

Progress in Applying Artificial Intelligence to ESG Data Acquisition and Financial Analysis

Xinyue Cheng

*International College, Jiangxi University of Finance and Economics, Nanchang, China
2213089930@qq.com*

Abstract. With the growing demand for sustainable finance and investment decision-making, the role of artificial intelligence (AI) in environmental, social, and governance (ESG) data collection and analysis has attracted increasing scholarly attention. Nevertheless, challenges such as fragmented data sources and inconsistent standards continue to limit its application. Against this backdrop, this study reviews prior research to outline recent progress in ESG data collection, risk identification, and financial analysis, while also evaluating AI's role in improving data reliability, guiding investment strategies, and strengthening risk control. The results reveal that techniques such as natural language processing (NLP), machine learning, and generative AI can strengthen ESG data collection and analytical capacity, with practical applications in risk identification, portfolio optimization, and asset pricing. Yet, practical applications are constrained by issues such as inconsistent data standards, algorithmic bias, and the lack of model interpretability. Future research should prioritize the establishment of unified data standards, the enhancement of algorithmic fairness and interpretability, and the advancement of multimodal data integration and cross-regional comparisons to develop a more intelligent and reliable ESG analytical framework.

Keywords: Artificial Intelligence (AI), ESG Data Collection, Financial Analysis, Machine Learning, Sustainable Finance

1. Introduction

Motivated by global carbon peaking and carbon neutrality targets, along with broader sustainable development initiatives, Environmental, Social, and Governance (ESG) factors have emerged as a central issue in the financial sector. The ESG performance of corporations influences market value, financing expenses, risk exposure, and sustainable returns. However, ESG data remain fragmented and complex, with limited transparency and inconsistent disclosure. Despite growing attention to ESG reporting, methodological discrepancies and selective corporate disclosure undermine the data credibility and comparability. In response, artificial intelligence (AI) offers new solutions to these challenges. Relevant technologies have been applied to annual reports, news, and public opinion analysis, demonstrating advantages in information extraction, risk identification, and rating, while also enhancing data quality and usability. Yet, current research has paid limited attention to the role of AI in investment optimization and risk management, and there remains need for improvement in algorithmic fairness and interpretability. Accordingly, through a review of the literature, this paper

maps the development of AI in ESG data collection, risk identification, and financial analysis. It investigates AI's role in improving data quality, aiding investment decision-making, and enhancing risk management, while highlighting challenges related to data standardization, regulatory oversight, and ethical governance. The study seeks to elucidate AI's applications and value in ESG, examine key challenges and improvements in the Chinese context, and provide references for policymakers and financial institutions to optimize ESG practices.

2. ESG data features and acquisition methods

2.1. ESG data structure and classification specification

ESG rating systems are composed of three key dimensions, namely Environmental (E), Social (S), and Governance (G), with scores ranging from 0 to 100, where higher scores reflect better corporate ESG performance [1]. The environmental dimension includes metrics like carbon emissions, energy usage, and pollution control, while the social dimension addresses employee rights, supply chain management, and community engagement, and the governance dimension covers board diversity, business ethics, and risk management [2]. However, variations in indicator selection and weighting among rating agencies can lead to discrepancies in a company's ratings across different systems. In addition, ESG data are primarily sourced from corporate annual reports, sustainability reports, news articles, and third-party rating agencies, encompassing both financial and extensive non-financial information [3].

2.2. Data collection challenges and quality assurance

The collection of ESG data experiences multiple challenges, like fragmented data sources, limited transparency, the absence of standardized disclosure criteria, and inconsistencies in ratings across different evaluation agencies. In China, ESG disclosure regulations are still being gradually refined, requiring regulators and investors to rely on diverse information sources for data acquisition and decision-making applications [4]. To address these issues, generative AI can improve the robustness of data processing and anomaly detection by implementing dynamic defense mechanisms, including adversarial sample detection and multi-agent verification [5]. Besides, natural language processing (NLP) and semantic analysis can extract unstructured information from corporate annual reports, sustainability reports, news articles, and social media, facilitating the identification of key indicators, monitoring of risk events, and analysis of trends [6]. Moreover, blockchain technology can establish traceable databases to record data provenance and track changes, thus improving data transparency and reliability. By integrating these approaches, the accuracy, reliability, and comparability of ESG data can be strengthened. This provides a solid foundation for evaluating corporate sustainability performance, supporting investment decisions and policy analysis, and facilitating multi-source data integration and cross-institutional comparisons.

