinary least squared (OLS) regression for the continuous dependent variables: *Restatement_Risk*, *MJones*, *PMJones*, and *F-score.* We apply a logistic regression for binary variables, *I_Rrisk1*, and *I_Rrisk2.* We include firm and year-fixed effects on the OLS regressions, and for the logistic regression, we control for year-fixed and industry-fixed effects based on Fama-French 48 classifications.

According to our hypothesis that machine readership disciplines managerial financial reporting behavior, we expect the coefficient on *Machine Download %* to be negative and significant for measures of *machine-sensitive misreporting.* However, we do not have clear expectations for the coefficient between machine readership and *TM-sensitive misreporting.*

4 Results

4.1 Machine Readership and Machine-sensitive Misreporting

Table 3 Panel A presents the regression results for machine readership and *machine-sensitive misreporting.* Columns (1) and (2) report OLS regression results between AI readers and the probability of misreporting (*Restatement_Risk*), controlling for firm-and-year fixed effects (column (1)) and industry-and-year fixed effects (column (2)). The coefficients on *Machine Downloads %* in the OLS regression model in columns (1) and (2) are −0.035 and −0.062, respectively, both negative and significant at 1% level. This suggests a negative association between machine

readership and *machine-sensitive misreporting*. Specifically, a one-standard-deviation increase in *Machine Downloads %* is associated with a 10% (column (1)) and 18% (column (2)) decrease in *Restatement_Risk*. Since the distribution of *Restatement_Risk* is concentrated on the left end, the economic significance can differ considerably across firms. We further investigate the relationship using two indicator variables *I_Rrisk1* and *I_Rrisk2*. These two measures represent substantial risks of restatement indicated by the machine learning method used in Bertomeu, Cheynel, Floyd, and Pan (2021). Columns (3) and (4) display the logistic regression results, which are consistent with previous findings. The coefficient on *Machine Downloads %* in the logistic regression model of *I_Rrisk1* in column (3) is -1.986, translating into an odds ratio of 0.137. In other words, a one-standard-deviation increase in *Machine Downloads %* is associated with a 31% decrease in the probability of being perceived as high risk by a machine reader. Similarly, the coefficient on the variable *Machine Downloads %* in column (4) is -1.784, further suggesting that firms with higher machine readership are less likely to experience high misreporting risks. Consistent with our hypothesis, all of the estimated coefficients of *Machine Downloads %* are negative and significant at the 1% confidence level.[9]

We also find consistent evidence that *Leverage, Sales Growth*, and *Loss* (*Size* and *RD*) are positively (negatively) associated with financial misreporting across all model specifications. In sum, the empirical evidence on the relationship between machine readers and *machine-sensitive misreporting* is consistent across machine-learning-based misreporting proxies. This evidence suggests that the increase in machine readership has an economically significant influence in that it reduces the likelihood of financial misstatements.

To further investigate the impact of machine readership on the quality of financial reporting, we conducted individual regressions of the top ten most important predictors for restatement according to Bertomeu, Cheynel, Floyd, and Pan (2021).[10] Our results, which are not presented in the draft, indicate that the coefficients of machine readership on accounting variables are not statistically significant. One key insight from Bertomeu, Cheynel, Floyd, and Pan (2021) is that accounting variables alone do not effectively detect misstatements; their significance in predicting misstatements arises when they are interacted properly with audit and market variables. Considering our findings with this added context, we infer that the reduction in

[9] The results are qualitatively similar when we use linear probability model (untabulated).

[10] The top ten predictors are: % of Soft Assets, Bid-ask spread, Non-audit fees/total fees, Qualified opinion, Change in operating lease activity, Short interest, Stock return volatility, Log of non-audit fee, Percentile rank of audit fee by auditor, and leverage (See Table 7 in Bertomeu, Cheynel, Floyd, and Pan (2021)

machine-sensitive misreporting is unlikely to be driven by changes in the individual levels of accounting items. Instead, it is more likely to occur through interactions among non-accounting variables.

Providing direct evidence of the actual change in interaction is challenging, as machine learning models generally do not offer an explicit functional form. However, certain examples can offer insights into these interactions. For example, *Change in cash sale*, which is one of the key predictors used by Bertomeu, Cheynel, Floyd, and Pan (2021), can be employed as an indicator for business expansion or retraction. A significant change in sales will often be accompanied by hiring or layoffs of employees. Therefore, we can expect a correlation between *Change in cash sale* and *Abnormal change in employees*, which is a significant non-accounting predictor of restatement. In fact, certain institutional investors may have access to proprietary sources of information. For example, they may utilize machine learning algorithms to monitor whether a firm's cost of goods sold aligns with the warehouses' truck traffic captured by satellite images. Machine learning algorithms can be trained to identify suitable interactions among these variables and effectively detect abnormal correlations. If any discrepancies are identified, these algorithms can also raise alerts for further investigation.

4.1.1 Alternative Measure of Machine Readership

To address potential concerns regarding the assumption that managers are aware of whether their firm's financial reports are being read by machines, we introduce a more direct measure: *AI Ownership.* This measure represents a firm's cumulative ownership by investment companies equipped with artificial intelligence (AI) capabilities. Given that a significant number of these investment companies are Form 13-F filers, the measure *AI Ownership* should render a more accurate representation of managers' awareness.

Table 3 Panel B presents the regression results. The coefficient of *AI Ownership* in the OLS regression model in column (2) is -0.026, negative and significant at 10% level, suggesting a negative association between *AI Ownership* and *machine-sensitive misreporting.* Specifically, a one-standard-deviation increase in *AI Ownership* is associated with a 2.6% decrease in *Restatement_ Risk.* Consistent with our main findings, the coefficients on *AI Ownership* in the logistic regression models in columns (3) and (4) are -2.699 and -3.420, negative and significant at the 5% and 1% level, respectively. In other words, a one-standard-deviation increase in *AI Ownership* is associated with an 12% (13%) decrease in the probability of being perceived as high risk

by a machine reader.

4.2 Instrument Variable Test

It is possible that higher financial reporting quality attracts more machine readers. To address this looming endogeneity problem, we explore a potential exogenous shock to machine readership. Institutional investors are the main sources of AI readers of financial reports (See Appendix D for the top machine download investors), and researchers have documented a recent trend in the finance industry emphasizing the recruitment of talent with experience in information technology and data analytics (Abis and Veldkamp, 2022). We use the ownership-weighted *AI Talent Supply* available to institutional investors as an instrumental variable to firms' machine readership, as the local talent supply with machine learning and AI experience can impact whether or not a given investor adopts AI and machine learning techniques. Our instrument, *AI Talent Supply*, is calculated based on the size of the AI-related employment pool local to the headquarters of a firm's investors, and thus is positively associated with *Machine Downloads %*. On the other hand, the *AI talent supply* for institutional investors is unlikely to be correlated with companies' financial reporting decisions. *AI Talent Supply* should be exogenous since the headquarters of an institutional investor will most likely have been established before the AI and machine learning trends took off. Following Jiang, Tang, Xiao, and Yao (2021) and Cao, Jiang, Yang, and Zhang (2023), we first obtain the number of people between 18 and 64 with undergraduate and/or graduate degrees in information technology in each state from 2011 to 2016 and calculate the talent supply using state-year population data from the Integrated Public Use Microdata Series (IPUMS) survey. Next, we match the headquarters states of the firm's institutional investors and thereby obtain the investor-state-level AI talent supply. Lastly, we take the average of the talent supply, weighted by the firm's level of institutional ownership, and create the local *AI Talent Supply* at the firm level.

We use a two-stage least-squares regression to conduct the analysis. In the first stage, we use an OLS model to estimate the correlation between the *AI Talent Supply* and machine readership. We expect a positive relationship between the instrumental variable and machine readership. The result is reported in column (1) of Table 4. *AI Talent Supply* is positively and significantly associated with *Machine Downloads %*. In the second stage, we regress the fitted value from the first stage, *Instrumented Machine Downloads %* on *machine-sensitive misreporting* measures. Columns (2) through (4) of Table 4 present the results of the *machine-sensitive misreporting* measures and *Instrumented Machine Downloads %*. The coefficients on *Instrumented Machine Downloads %* are all negative and significant, consistent with results in Table 3 and suggesting that firms with higher machine readership have a lower risk of misstatement.

