The use of AI in 10-K Filings

Marcelo S. Perlin[†, a]
Cristian R. Foguesatto[‡, a]
Aliki K. Galanos[*, a]
Felipe Affonso[**, a]

[a]Federal University of Rio Grande do Sul (UFRGS)

Abstract With the rise of AI use in corporate settings and the associated risks to disclosure quality and credibility, we examine whether organizations rely on AI-assisted writing in their annual reports. Using Desklib's Large Language Model and FinBERT, we estimate the probability of AI-generated content and analyze sentiment across three sections of 10-K reports from S&P 500 firms between 2018 and 2024. Results indicate a relatively stable use of AI over time, with distinct patterns across sections, with the Management Discussion and Analysis showing the highest probability of AI adoption. Younger and smaller firms are more likely to rely on AI-assisted writing. We also find a strong positive association between sentiment scores and AI detection, particularly in the Risk Factors section. These findings provide insights for researchers and important implications for investors and policymakers regarding this emerging organizational practice.

Keywords: Corporate Finance; 10-K Filings; Large Language Models (LLMs); Artificial Intelligence

JEL codes: G30, C80.

1. Introduction

The use of Large-Language Models (LLMs) is emerging as a powerful tool for textual analysis, with its human-like ability to generate reports (Kok, 2025), gaining significant relevance in financial reporting (Huang et al., 2023; Yang et al., 2024). Given the importance of corporate disclosure and governance, the reliance on these reports implies that firms should be wary of language selection, text preparation, length and readability (Hasan, 2020; Nguyen and Kimura, 2023). However, with ongoing technological advances,

How to cite: Perlin, M. S., Foguesatto, C. R., Galanos, A. K. & Affonso, F. (2026). The use of AI in 10-K Filings. *Revista Brasileira de Finanças, xxx*, xxx. https://doi.org/10.xxxx/xxxx
Submitted on February 24, 2026. Published online on xxx xxx, 2026.
Editor in charge: Mr. Editor.
† marceloperlin@gmail.com
‡ cristian.foguesatto@ufrgs.br
* aliki.g.kara@gmail.com
** felipe.affonso2@gmail.com

it is expected that LLMs will support professionals in generating reported information (Türegün, 2019; Csere et al., 2025) and possibly assist in writing (De Villiers et al., 2024) them. Thus far, researchers have provided evidence on the effects of AI strategies and integration on different firm outcomes such as, for example, corporate investment efficiency (Liu and Hu, 2025) and ESG performance (Weng, 2025). Moreover, the disclosure of automated decision-making has also been examined (Bonsón et al., 2023). However, as far as we are concerned, there is no study on the use of AI to assist managers in writing financial reports.

Although the implications of such assistance can be advantageous with professionals having more time to focus on decision making over data collection and manual analysis (Türegün, 2019), it could also bias reports by compromising their transparency and accountability (De Villiers et al., 2024). The use of LLMs can increase the risk of artificial and inconsistent information, with a direct impact on the quality of financial reports, and also on the governance and transparency of the information. Therefore, we aim to examine whether organizations use AI-assisted writing in their annual reports, covering three main section of the 10-K: Risk Factors, Management Discussion and Analysis (MD&A) and Business Description. These sections are relevant because provide a set of key companies' information. In addition, we investigate which firms are more likely to do so. In doing so, we extend the existing literature on the use of AI in the organizational spectrum and further discuss the implications of its use. Moreover, by providing evidence of the disclosure quality from a different perspective (AI use) and the possible indicatives of which companies might compromise their reports, this study opens new research avenues on LLM-assisted writing in Annual Reports.

2. Data and Methods

We collect 10-K reports and financial/price data for all firms that were part of the S&P500 index from 2018 to 2024. This time frame is particularly relevant as it coincides with the widespread adoption of AI-based natural language processing tools in corporate reporting. For the 10-K reports, we use EdgarTools, a Python library for downloading company data directly from the official SEC API. The S&P500 composition and all financial data is obtained from EODHD, a commercial data vendor and a common source of information for academic research.

Within each downloaded 10-K document, we save the text of three distinct sections that are later used for the analysis: business description, risk factors (de Jong, 2025), and management discussion of results (Fengler and Phan, 2025). Each of this section has a different format, length and purpose, which allows our study to present a more complete analysis in the use of AI for text generation in financial reports.