2.3. Artificial Intelligence methods and technical practices

AI demonstrates clear advantages in ESG data processing, risk identification, and score automation. By applying spatial Durbin models, it is possible to examine the spatial spillover effects of AI on regional economic sustainability and reveal its capability in integrating and analyzing multi-source data. In practice, machine learning algorithms are widely used to analyze unstructured text, such as corporate annual reports, news stories, and social media content, thus enabling the identification of environmental, social, and governance-related risk events. These systems can dynamically adjust

ESG scores in response to emerging controversies. In addition, NLP and semantic analysis enhance the understanding of textual information and the extraction of relevant indicators, enabling ESG data to more accurately reflect a company's underlying risks and trends [1]. By integrating these technologies, AI enhances the efficiency of data processing, improves the accuracy of scoring, and provides methodological support for building a quantifiable, dynamically updatable ESG evaluation framework. This helps reveal actual corporate performance disparities across environmental, social, and governance dimensions.

3. Core applications of Artificial Intelligence in ESG financial analysis

3.1. Portfolio optimization and intelligent decision-making

AI enhances the analysis and optimization of investment portfolios by integrating information from multiple sources, including corporate annual reports, news articles, social media content, and policy documents. First, unstructured data are translated into measurable indicators, such as environmental emission levels, employee rights incidents, and governance risks, via natural language processing and semantic analysis [3]. Machine learning models, such as random forests, gradient boosting trees, or deep neural networks, are then trained on historical data to identify potential risk events and forecast future performance trends. A scoring system decomposes corporate ESG performance into sub-indicators and adaptively adjusts their weights according to real-time data, hence enabling asset allocation to more accurately capture a company's risks and opportunities [7]. The system can also develop customized strategies based on an investor's risk tolerance and financial profile, achieving an optimal balance between risk and return within the portfolio [8]. By integrating macroeconomic indicators, labor quality metrics, and industry development trends, the model assesses the long-term effects of structural changes on the portfolio, facilitating dynamic optimization over time and across markets. The entire process supports real-time data updates and multi-dimensional integration, thus ensuring that investment decisions are not only grounded in quantitative rigor but also agile enough to adapt to evolving market and policy conditions.

3.2. Risk identification and dynamic management

ESG rating inconsistencies may exacerbate operational risks and impose financing constraints on firms, particularly in regions where financial systems are less developed. As such, companies with lower ratings tend to experience higher capital costs. To mitigate these risks, AI can be deployed to build dynamic risk early-warning systems that monitor corporate ESG performance and related risk events in real time. For instance, by using satellite imagery, IoT sensors, and news sentiment data, AI can gather environmental, social, and governance-related information and convert unstructured data into quantifiable indicators [7]. Machine learning models can then be trained on historical risk incidents and corporate behavior patterns to identify potential compliance breaches or risk events, estimating their probability and potential impact [4]. A scoring mechanism with adaptive thresholds can automatically revise alert levels based on the latest data, capturing shifts in a company's risk profile. These models can also utilize policy document analysis to monitor shifts in environmental compliance standards and regulatory pressures. To reduce algorithmic bias and ensure assessment fairness, measures such as interpretability analysis, data validation, and fairness constraints should be implemented. This helps guarantee that early-warning signals faithfully reflect a company's true risk exposure. By integrating multi-source data, dynamic scoring, and predictive analytics, financial

institutions can achieve more precise risk management and quicker response mechanisms, thereby strengthening overall risk control capabilities.

3.3. Asset pricing and scoring models

The accuracy and objectivity of ESG scoring and asset pricing have been significantly improved via AI. In particular, the data sources include China Securities Index (CSI) ESG Ratings, Bloomberg ESG Index, along with corporate annual reports, news articles, and market sentiment data. In terms of methodology, neural networks, support vector machines, and multi-factor scoring models break down corporate ESG performance into components such as environmental input, social impact, and governance effectiveness, and improve the differentiation of scores through weight optimization [9]. NLP further examines sentiment in news and social media texts, producing dynamic ESG sentiment factors that are integrated into pricing models. This allows asset pricing to capture real-time market sentiment and the influence of external events. And the data processing workflow includes raw data cleaning, feature extraction, metric normalization, model training and validation, as well as dynamic adjustment of scores and weights, converting unstructured information into quantifiable indicators. Through this methodology, ESG scoring boosts accuracy and differentiation while also improving asset pricing's responsiveness to market volatility and policy shifts. It provides a more quantifiable and traceable foundation for portfolio risk management and return optimization [6].