4.3 Machine Readership and TM-sensitive Misreporting

Table 5 shows the results of machine readership and *TM-sensitive misreporting* using *MJones, PMJones*, and *F-score* across the columns. None of the coefficients on the three measures are statistically significant, which indicates that machine readership is not associated with *TM-sensitive misreporting*. In addition, the coefficients on *Machine Downloads %* across columns (3) through (5) are positive. The results using the alternative measure of machine readership, *AI Ownership*, are consistent and not significant.

The insignificance of machine readership coefficients is consistent with our expectations. Prior to the introduction of machine readership, traditional methods for identifying misreporting were highly effective in identifying certain abnormal patterns in individual accounting variables using linear or logistic models. Hence the adoption of machine readership, along with its underlying AI and machine learning technologies, has not had a significant impact on investors' ability to detect such irregularities. As a result, we do not observe a significant correlation between machine readership and *TM-sensitive misreporting* in Table 5. However, the increased utilization of machine readership has improved the detection of irregular patterns in more complex structures, such as from interactions among accounting and non-accounting variables. Consequently, in response to increased and still increasing machine readership, managers have altered the aspects of financial reporting irregularities that are only detectable by machines. The findings from Table 3 and Table 5 collectively indicate managers' selective adjustments to their financial reporting in response to the unique capabilities of machine readers.

4.4 Machine Readership and the Actual Restatement Incidence

Building on the previous findings, our next objective is to investigate whether machine readership can contribute to an overall enhancement in the quality of financial reporting. To this end, we examine the correlation between machine readership and actual restatements. Table 6 presents our findings. Column (1) shows the results of a logit regression model after controlling for industry and year fixed effects. Columns (2) and (3) use a linear probability model. In all three model specifications, we observe negative and statistically significant coefficients on

Machine Downloads %. In particular, when we control for firm and year fixed effects in the linear probability model, a one-standard-deviation increase in *Machine Downloads %* is associated with a 9% reduction in restatements.

Based on the findings presented in the previous three tables, we can conclude that machine readership has a positive impact on the overall quality of financial reporting, as demonstrated by the decreased likelihood of restatements. This improvement in financial reporting quality is primarily observed in the realm of *machine-sensitive misreporting.*

5 Additional Analyses

In this section, we conduct several additional analyses to further support the idea that machine readership disciplines managers' financial reporting behavior. First, we explore settings wherein machine learning techniques can plausibly offer substantial advantages over traditional methods. The primary benefits of machine learning are its ability to handle more complex information structures and its capacity for incorporating a greater variety of information sources. We conduct two analyses to confirm the advantages of machine readership. We then examine the effect of machine readership by exploring variations in the costs of misreporting.

5.1 Advantages of Machines - Complex Information Structures

Machine learning excels at handling complex information structures and incorporating a wide variety of information sources. In this section, we specifically examine machine learning's capacity to process complex information structures by leveraging variations in financial statement complexity.

Financial statement complexity can arise from the intricacies of a firm's business transactions as well as the complications associated with reporting standards. Both forms of complexity can pose more significant challenges for traditional readers in processing valuable information than they would for machine readers. To capture this difference, we utilize the length of a firm's 10-K report as a proxy for financial statement complexity, calculated as the natural logarithm of the word count.[11] Previous research by Guay, Samuels, and Taylor (2016) demonstrates a positive relationship between voluntary disclosure and financial statement complexity, measured by both the length and readability of a firm's 10-K. Notably, the length of the 10-K encompasses the

[11] We obtain the dataset from the Loughran-McDonald 10-K summary files: link.

impact of readability, as the two measures are highly correlated. Consequently, we adopt the length of the firm's 10-K as our measure of financial statement complexity.

To incorporate this complexity measure into our analysis, we create an indicator variable named *FS Complexity.* This variable takes a value of one if the length of the firm's 10-K exceeds the sample average, and zero otherwise. Alternatively, we define a variable called *FS Complexity_Alt* as an indicator variable taking a value of one if the number of unique words in a 10-K is above the median, and zero otherwise. The results obtained using this alternative measure are qualitatively similar (untabulated). We also include an interaction term between *FS Complexity* and *Machine Downloads %.* The presence of this interaction term allows us to examine whether machine readership can mitigate the challenges associated with information processing frictions. If machine readership is effective in reducing these frictions, we expect the coefficient of the interaction term to be negative.

Table 7 presents the results. Across all model specifications, the coefficients on *Machine Downloads %* exhibit negative and significant values, aligning with our main findings. This consistency reinforces the notion that machine readership has a disciplining effect on financial reporting behavior. Furthermore, the coefficients on *FS Complexity* and the measures related to machine-sensitive misreporting consistently display positive and significant values. This suggests that firms with higher financial statement complexity have a higher likelihood of misstatement. This finding supports the notion that complexities in financial statements can pose challenges for accurate reporting.

Importantly, in line with our hypothesis, the coefficients on the interaction term between *Machine Downloads %* and *FS Complexity* are negative and significant. This indicates that the disciplining effect of machine readership is stronger among firms that produce more complex financial statements. For instance, based on the results presented in column (2), a one-standard-deviation increase in *Machine Downloads %* corresponds to a 36% decrease in the probability of being perceived as high risk by a machine reader, specifically for firms with higher *FS Complexity.* This implies that the impact of machine readership on reporting quality is more pronounced among firms with greater financial statement complexity.

5.2 Advantages of Machines - Alternative Data Coverage

Another significant advantage of machine learning and AI techniques is the ability to handle and incorporate vast amounts of data. To further strengthen our analysis of machine readership, we employ an event study that capitalizes on this capability. One valuable type of data frequently utilized by investors is data that captures consumer footprints, such as satellite images of retail parking lots. Due to the sheer volume of satellite images, such information has to be processed by machine learning models before it can be effectively used to inform investment decisions. Prior studies have shown that this kind of data contains incremental information for stock prices (Zhu, 2019; Kang, Stice-Lawrence, and Wong, 2021; Katona, Painter, Patatoukas, and Zeng, 2022). We build on the setting from Katona, Painter, Patatoukas, and Zeng (2022) on the staggered introduction of satellite image coverage, and conduct a difference-in-differences test of *machine sensitive misreporting.*[12] The underlying premise of our analysis is that the availability of satellite image coverage provides machine readers with a broader pool of information to incorporate into their analyses. This expanded information set increases the likelihood of detecting patterns and anomalies that may be indicative of potential misreporting. As a result, firms are less likely to engage in misreporting behavior due to the heightened scrutiny and detection capabilities of machine readers.

The findings presented in Table 8 demonstrate the association between alternative data availability and *Restatement_Risk*. Specifically, the negative coefficients of the interaction terms between treatment firms (*Alt_Data_Covered*) and the post-coverage period (*Post_Coverage*) indicate that the presence of alternative data is linked to a reduced likelihood of restatements. This suggests that the availability of alternative data serves as a mitigating factor in restatement risk. Moreover, even when accounting for alternative data coverage, the coefficients on *Machine Downloads %* remain negative and significant. This implies that the disciplining effect of machine readership is further reinforced by the presence of alternative data.

Taken together, these findings provide compelling evidence supporting the notion that the disciplining effect of machine readership is particularly pronounced in situations where machines offer substantial advantages. The availability of alternative data, in conjunction with machine readership, contributes to improved financial reporting quality and a reduced likelihood of restatements.

[12] We thank the authors of Katona, Painter, Patatoukas, and Zeng (2022) for sharing the list of treatment and control firms.

5.3 Potential Costs of Misstatement

We next explore whether firms that incur different restatement costs react differently to the adoption of machine readership. We utilize two scenarios: firms with a peer firm that recently issued a restatement and firms in high-litigation industries.