With the text data, we use Desklib's AI Text Detector (version 1.01) (DeskLib, 2024), a large language model that outputs the probability of AI use in writing text. We chose to use Desklib given is high accuracy in the RAID Benchmark Leaderboard (access in 2026-01-15) (Dugan et al., 2024) and also its free access to the public, greatly facilitating the reproduction of this study. Additionally, we compute the sentiment of all 10-K sections using FinBERT (Huang et al., 2023), a large language model for extracting sentiment from financial texts.

One caveat of using Desklib and FinBERT in analyzing large financial text is their maximum input tokens (5000 for Desklib and 512 for FinBert). We find that, for our dataset, this limit was significantly lower than the total number of tokens in the majority of the texts. To solve this issue, our estimation strategy was to use a sliding window and split large texts into smaller parts with an overlap of 10% of tokens for each split (Wang et al., 2019; Ding et al., 2020). For Desklib, we use a text chunk size of 2.500 tokens (overlap of 250 tokens) and for FinBert we use 500 as the chunk size and 50 as overlap. We then aggregate the results of AI detection by taking an average of all chunk probabilities. For the sentiment, we first ignore all *neutral* sentiment cases and, each chunk of text, assign +1 for positive sentiments and -1 for negative. Finally, we average the results for a final sentiment score for each of the 10-K sections.

After cleaning for missing data, our text and financial data consists of 3939 data points for 570 companies during the time period between 2018 and 2024. Table 1 presents the descriptive statistics for our final sample: Panel A details the company's financial and market characteristics, and Panel B provides statistics for the text analysis variables, including the AI detection scores and sentiment scores derived from the 10-K reports.

In Table 1, Panel A, we observe significant variability in performance, as indicated by the high standard deviation and wide range of Return on Equity (ROE). We can also see that the median IPO year is 1994, that is, most are well established companies. The average firm in the sample (Age) is 26.551 years old. Moreover, it is observed that firms do not considerably differ in terms of size. However, there are significant gaps among companies in terms of leverage and returns, as suggested by the means, minimum, and maximum values of these variables.

Panel B of Table 1 shows that the *Management Discussion of Results* section (MD Score) has the highest average probability of AI use at 0.581. This is substantially higher than the scores for the *Business Description* (BD Score) at 0.436 and *Risk Factors* (RF Score) at 0.411. The sentiment scores also behave as expected by content: *Risk Factors* (RF Sentiment) are almost uniformly negative, with a mean of -0.842. In contrast, *Business Description* (BD Sentiment) is positive on average (0.288), while *Management Discussion* (MD Sentiment) is slightly negative at -0.181.

Table 1
Descriptive Statistics of Data (2018-2024)

Variable	N	Mean	Std Dev	Min	Max
Panel A: Company Variables					
ROE	3,939	0.178	0.722	-3.095	4.049
ROA	3,939	0.065	0.080	-0.217	0.311
IPO Year*	3,799	1993	15	1919	2024
Age	3,799	26.6	14.8	-6.0	105.0
Log Size	3,939	10.347	0.592	9.083	12.028
Leverage	3,939	0.656	0.233	0.130	1.465
Beta	3,880	0.969	0.411	0.042	2.109
Asset Turnover	3,939	0.662	0.559	0.035	3.159
Accrual Ratio	3,939	0.006	0.056	-0.174	0.223
Panel B: Text Analysis Variables					
BD Score	2,847	0.436	0.178	0.030	0.979
RF Score	2,900	0.411	0.160	0.063	0.977
MD Score	2,918	0.581	0.155	0.085	1.000
BD Sentiment	3,362	0.288	0.623	-1.000	1.000
RF Sentiment	3,362	-0.842	0.352	-1.000	0.000
MD Sentiment	3,362	-0.181	0.420	-1.000	1.000

* Mean value for IPO Year is the median.