4. Major challenges and development path

4.1. Data standardization and quality assurance

The lack of standardization and variability in ESG data quality are key barriers to the development of China's ESG system. This issue manifests in two key ways. First, different rating agencies differ substantially in their indicator weightings and data sources, with some emphasizing environmental performance and others placing greater focus on governance structure. Thus, a single company may be assigned differing ratings across various frameworks, reducing comparability and complicating investment decisions as well as corporate practices. Second, corporate disclosures frequently exhibit "symbolic reporting," in which only favorable information is disclosed or key data is presented ambiguously. In addition, selective reporting bias is prevalent, complicating the verification of the authenticity and completeness of non-financial information. This issue is particularly pronounced among small and medium-sized enterprises, where data gaps are frequent [9].

Besides, the prevalence of data silos, where information remains fragmented across departments, further aggravates data quality issues. To address these challenges, three types of measures can be adopted. For example, China can draw on international standards such as GRI to establish a unified disclosure framework tailored to local conditions, clarifying definitions and calculation methods for core indicators like carbon emissions and employee training. This would provide clear guidance for firms. Moreover, a data quality evaluation system could be developed. For example, GPT-4 can be utilized to perform semantic analysis of disclosure texts alongside cross-verification of multi-source data, thereby allowing for the precise detection of greenwashing with an accuracy rate of up to 82%. Meanwhile, regulatory authorities could require the disclosure of core ESG indicators and promote third-party verification, establishing a "disclose-verify-supervise" closed loop [6]. At the same time, the implementation of distributed data platforms leveraging blockchain technology could enhance traceability and secure data sharing, aiding in the elimination of information silos.

4.2. Regulatory compliance and risk prevention

In ESG financial analysis, cross-regional and cross-industry policy differences, along with dynamic policy changes, heighten compliance risks for enterprises and financial institutions. Consequently, ESG non-compliance can lead to regulatory sanctions and reputational harm. These consequences may negatively affect a company's financing capabilities, borrowing costs, and market valuations. Research shows that strong auditing plays a key moderating role between ESG performance and corporate outcomes, boosting information transparency and reducing risk transmission. Meanwhile, by sidestepping direct environmental responsibilities during intermediation, financial institutions may raise compliance risks throughout the industry [10,11].

Against this backdrop, AI technology can play a vital role in addressing these challenges. NLP can be utilized to examine evolving policy texts and, together with corporate operational data, construct compliance risk early-warning models. This facilitates real-time detection, alerting, and automated mitigation of potential violations. Moreover, AI technology can strengthen the scope and accuracy of ESG reporting reviews, thus improving the quality of information disclosure. Effective risk mitigation requires cooperation among regulators, firms, and financial institutions. Regulatory authorities should develop standardized ESG compliance guidelines and evaluation criteria while instituting industry-wide dynamic monitoring mechanisms. Firms and financial institutions must assume their core responsibilities by employing technological tools to reinforce risk management. In sum, this unified framework strengthens governance performance and policy adaptability in ESG finance, lowers compliance risks, and promotes the sector's long-term sustainability.