5.3.1 Restatement costs: Peer Restatements

Gleason, Jenkins, and Johnson (2008) examine the spillover effect of unfavorable information on non-misstating firms in the same industry. They find that financial misreporting by one firm can lead capital market participants to scrutinize the financial statements of other firms in the same industry. In this section, we examine whether the association between machine readership and financial misreporting is stronger among firms whose peers recently issued restatements. To test this prediction, we define *Peer Restate* as an indicator variable with value of one for firms that have restating peers in the same year, and zero otherwise. We run an OLS regression of the main analysis and include an interaction term of *Machine Downloads %* × *Peer Restate.*

Table 9 reports the results. The coefficients on the interaction between *Machine Downloads %* and *Peer Restate* are negative and significant across all columns, indicating that the disciplinary effect of machine readers is stronger among firms exposed to a spillover effect from a restating peer than firms without recent restating peers. The effect is also economically significant. For example, in column (1), the effect of machine readership on *Restatement_ Risk* is 12.5% higher for firms with a restating peer. We also find that *Peer Restate* is positively correlated with *machine-sensitive misreporting.* This correlation is consistent with findings in prior literature that firms with restating peers are subject to higher scrutiny by capital market participants as well as regulators, which manifests in a positive relation between *Peer Restate* and our measures of misreporting. Overall, these results provide evidence that when firms are subjected to greater scrutiny for potential misstatements, they react more strongly to machine readership and experience a stronger disciplining effect in the context of *machine-sensitive misreporting.*

5.3.2 Litigation Costs: High Litigation Industries

In large part because they tend to be high risk, high technology (HT) firms are exposed to an above-average risk of shareholder lawsuits, resulting in large price fluctuations and potential losses to investors (Kasznik and Lev, 1995). The cost of misreporting for high-tech firms is thus higher than it is for non-high-tech firms. A marginal increase in the likelihood of managerial misreporting being revealed by machine readership thus has a stronger effect on high-tech firms' reporting decisions. Following the definitions in Kasznik and Lev (1995) and Ajinkya, Bhojraj, and Sengupta (2005), we define an indicator variable *Litigate* that has a value of one if the firm is in the following industries: biotechnology (SIC codes $2833-2836$), R&D services ($8731-8734$), programming ($7371-7379$), computers ($3570-3577$), or electronics ($3600-3674$), and zero otherwise. We partition the sample based on the values of *Litigate* and conduct the main analysis on *machine-sensitive misreporting.*

The results are reported in Table 10. Consistent with our previous findings, the coefficients on *Machine Downloads %* are negative and significant, and the coefficients on *Litigate* are positive and significant. For the interaction term, *Machine Downloads %* × *Litigate*, the coefficient is significant and negative, confirming our hypothesis that firms facing higher litigation costs react more strongly to machine readership and reduce their machine sensitive misreporting to a greater degree.

6 Robustness Tests – Alternative Measures

6.1 Alternative Measure for Machine Readership

Cao, Jiang, Yang, and Zhang (2023) aggregate machine-generated and other requests for each filing in EDGAR within a seven-day window because the majority of requests occur within seven days of the filing being made accessible. As a robustness check, we extend this period to fourteen days (thirty days) to verify whether the aggregation method influences the results of this study. Specifically, we define *Machine Downloads %_ 14D (30D)* similarly to the original *Machine Downloads %* variable, except that we aggregate machine-generated requests and total requests within fourteen (thirty) days, respectively. Based on this alternative approach, we expect the results to be consistent with our main analysis.

Panel A in Table 11 displays the results. Consistent with our prediction, we find that the coefficients on *Machine Downloads %_ 14D (30D)* for the OLS and logistic regressions are negative and significant. Moreover, the magnitude of the coefficients are very close to those in the main analysis, which are reported in Panel A of Table 3. These results provide consistent evidence of the disciplining effect of machine readership on firms' financial reporting decisions.

6.2 Alternative Measures of Machine-sensitive Misreporting

Our main *machine-sensitive misreporting* measure is developed and adapted from Bertomeu, Cheynel, Floyd, and Pan (2021) to predict accounting misstatements. Another recent paper, Bao, Ke, Li, Yu, and Zhang (2020), also uses a machine learning approach to predict accounting fraud. They employ the raw accounting numbers that are used to construct the ratios in Dechow, Ge, Larson, and Sloan (2011), along with an ensemble learning method, to predict the detected material accounting misstatements disclosed in the SEC's Accounting and Auditing Enforcement Releases (AAERs). Their model outperforms the logistic regression model by a large margin.

We construct a variable, *Fraud Score*, following the codes provided by Bao, Ke, Li, Yu, and Zhang (2020). The results are reported in Panel B of Table 11. We find that the coefficients on *Machine Downloads %* are negative, and the coefficients are significant at a 10% confidence interval. Note that the prediction model for Bao, Ke, Li, Yu, and Zhang (2020) is designed for detected material accounting misstatements, which include more severe misreporting behaviors and are also subject to the SEC detection decisions. In general, the results are consistent with our main finding that machine readership is negatively related to *machine-sensitive misreporting.*

7 Conclusion

Recent developments in AI and machine learning have garnered significant attention within the finance literature. However, research exploring the feedback effects of these emerging technologies on firms financial reporting remains somewhat limited. Our contributions align with multiple research streams in this domain (Zhu, 2019; Cao, Jiang, Yang, and Zhang, 2023), as we delve into the intricate relationship between new technology adoptions by financial statement users and firms' decisions pertaining to financial reporting. Our findings indicate that managers tend to reduce overall misreporting when machine readership is higher, and the reduction is concentrated on patterns that are sensitive to machine analysis. Our findings suggest a disciplining effect from new technologies.

This is also the first study to provide evidence on how the adoption of AI and machine learning affects financial statement preparation, an underexplored yet essential topic in financial accounting. Future studies can further our findings by exploiting different aspects of AI adoption and examining the different channels through which AI adoption affects managerial decision-making.

We acknowledge that our conclusions may be constrained by the potential noise and Type-II errors in the measures of *TM-sensitive misreporting* (Leone, 2022). In this regard, we also advocate for the adoption of more precise misreporting measures, either through the use of clear identification strategies or measures conducted using machine learning approaches in future research.

References

Abis, S. and L. Veldkamp (2022). The Changing Economics of Knowledge Production.

Ahmed, A. S., Y. Li, and N. Xu (2020). Tick size and financial reporting quality in small-cap firms: Evidence from a natural experiment. *Journal of Accounting Research 58*(4), 869–914.

Ajinkya, B., S. Bhojraj, and P. Sengupta (2005). The Association between Outside Directors, Institutional Investors and the Properties of Management Earnings Forecasts. *Journal of Accounting Research 43*(3), 343–376.

Azarmsa, E., L. Y. Liu, and S. Noh (2022). (In)Consistent Internal-External Communication? How Internal Communication Technology Affects Voluntary Disclosure. *SSRN Electronic Journal*.

Bao, Y., B. Ke, B. Li, Y. J. Yu, and J. Zhang (2020). Detecting Accounting Fraud in Publicly Traded U.S. Firms Using a Machine Learning Approach. *Journal of Accounting Research 58*(1), 199–235.

Bauguess, S. W. (2017, June). The Role of Big Data, Machine Learning, and AI in Assessing Risks: a Regulatory Perspective.

Ben-Rephael, A., Z. Da, P. D. Easton, and R. D. Israelsen (2022, September). Who Pays Attention to SEC Form 8-K? *Accounting Review 97*(5), 59–88.

Bernard, D., T. Blackburne, and J. Thornock (2020, June). Information flows among rivals and corporate investment. *Journal of Financial Economics 136*(3), 760–779.

Bertomeu, J., E. Cheynel, E. Floyd, and W. Pan (2021, June). Using machine learning to detect misstatements. *Review of Accounting Studies 26*(2), 468–519.

Bloomfield, R. J. (2002). The "incomplete revelation hypothesis" and financial reporting. *Accounting Horizon 16*(3), 233–243.

Bozanic, Z., J. L. Hoopes, J. R. Thornock, and B. M. Williams (2017). IRS Attention. *Journal of Accounting Research 55*(1), 79–114.

Cao, S., W. Jiang, B. Yang, and A. L. Zhang (2023). How to Talk When a Machine is Listening?: Corporate Disclosure in the Age of AI. *Review of Financial Studies forthcoming*, 58.

Cao, S. S., K. Du, B. Yang, and A. L. Zhang (2021). Copycat Skills and Disclosure Costs: Evidence from Peer Companies Digital Footprints. *Journal of Accounting Research 59*(4), 1261–1302.

Cecchini, M., H. Aytug, G. J. Koehler, and P. Pathak (2010, July). Detecting Management Fraud in Public Companies. *Management Science 56*(7), 1146–1160.

Charoenwong, B., Z. T. Kowaleski, A. Kwan, and A. Sutherland (2022, September). RegTech.

Chen, H., L. Cohen, U. Gurun, D. Lou, and C. Malloy (2020, October). IQ from IP: Simplifying search in portfolio choice. *Journal of Financial Economics 138*(1), 118–137.

Chen, J. V. (2022, November). The wisdom of crowds and the market's response to earnings news: Evidence using the geographic dispersion of investors. *Journal of Accounting and Economics*, 101567.

Crane, A., K. Crotty, and T. Umar (2022, June). Hedge Funds and Public Information Acquisition. *Management Science*, mnsc.2022.4466.