2.1 Models

We use econometric models to investigate the profile of companies that are mostly using AI in their written financial reports. For that, we use the probabilities of AI written text from Desklib AI Text Detector (DeskLib, 2024) as the dependent variable. Given its truncated nature ([0 - 1]) and panel format, we estimate a Panel Generalized Linear Model (PGLM) with random effects

Table 2
Description and Calculation of Explanatory Variables

Variable	Calculation	Hypothesis (sign)
Panel A: Hypothesis Variables		
ROE	Total Net Income divided by Total Equity	Positive
Age	Number of years since IPO date	Negative
Risk (Beta)	Beta from market model, calculated using SP500 returns	Positive
Leverage	Total debt divided by total assets	Positive
Sentiment	Sentiment value of text, based on FinBert	Positive
Asset Turnover	Total Revenue divided by Total Assets	Positive
Accrual Ratio	Free cash flow minus net income, divided by total assets	Positive
Panel B: Control Variables		
Sector	Sector of company	-
Year	Year of data	-

and quasibinomial distribution to model the data[1], Equation 1. [2]

$$E(AI_{i,t}) = f(z_{i,t}) \tag{1}$$

$$pAI_{i,t} = \alpha + \theta X_{i,t} \tag{2}$$

where pAI_i is the dependent variable, measuring the probability of AI use in text generation, α is the intercept, and the θ is a vector of parameters attached to $X_{i,t}$. As for the explanatory variables, we follow the work of Fengler and Phan (2025) and set our hypothesis and control variables as show in Table 2.

Following Table 2, we expect a positive relationship between Return on Equity (ROE) and AI usage, suggesting that more profitable firms are more likely to adopt AI-assisted writing (Shiyyab et al., 2023). High-performing companies typically possess greater financial resources, allowing them to invest in emerging technologies and absorb the costs associated with implementing new software systems. Furthermore, managers of profitable firms are often incentivized to adopt cutting-edge tools to signal efficiency and innovation to shareholders, maintaining a reputation for operational excellence.

[1] We thank the authors of R package **fixest** (Bergé, 2018), which was used in all estimations.

[2] Alternatively, we also estimate a vanilla probit model with binomial distributions and dummies for sector and year. The results are very comparable to the ones provided here. To save space, we omit the results of this estimations, but are available upon request.

As for Age, we believe in a negative relationship with AI usage, implying that younger firms are more prone to using AI than their older counterparts. We suspect that established corporations often have deeply entrenched manual reporting processes and bureaucratic layers that resist technological change. Conversely, younger companies can be viewed as more agile and innovative (Protogerou et al., 2017), often lacking legacy systems, which makes them more open to experimenting with and integrating generative AI tools into their financial reporting workflows.

Both Risk (Beta) and Leverage carry a positive hypothesis, suggesting that firms facing higher market volatility or debt burdens rely more heavily on AI. Companies with high leverage or market risk face intense scrutiny from creditors and investors, requiring extensive and careful disclosure of risk factors (Zhu et al., 2024). The use of AI in these scenarios is justified as a mechanism to reduce the high administrative costs of drafting these mandatory, voluminous, and legally sensitive disclosures, allowing managers to meet transparency demands efficiently. Likewise, we expect Asset Turnover to be positively associated with the use of AI in corporate reports. Firms that employ their assets more efficiently tend to place greater value on operational scalability and process optimization. Prior research shows that more efficient firms are better positioned to benefit from information technologies, as automation helps reduce administrative frictions and managerial overload by reallocating routine tasks away from senior management (Brynjolfsson et al., 2017).

The variables for Sentiment and Accrual Ratio are both hypothesized to be positive. For Accrual Ratio, higher accruals often indicate complex accounting estimates that require detailed technical explanations; firms may use LLMs to help generate this dense, technical text more quickly. Regarding Sentiment, the positive link suggests that AI tools may be used to polish the report of positive reports. That is, managers might use the technology's tendency to produce smoother, more professional, and generally more optimistic or neutral-positive corporate prose compared to raw human drafts (Huang et al., 2023).

3. Results

3.1 Descriptive Statistics

In Figure 1 we see that the average AI score across all three sections (MD&A, Business Description, Risk Factors) have remained relatively steady from 2018 to 2024. This indicates that AI-assisted writing has not seen a dramatic, widespread increase in these specific report sections during this period. Nevertheless, a small peak is observed in 2021 for all three sections.