4.3. Technological innovation and development trends

ESG data collection and financial analysis are currently undergoing rapid transformation driven by technological innovation. The integration of multiple technologies and the expansion of application scenarios are enhancing data processing and decision-making capabilities. In data collection, the combination of artificial intelligence with the IoT and satellite remote sensing enables real-time and precise ESG data acquisition. For example, applying remote sensing to observe corporate carbon emissions or deploying IoT devices to monitor resource efficiency can alleviate the delays and errors linked to conventional manual procedures [9]. In data analysis, large language models (LLMs) can perform in-depth evaluations of corporate social responsibility reports and news sentiment, thus revealing latent risks, such as those related to supply chain social accountability. Simultaneously, reinforcement learning algorithms can be used to dynamically optimize ESG investment portfolios, enhancing their resilience to market volatility. Furthermore, AI technologies are increasingly being used to generate ESG ratings directly, integrating multimodal data such as satellite imagery and social media to enhance the timeliness and accuracy of assessments [12]. Looking ahead, practical strategies should work to close gaps in technology use across firms and regions. In technologically advanced companies, emphasis should be placed on applying AI to support green innovation and emission reduction strategies. In regions with underdeveloped digital infrastructure, priority should be given to building foundational technologies such as 5G and cloud computing to bridge the ESG data gap. Besides, focus should be placed on developing localized ESG rating systems, ensuring algorithmic fairness and transparency, and implementing ethical governance frameworks. Moreover, tools like dynamic scoring systems and automated investment platforms can enhance the accuracy and applicability of ESG investment analysis [5,8].

5. Conclusion

This paper reviews the technological advancements in ESG data collection and financial analysis, finding that methods such as natural language processing, machine learning, and generative models have markedly increased the effectiveness of ESG data gathering, analytical accuracy, and support for decision-making. The results indicate that AI can integrate multi-source heterogeneous data and construct dynamic risk assessment models, while demonstrating considerable potential in portfolio optimization, asset pricing, and compliance monitoring, thereby providing critical technical support for the advancement of sustainable finance practices. Nevertheless, the real-world influence of these technologies on ESG applications has not been evaluated. Furthermore, issues such as ESG data standardization, model bias, and interpretability have not been fully resolved, and their applicability in emerging markets remains to be further validated. Future research should focus on promoting the establishment of ESG data standardization and quality assessment systems, integrating technologies such as blockchain to enhance data credibility, developing more interpretable and fair AI models to mitigate algorithmic bias, and strengthening cross-regional and cross-industry comparative studies to explore AI adaptation pathways under different regulatory environments. Besides, the application of multimodal data integration and real-time analytical technologies should be strengthened to build more responsive and intelligent ESG assessment systems and investment support tools.

References

- [1] Birindelli, G., et al. (2025). How important are ESG ratings for financial institutions? Evidence from corporate leverage ratios across Europe. *International Review of Economics and Finance*, 102, 104398.
- [2] Zou, Y., & Tian, X. (2025). Influencing factors and economic consequences of ESG rating disagreements: A literature-based study. *Finance and Accounting Research*, (05), 65-71.
- [3] Yu, S., & Shi, H. (2025). The impact of artificial intelligence policy on corporate ESG performance: Empirical evidence from National AI Innovative Application Pilot Zones. *Western Forum*, 35(04), 52-67.
- [4] Wang, J. (2025). Digital transformation, environmental regulation and corporate ESG performance: Evidence from China. *Corporate Social Responsibility and Environmental Management*, 32(2), 1567-1582.
- [5] Chen, P., & Li, C. (2025). Generative AI empowers financial data governance: Value implication, evolutionary risks, and realization path. *Academic Exploration*, 1-12.
- [6] Zheng, Y., & Wang, Z. (2025). Research on the implementation mechanism of ESG rating system based on ChatGPT. *Friends of Accounting*, (01), 87-93.
- [7] Dash, A., & Mohanta, G. (2025). Drivers of sustainable financial consumerism: Exploring the impact of artificial intelligence, influencers, financial literacy, and product quality on sustainable development. *Cleaner and Responsible Consumption*, 18, 100306.
- [8] Liu, H. (2025). Artificial intelligence development and household financial asset allocation. *International Review of Economics and Finance*, 102, 104365.
- [9] Cheng, Y., & Li, H. (2025). The impact of ESG performance on corporate digital transformation. *Environment, Development and Sustainability*, 1-28.
- [10] Attia, E. F., & Almoneef, A. (2025). Impact of ESG on firm performance in the MENAT region: Does audit quality matter? *Sustainability*, 17(13), 6151.
- [11] Sklavos, G., et al. (2025). Unmasking greenwashing in finance: A PROMETHEE II-based evaluation of ESG disclosure and green accounting alignment. *Risks*, 13(7), 134.
- [12] Zhao, Y., Dai, R., & Nagayasu, J. (2025). Generative AI: The transformative impact of ChatGPT on systemic financial risk in Chinese banks. *Pacific-Basin Finance Journal*, 93, 102829.