Dechow, P., W. Ge, and C. Schrand (2010, December). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics 50*(2), 344–401.

Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan (2011, March). Predicting Material Accounting Misstatements: Predicting Material Accounting Misstatements. *Contemporary Accounting Research 28*(1), 17–82.

Dechow, P. M., R. G. Sloan, and A. P. Sweeney (1995). Detecting Earnings Management. *The Accounting Review 70*(2), 193–225.

Drake, M. S., B. A. Johnson, D. T. Roulstone, and J. R. Thornock (2020, March). Is There Information Content in Information Acquisition? *Accounting Review 95*(2), 113–139.

Drake, M. S., P. J. Quinn, and J. R. Thornock (2017, September). Who Uses Financial Statements? A Demographic Analysis of Financial Statement Downloads from EDGAR. *Accounting Horizons 31*(3), 55–68.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2015). The Determinants and Consequences of Information Acquisition via EDGAR. *Contemporary Accounting Research 32*(3), 1128–1161.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2016, April). The usefulness of historical accounting reports. *Journal of Accounting and Economics 61*(2), 448–464.

Dyer, T. A. (2021, August). The demand for public information by local and nonlocal investors: Evidence from investor-level data. *Journal of Accounting and Economics 72*(1), 101417.

Gibbons, B., P. Iliev, and J. Kalodimos (2021, February). Analyst Information Acquisition via EDGAR. *Management Science 67*(2), 769–793.

Gleason, C. A., N. T. Jenkins, and W. B. Johnson (2008). The Contagion Effects of Accounting Restatements. *The Accounting Review 83*(1), 83–110.

Guay, W., D. Samuels, and D. Taylor (2016, November). Guiding through the Fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics 62*(2), 234–269.

Ham, C. C., R. N. Hann, M. Rabier, and W. Wang (2022, June). Auditor Skill Demands and Audit Quality: Evidence from Job Postings.

Iliev, P., J. Kalodimos, and M. Lowry (2021, November). Investors Attention to Corporate Governance. *The Review of Financial Studies 34*(12), 5581–5628.

Jiang, W., Y. Tang, R. Xiao, and V. Yao (2021, April). Surviving the Fintech Disruption.

Jones, J. J. (1991). Earnings Management During Import Relief Investigations. *Journal of Accounting Research 29*(2), 193–228.

Kang, J. K., L. Stice-Lawrence, and Y. T. F. Wong (2021). The firm next door: Using satellite images to study local information advantage. *Journal of Accounting Research 59*(2), 713–750.

Kasznik, R. and B. Lev (1995). To Warn or Not to Warn: Management Disclosures in the Face of an Earnings Surprise. *The Accounting Review 70*(1), 113–134.

Katona, Z., M. Painter, P. N. Patatoukas, and J. J. Zeng (2022, July). On the Capital Market Consequences of Big Data: Evidence from Outer Space. *SSRN Scholarly Paper*.

Kelly, B. T. and D. Xiu (2023). Financial machine learning. *NBER Working Paper* (w31502).

Kothari, S. P., A. J. Leone, and C. E. Wasley (2005, February). Performance matched discretionary accrual measures. *Journal of Accounting and Economics 39*(1), 163–197.

Kothari, S. P., N. Mizik, and S. Roychowdhury (2016, March). Managing for the Moment: The Role of Earnings Management via Real Activities versus Accruals in SEO Valuation. *The Accounting Review 91*(2), 559–586.

Law, K. and M. Shen (2020, July). How Does Artificial Intelligence Shape Audit Firms?

Lee, C. M. C., P. Ma, and C. C. Y. Wang (2015, May). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics 116*(2), 410–431.

Leone, A. J. (2022). Is construct validity being overlooked in accounting research? A critical review of earnings management proxies.

Li, E. X., G. Lind, K. Ramesh, and M. Shen (2023). Externalities of accounting disclosures: evidence from the federal reserve. *The Accounting Review*, 1–27.

Loughran, T. and B. McDonald (2017, April). The Use of EDGAR Filings by Investors. *Journal of Behavioral Finance 18*(2), 231–248.

Yan, X. and L. Zheng (2017). Fundamental analysis and the cross-section of stock returns: A data-mining approach. *The Review of Financial Studies 30*(4), 1382–1423.

Zhu, C. (2019, May). Big Data as a Governance Mechanism. *The Review of Financial Studies 32*(5), 2021–2061.

Appendices

A Sample Selection

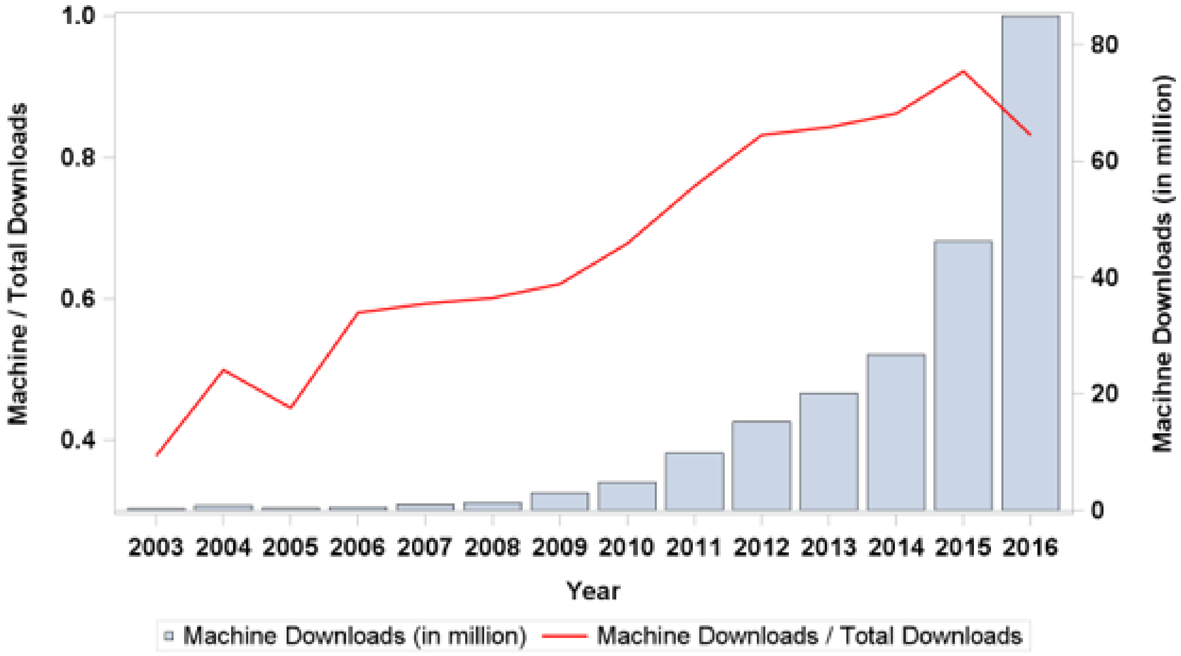
	Firm-years
Firm-years from universe of Compustat (with total assets)	
(fiscal years 2004 – 2016**)**	97,443
Observations with missing control variables (from Compustat and CRSP)	(19,196)
Observations with missing machine downloads information (from EDGAR)	(17,752)
Observations with missing analyst and institutional ownership information (from I/B/E/S and Thomson Reuters Ownership database)	(15,967)
Total Observations	**44,528**

B Variable Definitions

Variable	Description
Restatement variables	
Restatement_ Risk	The probability of Restatement Risk following Bertomeu, Cheynel, Floyd, and Pan (2021), using Audit Analytics' Restatement data and GBRT model.
I_ Rrisk1	Indicator variable equal to one if R-Risk is above a certain threshold – a yearly cutoff point above which Type I error is at five percent, and zero otherwise.
I_ Rrisk2	Indicator variable equal to one if R-Risk is above a certain threshold – a yearly cutoff point above which Type I error is at ten percent, and zero otherwise.
Machine Readership variables	
AI Ownership	Following Cao, Jiang, Yang, and Zhang (2023), we identify investment companies as AI-equipped if they have posted job openings related to AI technology in the past five years, according to Burning Glass job posting data between 2011 and 2016. Then, we aggregate the ownership of AI-equipped investment company shareholders at the firm-year-level.
Machine Downloads	The natural logarithm of the average number of machine downloads of a firm's filing at t that were filed during the [t- 4, t-1] quarters. To measure Machine Downloads, we identify an IP address downloading more than 50 unique firms' filings (Lee, Ma, and Wang, 2015). Next, we aggregate the daily raw downloads data for each filing within seven days after it becomes available on EDGAR.
Machine Downloads %	The ratio of Machine Downloads to Total Downloads before taking the natural logarithm.
Other variables	
AI Talent Supply	Following Jiang, Tang, Xiao, and Yao (2021) and Cao, Jiang, Yang, and Zhang (2023), we obtain the number of people between 18 and 64 with college or graduate degree in information technology, from Integrated Public Use Microdata Series (IPUMS), a state-level data from 2011 to 2016 scaled by state population. Then, we match the headquarters of the institutional investors and states into an investor-state-year-level AI talents supplies. Last, we aggregate at the firm-level.
Analysts Following	The natural logarithm of 1+ the number of analysts following a firm.
Big 4	Indicator variable equal to one if the firm is audited by the big 4 auditing firm, and zero otherwise.
F-score	Score variable following Dechow, Ge, Larson, and Sloan (2011), impying that a score of 1.00 indicates that the firm has the same probability of misstating than the unconditional probability.
FS Complexity	Indicator variable equal to one if the number of words in a 10-K is above median, and zero otherwise.
Institutional Ownership	The proportion of shares held by institutional investors.