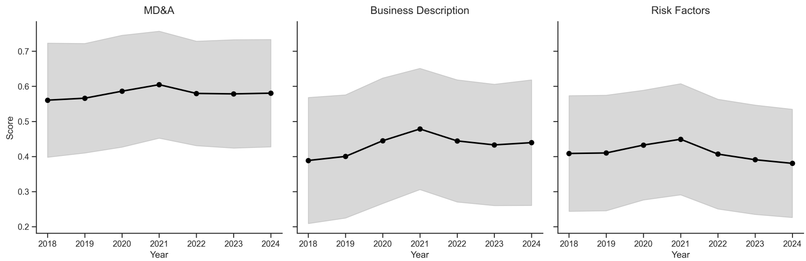

Figure 1
Average AI Score per 10-K text segment - per year

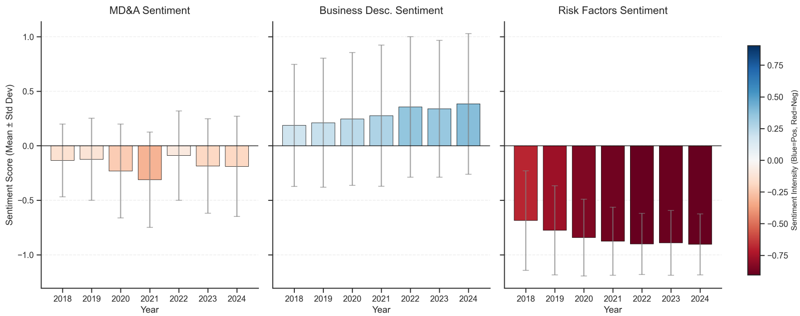

Figure 2
Average Sentiment Score per 10-K text segment

AI came into use and spread around the end of 2022 with the release of ChatGPT. Moreover, the variation around the mean (reflected by the standard error) suggests that in 2024 the use of AI rose in the MD&A and Business Description sections. However, this growth in the variation also means that while the average firm did not change its AI-assisted writing level, some others did, reflecting in the slight divergence between firms in our sample. Overall, based in Figure 1, we find that the MD&A section is consistently presenting higher AI usage than the other two.

In Figure 2 we present the average sentiment per document and year. First, we see that, as expected, the average Risk Factor sentiment is far lower than for the MD&A and Business Description. This is justified by the purpose of this sections, which is to alert potential investors about all risks of the company. Likewise, Business Description presents a more positive tone, which is also explained by its nature, to convince possible investors about the feasibility of the business, usually on a suggestive and positive tone.

Table 3
Estimation Results for GLM Model (Equation 2)

	Business Description	Risk Factors	Management Discussion
ROE	0.001 (0.010)	-0.035 (0.016)**	0.005 (0.009)
Age	-0.016 (0.003)***	-0.017 (0.004)***	-0.005 (0.003)
Log Size	-0.158 (0.102)	-0.064 (0.122)	0.000 (0.102)
Risk (Beta)	0.188 (0.125)	-0.053 (0.148)	-0.185 (0.125)
Leverage	0.273 (0.164)*	-0.632 (0.221)***	0.418 (0.203)**
Sentiment	0.263 (0.071)***	7.071 (0.972)***	0.161 (0.099)
Asset Turnover	-0.445 (0.111)***	-0.438 (0.141)***	-0.008 (0.093)
Accrual Ratio	0.109 (0.964)	2.270 (1.149)**	1.622 (0.988)
Year effect	Yes	Yes	Yes
Sector dummies	Yes	Yes	Yes
Num. obs.	2833	2833	2833

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

3.2 Econometric Results

Table **??** presents the results from the estimation of Equation 2 for each of the 10-K sections. First, looking at parameter *Age*, we find that established, older firms are significantly less likely to utilize AI for writing their 10-Ks, as evidenced by the negative coefficients for Age, which are statistically significant for the *Business Description* and *Management Discussion* sections (p<0.01).

Similarly, the coefficient for (*Log Size*) is negative in all three cases, but statistically significant in only one (*Risk Factor*). This mildly suggests that larger corporations, which often face more complex legal scrutiny, may prefer manual human drafting over AI automation for disclosing risks. Furthermore, the coefficient for *Risk (Beta)* is also negative for two out of the three sections (*Business Description* and *Risk Factors*). This is unlike what we expected, with high market volatility firms being less likely to count on AI assisted-writing in their reports.

Additionally, we also find that the coefficient for *Sentiment* is a strong positive predictor of AI use. That is, reports with more positive sentiment scores show a significantly higher probability of AI generation in the *Business Description* and *Risk Factors* sections, with the latter showing a particularly large coefficient with respect to others. Furthermore, *Asset Turnover* has a significantly positive relationship with the probability of AI use in the *Management Discussion* section. This suggests that more effective companies are more likely to AI assisted-writing.