Leverage	Total liabilities divided by total assets.
Litigate	Indicator variable equal to one if the firm belongs to the biotechnology (SIC codes 2833–2836), R&D services (8731–8734), programming (7371–7379), computers (3570–3577), electronics (3600–3674), and zero otherwise.
Loss	Indicator variable equal to one if net income is negative, and zero otherwise.
Market-to-Book	The ratio of the market value of equity to the book value of equity.
MJones	The absolute value of residuals from the modified Jones model following Dechow, Sloan, and Sweeney (1995).
Peer Restate	Indicator variable equal to one if a firm has peer firm in the same four-digit sic code that announced restatement in the same year, and zero otherwise.
Performance-matched Jones	The absolute value of residuals from the performance-matched modified Jones model following Kothari, Leone, and Wasley (2005).
Restatement	An indicator variable is the firm issued a restatement filing in the Form $8-K$ Item 4.02.
RD	Total research and development expenses scaled by sales.
ROA	Firms' net income over lagged total assets.
Sales Growth	The one-year percentage change in sales for the year prior to the current fiscal year.
Size	The natural logarithm of market capitalization.
Total Downloads	The natural logarithm of the average number of total downloads of a firm's filing at t that were filed during the [t- 4, t-1] quarters.

C Trend of Machine Downloads



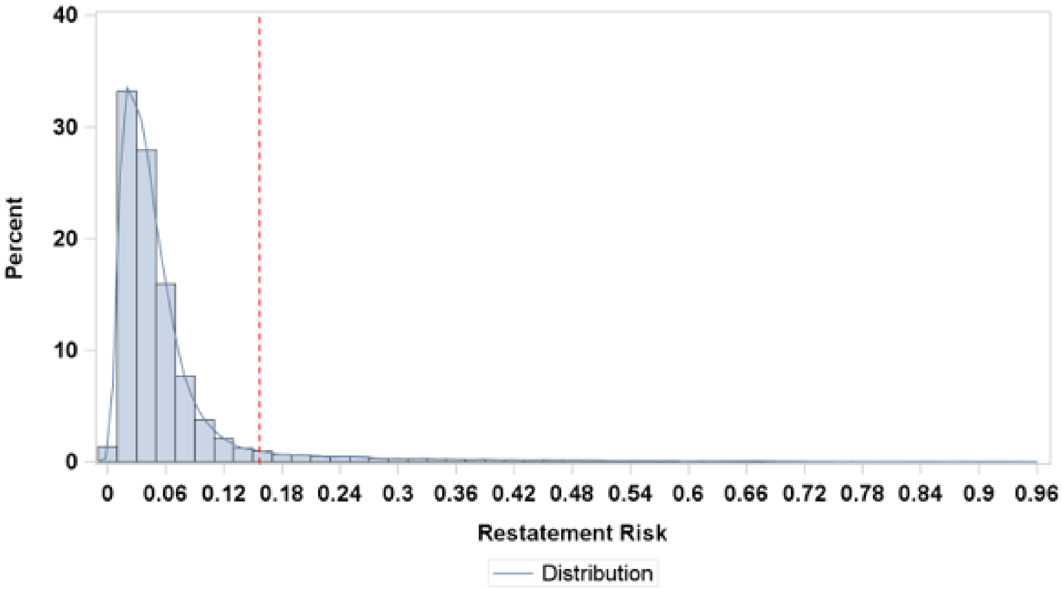
Source: Cao, Jiang, Yang, and Zhang (2023).

D Top 20 Machine Downloaders

Rank	Name of institution	#MD	Type of institution
1	Renaissance Technologies	536,753	Quantitative hedge fund
2	Two Sigma Investments	515,255	Quantitative hedge fund
3	Barclays Capital	377,280	Financial conglomerate with asset management
4	JPMorgan Chase	154,475	Financial conglomerate with asset management
5	Point72 Asset Management	104,337	Quantitative hedge fund
6	Wells Fargo	94,261	Financial conglomerate with asset management
7	Morgan Stanley	91,522	Investment bank with asset management
8	Citadel LLC	82,375	Quantitative hedge fund
9	RBC Capital Markets	79,469	Financial conglomerate with asset management
10	D. E. Shaw Co.	67,838	Quantitative hedge fund
11	UBS AG	64,029	Financial conglomerate with asset management
12	Deutsche Bank AG	55,825	Investment bank with asset management
13	Union Bank of California	50,938	Full-service bank with private wealth management
14	Squarepoint Ops	48,678	Quantitative hedge fund
15	Jefferies Group	47,926	Investment bank with asset management
16	Stifel, Nicolaus Company	24,759	Investment bank with asset management
17	Piper Jaffray	18,604	Investment bank with asset management
18	Lazard	18,290	Investment bank with asset management
19	Oppenheimer Co.	15,203	Investment bank with asset management
20	Northern Trust Corporation	11,916	Financial conglomerate with asset management

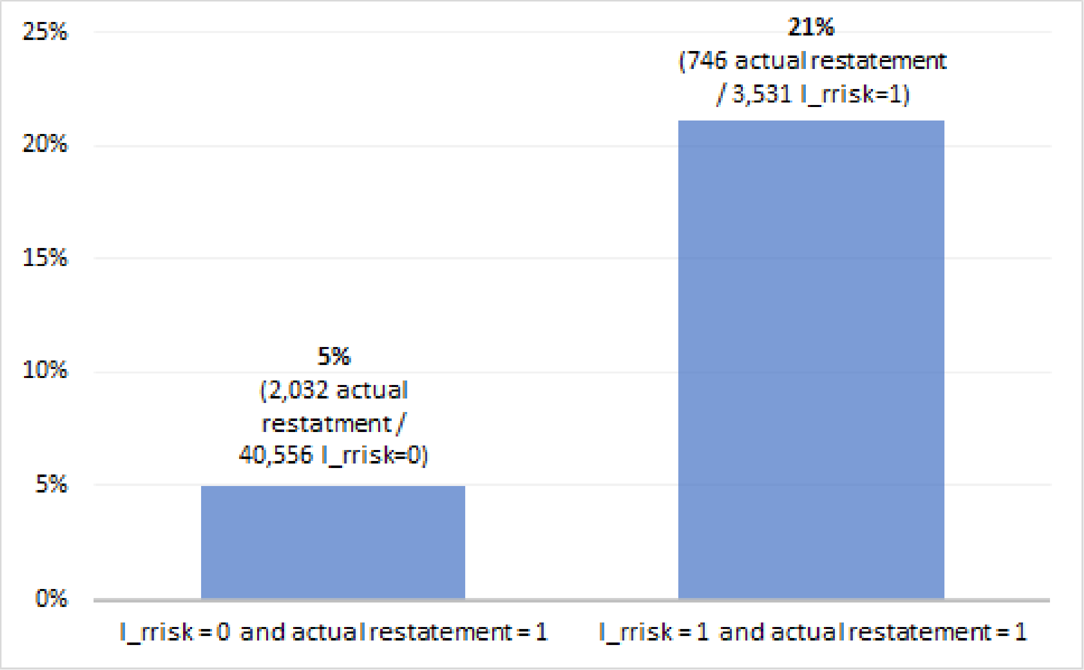
Source: Cao, Jiang, Yang, and Zhang (2023).

Figure 2: Distribution of Machine-sensitive Misreporting (*Restatement_ Risk*)



Notes: This figure plots the distribution of *Restatement_ Risk* measure provided by Bertomeu, Cheynel, Floyd, and Pan (2021) . Following the method in Dechow, Ge, Larson, and Sloan (2011), we create an indicator variable of *Restatement_ Risk.* The red vertical dash line, where = 0.157, indicates that if an indicator variable for misstatement takes value of 1 for area above the line, the Type 1 error for such indicator variable is five percent.

Figure 3: Validation of *I_ Rrisk1* using actual restatement



Notes: This figure displays an evaluation of *I_ Rrisk1* and the actual misstatement. *I_ Rrisk1* is an indicator variable equals one when Restatement Risk is greater than the cutoff values where the Type 1 error rate is at 5%. Cutoff value is calculated at an annual basis. We calculate the percentage of actual restatements among $I\_Rrisk1 = 1$ and $I\_Rrisk1 = 0$.

Table 1: Descriptive Statistics

This table presents descriptive statistics for the variables used in our analysis. The sample is 44,087 firm-year observations and covers firms over 2004 – 2016 with non-missing financial data from Compustat, CRSP, Thomson Reuters Ownership database, I/B/E/S, IPUMS, and Burning Glass. Variable definitions are listed in Appendix B.

Variable	N	Std Dev	Mean	25th Pctile	Median	75th Pctile
AI Ownership	19,698	0.05	0.04	0.00	0.02	0.08
AI Talent Supply	18,520	1.81	10.03	9.59	10.67	11.13
Analysts Following	44,528	1.01	1.57	0.69	1.61	2.40
Big4	44,528	0.45	0.72	0	1	1
Institutional Ownership	44,528	0.34	0.56	0.25	0.63	0.85
Leverage	44,528	0.19	0.20	0.03	0.15	0.32
Loss	44,528	0.46	0.29	0	0	1
Machine Downloads %	44,528	0.18	0.72	0.59	0.74	0.86
MTB	44,528	2.15	2.54	1.11	1.83	3.18
Restatement_ Risk	44,528	0.08	0.06	0.02	0.04	0.06
I_ Rrisk1	44,528	0.24	0.06	0	0	0
I_ Rrisk2	44,528	0.32	0.12	0	0	0
RD	44,528	1.38	0.26	0.00	0.00	0.05
ROA	44,528	0.13	0.00	-0.02	0.02	0.07
Sales Growth	44,528	0.38	0.11	-0.03	0.06	0.18
Size	44,528	1.92	6.31	4.84	6.29	7.72
Total Downloads	44,528	1.57	5.14	3.83	4.98	6.57

Table 2: Pearson and Spearman correlations

Table 2 presents the correlation matrix between our measure of misreporting, machine downloads, and key control variables. Pearson (Spearman) correlations appear above (below) the diagonal. Coefficients that are significant at the 10% level or better are in bold. The sample is 44,528 observations for the period $2004 - 2016$. Variable definitions are listed in Appendix B.

	Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
(1)	*Restatement_ Risk*		**0.77**	**0.71**	**-0.18**	**-0.15**	**-0.22**	**-0.08**	**-0.08**	0.00	**-0.11**	**0.06**	0.01	**0.19**	0.00	**-0.07**	**-0.02**	**-0.03**
(2)	*I_ Rrisk1*	**0.42**		**0.71**	**-0.01**	0.00	**-0.05**	**-0.06**	**-0.06**	0.00	**-0.07**	**0.05**	0.00	**0.13**	0.00	**-0.06**	**-0.04**	**-0.06**
(3)	*I_ Rrisk2*	**0.55**	**0.72**		**-0.01**	0.00	**-0.06**	**-0.10**	**-0.08**	0.00	**-0.12**	**0.07**	0.00	**0.19**	-0.01	**-0.09**	**-0.07**	**-0.09**
(4)	*Machine Downloads*	**-0.12**	**-0.01**	**-0.01**		**0.99**	**0.61**	**-0.11**	0.01	0.00	**0.23**	**0.06**	-0.01	**-0.02**	**-0.01**	**0.20**	**0.12**	**0.03**
(5)	*Total Downloads*	**-0.12**	-0.01	**-0.02**	**0.98**		**0.48**	**-0.09**	**0.02**	0.00	**0.30**	**0.07**	**-0.01**	**-0.02**	**-0.01**	**0.26**	**0.15**	**0.07**
(6)	*Machine Downloads %*	**-0.10**	**-0.03**	**-0.02**	**0.38**	**0.24**		**-0.17**	**-0.05**	-0.01	**-0.20**	**-0.02**	0.00	**0.02**	0.00	**-0.17**	**-0.11**	**-0.20**
(7)	*AI Ownership*	**-0.12**	**-0.07**	**-0.12**	**-0.08**	**-0.06**	**-0.18**		**0.15**	**0.02**	**0.34**	-0.01	-0.01	**-0.18**	**-0.03**	**0.33**	**0.53**	**0.28**
(8)	*ROA*	**-0.24**	**-0.11**	**-0.17**	**0.06**	**0.09**	**-0.15**	**0.25**		0.01	**0.22**	**-0.09**	**0.02**	**-0.42**	**-0.05**	**0.13**	**0.17**	**0.08**
(9)	*MTB*	**-0.03**	**-0.04**	**-0.06**	**0.16**	**0.17**	-0.01	**0.18**	**0.29**		**0.03**	-0.01	**0.01**	0.00	0.00	**0.02**	**0.01**	**0.01**
(10)	*Size*	**-0.38**	**-0.11**	**-0.19**	**0.27**	**0.33**	**-0.24**	**0.41**	**0.44**	**0.40**		**0.07**	0.00	**-0.38**	**-0.01**	**0.80**	**0.55**	**0.54**
(11)	*Leverage*	**0.05**	**0.06**	**0.09**	**0.13**	**0.15**	**-0.10**	**0.04**	**-0.04**	**-0.02**	**0.23**		0.00	**0.09**	**0.01**	**0.07**	**0.05**	**0.09**
(12)	*Sales Growth*	**0.04**	-0.01	**-0.01**	**-0.07**	**-0.07**	**-0.04**	**0.08**	**0.21**	**0.23**	**0.14**	**-0.01**		0.00	0.00	-0.01	-0.01	-0.01
(13)	*Loss*	**0.41**	**0.15**	**0.23**	**-0.02**	**-0.04**	**0.05**	**-0.22**	**-0.79**	**-0.09**	**-0.37**	**0.02**	**-0.15**		**0.03**	**-0.21**	**-0.22**	**-0.11**
(14)	*RD*	**0.20**	**0.02**	**0.02**	**0.03**	**0.03**	**0.03**	**-0.02**	**-0.14**	**0.28**	**-0.05**	**-0.21**	**0.06**	**0.30**		0.00	**-0.01**	0.00
(15)	*Analysts Following*	**-0.24**	**-0.09**	**-0.15**	**0.19**	**0.25**	**-0.27**	**0.37**	**0.28**	**0.34**	**0.83**	**0.21**	**0.13**	**-0.20**	**0.05**		**0.57**	**0.51**
(16)	*Institutional Ownership*	**-0.12**	**-0.08**	**-0.13**	**0.18**	**0.19**	**-0.07**	**0.63**	**0.29**	**0.26**	**0.54**	**0.12**	**0.11**	**-0.23**	**0.01**	**0.51**		**0.45**
(17)	*Big4*	**-0.16**	**-0.07**	**-0.13**	**0.11**	**0.14**	**-0.19**	**0.32**	**0.19**	**0.20**	**0.59**	**0.19**	**0.03**	**-0.12**	**0.05**	**0.55**	**0.42**	