4. Discussion of Results

Our empirical analysis reveals that, while the aggregate adoption of AI in S&P 500 annual reports has remained relatively stable between 2018 and 2024, there is a distinct heterogeneity in usage across different report sections. The Management Discussion and Analysis (MD&A) section consistently exhibits the highest probability of AI-assisted writing. This finding aligns with the functional strengths of Large Language Models (LLMs) in summarizing complex financial performance data into coherent narratives (Kok, 2025).

Regarding firm characteristics, the negative and statistically significant relationship between firm age and AI utilization supports the structural inertia hypothesis (Liu and Hu, 2025). Older, established firms appear more entrenched in legacy reporting processes and exhibit resistance to integrating generative AI into their disclosure workflows. In contrast, younger firms, without layers of legacy IT systems, demonstrate higher agility in adopting these tools. This corroborates the notion that newer market entrants are more likely to leverage emerging technologies to gain operational efficiencies and signal innovation to stakeholders, a trend also observed in corporate investment efficiency contexts (Liu and Hu, 2025).

Contrary to the expectation that resource-rich firms would pioneer AI adoption, our results regarding firm size indicate a conservative approach among larger corporations, particularly in the drafting of *Risk Factors*. The negative correlation suggests that large-cap companies, which face intense regulatory scrutiny and higher litigation risks, prefer human oversight over automated generation for legally sensitive disclosures. This highlights a critical boundary in current AI assistance: while efficiency is desirable, the potential for inaccuracies or generic disclosures in risk reporting (De Villiers et al., 2024) likely deters full automation in high-stakes areas where precise legal language is paramount.

A critical finding of this study is the strong positive association between sentiment scores and AI probability, particularly within the Business Description and Risk Factors sections. This implies that managers may be utilizing LLMs not merely for text generation, but as a strategic mechanism for impression management to "polish" the tone of disclosures. By leveraging the tendency of LLMs to produce neutral-to-positive professional prose (Huang et al., 2023), firms might inadvertently or intentionally soften the language of negative disclosures. This raises concerns that AI-assisted writing could compromise the transparency of 10-K reports by masking underlying risks with synthetically smoothed narratives.

Collectively, these results portray a nuanced landscape of AI integration in financial reporting. While widespread, unchecked automation is not yet the norm, specific subsets of firms—primarily younger and more agile entities—are leveraging these tools to reshape the narrative tone and efficiency of their filings. The strategic distinction in usage across report sections suggests that managers are deploying AI where summarization is needed (MD&A) while exercising caution where legal precision is critical. As noted by (Yang et al., 2024), this evolution necessitates that auditors and regulators adapt their monitoring frameworks to effectively detect and evaluate machine-generated nuances in financial disclosures.

5. Conclusion

This study aimed to examine the prevalence and determinants of AI-assisted writing within the annual reports of S&P 500 companies. We use Desklib's large language model to detect probabilities of AI-generated content and FinBERT to assess sentiment across three key sections: MD&A, Business Description, and Risk Factors. This approach allowed us to extend existing literature by providing empirical evidence on how organizations are currently integrating LLMs into their financial reporting workflows.

Our descriptive analysis reveals that, while the average probability of AI use has remained relatively stable over the sample period, distinct patterns have emerged regarding specific report sections. The MD&A section consistently displayed the highest probability of AI utilization compared to Business Descriptions and Risk Factors. In the econometric side, our results identify specific firm characteristics that drive this adoption. We found a significant negative relationship between firm age and AI usage, supporting the hypothesis that younger, more agile firms are more prone to using AI than established, older corporations which may be bound by structural inertia. Additionally, we observed a strong positive association between sentiment scores and AI detection, particularly within the Risk Factors section. This suggests that managers may be leveraging LLMs to refine the tone of their reports, utilizing the technology to produce smoother, more professional, and potentially more optimistic prose.

These findings have significant implications for investors and regulators concerned with disclosure quality. The correlation between AI assistance and positive sentiment raises the critical question of whether LLMs are being used to enhance readability or to obscure material risks, potentially aggravating agency problems. As technological advancements continue to make these tools more accessible, future research should investigate the qualitative impact of AI-generated text on information asymmetry and whether the polishing effect of AI compromises the transparency and accountability essential to financial reporting. Future research could also extend our research by investigating the difference between large and small firms eliminating the bias we are not free from when investigating S&P 500, which are typically large and subject to strict inclusion criteria.