Table 3: Panel A: Machine readership and Machine-sensitive Misreporting

This table reports the relationship between machine readership (measured by *Machine Downloads %*) and machine-sensitive misreporting proxies (1) *Restatement_ Risk*, (2) *I_ Rrisk1*, and (3) *I_ Rrisk2*. Variable definitions are listed in Appendix B. All continuous variables except *Restatement_ Risk* and *Machine Downloads %* are winsorized at 1% and 99%. Column (1) presents OLS regression result of *Restatement_ Risk* with firm and year fixed effects. Column (2) presents OLS regression result of *Restatement_ Risk* with industry and year fixed effects. Column (3) presents logistic regression result of *I_ Rrisk1* controlling industry and year fixed effects. Column (4) presents logistic regression result of *I_ Rrisk2* controlling industry and year fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)	(4)
VARIABLES	*Restatement Risk*	*Restatement Risk*	*I_ Rrisk1*	*I_ Rrisk2*
Machine Downloads %	-0.035***	-0.063***	-1.986***	-1.784***
	(0.007)	(0.008)	(0.201)	(0.159)
Total Downloads	0.001	0.001	0.048	0.091***
	(0.001)	(0.002)	(0.040)	(0.031)
ROA	-0.006	-0.015	-0.281	-0.158
	(0.010)	(0.009)	(0.232)	(0.180)
Market-to-Book	0.000	-0.000	0.009	-0.005
	0.000	0.000	(0.010)	(0.008)
Size	-0.001	-0.003***	-0.031	-0.085***
	(0.001)	(0.001)	(0.025)	(0.019)
Leverage	0.025***	0.021***	0.886***	1.141***
	(0.006)	(0.004)	(0.108)	(0.084)
Sales_ Growth	0.005*	0.009***	0.250***	0.271***
	(0.002)	(0.002)	(0.045)	(0.036)
RD	-0.002**	-0.003***	-0.085***	-0.088***
	(0.001)	(0.001)	(0.022)	(0.017)
Loss	0.016***	0.024***	0.807***	0.858***
	(0.002)	(0.003)	(0.065)	(0.051)
Analyst_ Following	0.001	-0.002	-0.137***	-0.118***
	(0.002)	(0.002)	(0.038)	(0.029)
Institutional Ownership	0.012***	0.009**	0.004	0.003
	(0.003)	(0.004)	(0.084)	(0.065)
Big 4	0.002	-0.012***	-0.467***	-0.532***
	(0.004)	(0.002)	(0.056)	(0.043)
Constant	0.067***	0.118***	-1.537***	-0.803**
	(0.013)	(0.011)	(0.431)	(0.348)
Observations	44,528	44,528	44,528	44,528
Adj-R^2/ Pseudo R^2	0.373	0.149	0.088	0.123
Method	OLS	OLS	Logistic	Logistic
Firm FE	Yes	No	No	No
Ind. FE	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Panel B. Alternative machine readership measure: AI Ownership

This table reports the relationship between machine readership (measured by *AI Ownership*) and machine-sensitive misreporting proxies (1) *Restatement_ Risk*, (2) *I_ Rrisk1*, and (3) *I_ Rrisk2*. Variable definitions are listed in Appendix B. All continuous variables except *Restatement_ Risk* are winsorized at 1% and 99%. Column (1) presents OLS regression result of *Restatement_ Risk* with firm and year fixed effects. Column (2) presents OLS regression result of *Restatement_ Risk* with industry and year fixed effects. Column (3) presents logistic regression result of *I_ Rrisk1* controlling industry and year fixed effects. Column (4) presents logistic regression result of *I_ Rrisk2* controlling industry and year fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)	(4)
VARIABLES	*Restatement Risk*	*Restatement Risk*	*I_ Rrisk1*	*I_ Rrisk2*
AI Ownership	-0.021	-0.026*	-2.699**	-3.420***
	(0.011)	(0.012)	(1.114)	(0.888)
Total Downloads	-0.001	-0.000	-0.023	0.054
	(0.001)	(0.001)	(0.062)	(0.049)
ROA	-0.017	-0.003	0.048	0.178
	(0.010)	(0.016)	(0.360)	(0.277)
Market-to-Book	-0.000	-0.000	-0.008	-0.018
	(0.001)	(0.000)	(0.015)	(0.012)
Size	0.000	-0.002*	0.011	-0.073**
	(0.002)	(0.001)	(0.036)	(0.029)
Leverage	0.025**	0.028***	1.269***	1.594***
	(0.007)	(0.005)	(0.156)	(0.125)
Sales_ Growth	0.003	0.005*	0.225***	0.268***
	(0.003)	(0.002)	(0.070)	(0.056)
RD	-0.002*	-0.003***	-0.136***	-0.128***
	(0.001)	(0.001)	(0.037)	(0.027)
Loss	0.010***	0.022***	0.892***	1.026***
	(0.002)	(0.003)	(0.096)	(0.076)
Analyst_ Following	-0.001	-0.003*	-0.213***	-0.214***
	(0.003)	(0.001)	(0.056)	(0.045)
Institutional Ownership	0.010	0.001	-0.105	-0.039
	(0.005)	(0.003)	(0.134)	(0.106)
Big 4	0.001	-0.009**	-0.361***	-0.536***
	(0.004)	(0.002)	(0.087)	(0.066)
Constant	0.037**	0.062***	-2.394***	-1.780***
	(0.011)	(0.007)	(0.635)	(0.510)
Observations	19,698	19,698	19,624	19,624
Adj-R^2/ Pseudo R^2	0.453	0.115	0.097	0.161
Method	OLS	OLS	Logit	Logit
Firm FE	Yes	No	No	No
Ind. FE	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Table 4: Exogenous Variation Using Local AI Talent Supply

This table reports the 2SLS analysis using AI Talent Supply as Instrumented Machine Downloads %. Machine-sensitive misreporting measures include Columns (1) *Restatement_ Risk*, (2) *I_ Rrisk1*, and (3) *I_ Rrisk2*. Instrumented *Machine Downloads %* is the standardized value of predicted Machine Downloads % from the first stage. The sample covers firms over 2011 – 2016 with non-missing financial data. Variable definitions are listed in Appendix B. All continuous variables (excluding Restatement_Risk and Machine Downloads %) are winsorized at 1% and 99%. Control variables include ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, Institutional ownership, and Big4. OLS regressions are estimated with firm, industry, and year fixed effects. Logistic regressions are estimated with year and industry fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)	(4)
VARIABLES	*Machine Downloads %*	*Restatement Risk*	*I_ Rrisk1*	*I_ Rrisk2*
AI Talent Supply	0.001**			
	(0.000)			
Instrumented		-1.261**	-38.947**	-51.310***
Machine Downloads %		(0.396)	(17.202)	(13.707)
Total Downloads		-0.000	-0.050	0.033
		(0.001)	(0.066)	(0.051)
Observations	18,520	18,520	18,452	18,452
Adj-R^2/ Pseudo R^2	0.726	0.114	0.093	0.158
Method	OLS	OLS	Logistic	Logistic
Constant	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	No	No	No
Ind. FE	No	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Table 5: Machine readership and TM-sensitive misreporting

This table reports the regression results of *TM-sensitive misreporting* on machine readership (measured by *Machine Downloads %* and *AI Ownership*). Proxies for *TM-sensitive misreporting* are (1) Discretionary accruals from the modified Jones model (*MJones*), (2) Discretionary accruals from performance-matched modified Jones model (*PM-Jones*), and (3) F-score as a prediction of material misstatements (*F-score*). Variable definitions are listed in Appendix B. All continuous variables except *Machine Downloads %* are winsorized at 1% and 99%. Control variables include *ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, Institutional ownership*, and *Big4*. OLS regressions are estimated with year and firm fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	*MJones*		*PM-Jones*		*F-score*	
Machine Downloads %	-0.004		0.002		0.034	
	(0.005)		(0.004)		(0.026)	
AI Ownership		0.012		0.012		-0.018
		(0.012)		(0.016)		(0.077)
Observations	30,585	13,153	30,511	13,109	30,585	13,153
Adj-R^2/ Pseudo R^2	0.247	0.289	0.109	0.143	0.665	0.714
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 6: Machine readership and Restatements

This table examines and reports the regression results between restatement incidence and machine readership. *Restatement* is an indicator variable that equals to one if a firm filed a restatement through Form $8 - K$ Item 4.02. *Machine Download %* is a proxy for machine readership. The sample covers firms over 2004 – 2016 with non-missing financial data. Variable definitions are listed in Appendix B. Column (1) employs logit regression, and Column (2) and (3) use linear probability model, controling for firm and year fixed effects,and industry and year fixed effects, respectively. All continuous variables are winsorized at 1% and 99%. Control variables include ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, and Institutional ownership. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)
VARIABLES		*Restatement*	
Machine Downloads %	-1.276***	-0.042**	-0.100***
	(0.182)	(0.015)	(0.019)
Total Downloads	-0.128***	-0.011**	-0.010**
	(0.036)	(0.004)	(0.004)
Observations	44,528	44,528	44,528
Adj-R^2/ Pseudo R^2	0.029	0.073	0.015
Method	Logit	OLS	OLS
Constant	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Firm FE	No	Yes	No
Ind. FE	Yes	No	Yes
Year FE	Yes	Yes	Yes

Table 7: Machine readership advantage: financial statement complexity

This table reports the relationship between our measure of misreporting and machine readership: between *Machine Downloads %* and machine-sensitive misreporting proxies (1) *Restatement_Risk*, (2) *I_Rrisk1*, and (3) *I_Rrisk2*. Variable definitions are listed in Appendix B. All continuous variables (excluding Restatement_Risk and Machine Downloads %) are winsorized at 1% and 99%. Control variables include *ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, Institutional ownership,* and *Big4*. Regressions are estimated with industry and year fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)
VARIABLES	*Restatement Risk*	*I_Rrisk1*	*I_Rrisk2*
Machine Downloads %	-0.023**	-1.612***	-1.520***
	(0.008)	(0.244)	(0.190)
FS Complexity	0.026***	0.860***	0.689***
	(0.005)	(0.165)	(0.130)
Machine Downloads %	-0.021***	-0.472**	-0.282
× *FS Complexity*	(0.006)	(0.229)	(0.179)
Total Downloads	0.001	0.038	0.083***
	(0.001)	(0.041)	(0.032)
Observations	44,052	44,052	44,052
Adj-R^2/ Pseudo R^2	0.376	0.095	0.129
Method	OLS	Logistic	Logistic
Constant	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Ind. FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Table 8: Machine readership advantage: Alternative Data Coverage

This table reports the relationship between satellite data coverage and machine-sensitive misreporting: between *Post_ Coverage*, *Alt_ Data_ Covered*, and machine-sensitive misreporting proxies *Restatement_ Risk*. Variable definitions are listed in Appendix B. All continuous variables (excluding *Restatement_ Risk* and *Machine Downloads %*) are winsorized at 1% and 99%. Control variables include *ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, Institutional ownership,* and *Big4*. Column (1) and (2) present OLS regression results with firm and year fixed effects. Column (3) and (4) present OLS regression results with industry and year fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)	(4)
VARIABLES	*Restatement_ Risk*			
Machine Downloads %	-0.035***	-0.035***	-0.063***	-0.064***
	(0.007)	(0.005)	(0.008)	(0.004)
Post_ Coverage		-0.000		-0.003
		(0.003)		(0.003)
Post_ Coverage × *Alt_ Data_ Covered*		-0.013***		-0.009*
		(0.004)		(0.005)
Observations	44,528	44,528	44,528	44,528
Adj-R^2	0.373	0.373	0.149	0.150
Method	OLS	OLS	OLS	OLS
Constant	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	No	No
Ind. FE	No	No	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Table 9: AI-enabled detection: increased misreporting cost from peer restatement

This table reports whether firms with *Peer Restate* displays a different relationship between machine readership and machine-sensitive misreporting. between Machine Downloads % and misreporting proxies (1) *Restatement_Risk*, (2) *I_Rrisk1*, and (3) *I_Rrisk2*. The sample covers firms over 2004 − 2016 with non-missing financial data and partitioned based on the indicator variable *Peer Restate*. Variable definitions are listed in Appendix B. All continuous variables are winsorized at 1% and 99% except *Restatement_Risk* and *Machine Downloads %*. Control variables include *ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, Institutional ownership*, and *Big4*. Regressions are estimated with industry and year fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)
VARIABLES	*Restatement Risk*	*I_Rrisk1*	*I_Rrisk2*
Machine Downloads %	-0.033***	-1.090***	-1.017***
	(0.007)	(0.301)	(0.227)
Peer Restate	0.037***	1.085***	0.946***
	(0.006)	(0.202)	(0.151)
Machine Downloads %	-0.040***	-1.081***	-0.955***
× *Peer Restate*	(0.009)	(0.278)	(0.207)
Total Downloads	0.001	0.051	0.092***
	(0.002)	(0.040)	(0.031)
Observations	44,528	44,528	44,528
Adj-R^2/ Pseudo R^2	0.152	0.091	0.125
Method	OLS	Logistic	Logistic
Constant	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Ind. FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Table 10: AI-enabled detection: High-litigation industry

This table reports the relationship between our measure of misreporting and machine readership: between *Machine Downloads %* and misreporting proxies (1) *Restatement_Risk*, (2) *I_Rrisk1*, and (3) *I_Rrisk2*. The sample covers firms over 2004 – 2016 with non-missing financial data and partitioned based on the indicator variable *Litigate*. Variable definitions are listed in Appendix B. All continuous variables (excluding Restatement_Risk and Machine Downloads %) are winsorized at 1% and 99%. Control variables include *ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, Institutional ownership*, and *Big4*. Regressions are estimated with industry and year fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)	(3)
VARIABLES	*Restatement Risk*	*I_Rrisk1*	*I_Rrisk2*
Machine Downloads %	-0.052***	-1.843***	-1.548***
	(0.007)	(0.212)	(0.168)
Litigate	0.042***	0.462**	0.679***
	(0.013)	(0.193)	(0.154)
Machine Downloads %	-0.051***	-0.524**	-0.890***
× *Litigate*	(0.016)	(0.254)	(0.202)
Total Downloads	0.001	0.048	0.091***
	(0.002)	(0.040)	(0.031)
Observations	44,528	44,528	44,528
Adj-R^2/ Pseudo R^2	0.152	0.088	0.124
Method	OLS	Logistic	Logistic
Constant	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Ind. FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Table 11: Alternative measures

This table reports the relationship between our measure of misreporting and machine readership: between Machine Downloads % (alternative definition) and misreporting proxies (1) *Restatement_ Risk*, (2) *I_ Rrisk1*, and (3) *I_ Rrisk2*. We define *Machine Downloads %_ 14Days (30Days)* similarly to the original *Machine Downloads %* variable, except that we aggregate machine-generated requests and total requests within 14 (30) days, respectively. The sample covers firms over $2004-2016$ with non-missing financial data. Variable definitions are listed in Appendix B. All continuous variables are winsorized at 1% and 99% except *Restatement_ Risk* and *Machine Downloads %*. Control variables include *ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, Institutional ownership*, and *Big4*. OLS regressions are estimated with firm (industry) and year fixed effects, and logistic regressions are estimated with industry and year fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

Panel A: Alternative measures of machine download

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	*Restatement_ Risk*		*Restatement_ Risk*		*I_ Rrisk1*		*I_ Rrisk2*	
Machine Downloads %_ 14Days	-0.041***		-0.068***		-2.225***		-1.951***	
	(0.007)		(0.008)		(0.202)		(0.161)	
Total Downloads_ 14Days	0.002		0.002		0.077*		0.125***	
	(0.002)		(0.002)		(0.041)		(0.032)	
Machine Downloads %_ 30Days		-0.041***		-0.068***		-2.308***		-1.932***
		(0.007)		(0.008)		(0.213)		(0.169)
Total Downloads_ 30Days		0.004*		0.003		0.123***		0.181***
		(0.002)		(0.002)		(0.044)		(0.034)
Observations	44,528	44,528	44,528	44,528	44,528	44,528	44,528	44,528
Adj-R^2/ Pseudo R^2	0.374	0.374	0.151	0.151	0.090	0.090	0.124	0.125
Method	OLS	OLS	OLS	OLS	Logistic	Logistic	Logistic	Logistic
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	No	No	No	No	No	No
Ind. FE	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Alternative measures of machine-sensitive misreporting

This table reports the relationship between machine readership and machine-sensitive misreporting proxies (1) *Fraud Score*, We define *Fraud Score* following the machine learning method in Bao, Ke, Li, Yu, and Zhang (2020). The sample covers firms over 2004 – 2016 with non-missing financial data. Variable definitions are listed in Appendix B. All continuous variables are winsorized at 1% and 99% except *Fraud Score* and *Machine Downloads %*. Control variables include *ROA, Market-to-Book, Size, Leverage, Sales growth, RD, Loss, Analysts following, Institutional ownership*, and *Big4*. Column (1) controls for firm and year fixed effects, and Column (2) controls for industry and year fixed effects. ***, **, and * represent statistical significance at 1%, 5%, and 10%. The standard errors, clustered by firm and year, are reported in parentheses.

	(1)	(2)
VARIABLES	*Fraud Score*	*Fraud Score*
Machine Downloads %	-0.144*	-0.188*
	(0.076)	(0.095)
Total Downloads	0.004	-0.013
	(0.019)	(0.017)
Observations	30,822	31,203
Adj-R^2	0.706	0.472
Method	OLS	OLS
Constant	Yes	Yes
Controls	Yes	Yes
Firm FE	Yes	No
Ind. FE	No	Yes
Year FE	Yes	Yes