References

Bergé, L. (2018). Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm, *CREA Discussion Papers*, (13).

Bonsón, E., Bednárová, M. and Perea, D. (2023). Disclosures about algorithmic decision making in the corporate reports of western european companies, *International Journal of Accounting Information Systems*, *48*, 100596.

Brynjolfsson, E., Rock, D. and Syverson, C. (2017). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics, *Technical report*, National Bureau of Economic Research.

Csere, P., Estep, C., Griffith, E. E. and MacKenzie, N. L. (2025). Financial executives' responses to the use of artificial intelligence in financial reporting and auditing, *Current Issues in Auditing*, pp. 1–9.

de Jong, P. (2025). Do risk warnings predict capital raising? evidence from 10-k disclosures and subsequent 8-k filings, *Finance Research Letters*, p. 108000.

De Villiers, C., Dimes, R. and Molinari, M. (2024). How will ai text generation and processing impact sustainability reporting? critical analysis, a conceptual framework and avenues for future research, *Sustainability Accounting, Management and Policy Journal*, *15*(1), 96–118.

DeskLib (2024). Ai text detector v1.01, https://huggingface.co/desklib/ai-text-detector-v1.01.

Ding, M., Zhou, C., Yang, H. and Tang, J. (2020). Cogltx: Applying bert to long texts, *Advances in Neural Information Processing Systems*, *33*, 12792–12804.

Dugan, L., Hwang, A., Trhlík, F., Zhu, A., Ludan, J. M., Xu, H., Ippolito, D. and Callison-Burch, C. (2024). Raid: A shared benchmark for robust evaluation of machine-generated text detectors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 12463–12492.

Fengler, M. R. and Phan, T. M. (2025). Unveiling themes in 10-k disclosures: A new topic modeling perspective, *International Review of Financial Analysis*, *103*, 104121.

Hasan, M. M. (2020). Readability of narrative disclosures in 10-k reports: does managerial ability matter?, *European Accounting Review*, *29*(1), 147–168.

Huang, A. H., Wang, H. and Yang, Y. (2023). Finbert: A large language model for extracting information from financial text, *Contemporary Accounting Research*, *40*(2), 806–841.

Kok, T. (2025). Chatgpt for textual analysis? how to use generative llms in accounting research, *Management Science*, .

Liu, Y. and Hu, R. (2025). Artificial intelligence, supply chain finance, and corporate investment efficiency, *Finance Research Letters*, p. 107851.

Nguyen, P. T. T. and Kimura, A. (2023). Disclosure characteristics of annual reports and information asymmetry: Evidence from foreign firms listed on the us stock exchange, *Finance Research Letters*, *54*, 103776.

Protogerou, A., Caloghirou, Y. and Vonortas, N. S. (2017). Determinants of young firms' innovative performance: Empirical evidence from europe, *Research Policy*, *46*(7), 1312–1326.
URL: https://www.sciencedirect.com/science/article/pii/S004873331730094X

Shiyyab, F. S., Alzoubi, A. B., Obidat, Q. M. and Alshurafat, H. (2023). The impact of artificial intelligence disclosure on financial performance, *International Journal of Financial Studies*, *11*(3).
URL: https://www.mdpi.com/2227-7072/11/3/115

Türegün, N. (2019). Impact of technology in financial reporting: The case of amazon go, *Journal of Corporate Accounting & Finance*, *30*(3), 90–95.

Wang, Z., Ng, P., Ma, X., Nallapati, R. and Xiang, B. (2019). Multi-passage bert: A globally normalized bert model for open-domain question answering, *arXiv preprint arXiv:1908.08167*, .

Weng, M. (2025). Green innovation through artificial intelligence technology: Enhancing environmental, social, and governance performance, *Finance Research Letters*, *75*, 106921.

Yang, X., Zang, S., Ren, Y., Peng, D. and Wen, Z. (2024). Evaluating large language models on financial report summarization: An empirical study, *arXiv preprint arXiv:2411.06852*, .

Zhu, X., Li, J. and Wang, Y. (2024). Are risk disclosures in financial reports informative? a text mining-based perspective, *Humanities and Social Sciences Communications*, *11*(1), 1–18.