# Addressing Heterogeneity and Improving Sensitivity and Accuracy of Existing ESG Ratings Based on NLP Large Models

Principal researcher: Qijia He
Department: College of Science & Engineering, Department of Mathematics & Statistics, the University of Glasgow
Contact: qhe@skidmore.edu / Qijia_He@outlook.com

## Research Motivation

ESG ratings are becoming a source indicator for investors to assess corporate sustainability performance along with the recent increasing public awareness in sustainable investment. However, since each of the rating agencies in incorporate their customized rating methods, data source, metric usages differently, thus customized the heterogeneity in ESG ratings has been arising against the increasing number of ESG rating agencies. The presence of this inconsistency can diminish corporate motivation to enhance ESG practices and impede the precise valuation of their ESG conducts in the market. This heterogeneity also undermines investors' confidence in ESG metrics as reliable investment criteria, thereby obstacling investment in companies that are genuinely working towards sustainability, as investors may find it challenging to accurately assess and compare their ESG performance.

The existing literature explores the implications of such inconsistency on investment decisions and corporate incentives to pursue sustainable practices. For instance, Berg et al. (2022) reveal a significant variation in the correlation of ESG scores across different rating agencies, with associations ranging from 0.38 to 0.71. Given a prior consideration of an agency-specific rating methodology, Billio (2021) emphasises that the inconsistency in ESG rating standards can result in biased outcomes. Such biases could, in turn, lead investors to misinformed decisions if they rely on these ratings in their investment strategies. Despite the efforts by agencies to improve rating processes, Dobrick (2021) mentions a substantial heterogeneity that remains in methodological discrepancies, firm-specific characteristics, and all other outside effects.

Considering the notable heterogeneity in ESG ratings, this research intends to investigate the intricacy of ESG rating heterogeneity. In particular, it aims to examine how various legal contexts and cultural idiosyncrasies may contribute to the rating discrepancies, in addition to the diverse rating methodologies employed by rating agencies. To accomplish this, two primary research questions will be addressed, each question of which forms the basis of a separate research paper.

1) **Research question 1:** Assessing the impact of country-specific regulatory frameworks and cultural practices on ESG ratings divergence.
2) **Research question 2:** Developing a new rating mechanism that uses natural language processing (NLP) and textual analysis to address ESG rating heterogeneity.

## Literature review

As ESG is gaining prominence in global financial markets, the issue of rating heterogeneity emerges together with the continuous involvement of ESG rating agencies. This heterogeneity manifests itself as disparity in the ratings assigned by different ESG rating agencies when evaluating similar or identical criteria, arising from variations in rating methodologies, data sources, and evaluation criteria. Specifically, Christensen (2019) reveals that E(environmental) and S(social) disclosures have a stronger determine on ESG rating result compared to G (governance).

Berg et al. (2022) disclose substantial disparities in the correlation of ESG scores across different rating agencies, spanning from 0.38 to 0.71. Meanwhile, Dobrick (2023) uses a linear mixed model analysis to conclude the presence of sizing bias in the ESG rating data from Refinitiv's ASSET4 ESG database. This finding primarily argues that larger firms tend to have higher ESG ratings. Furthermore, Breg et al., (2022) summarise the primary drivers of ESG divergence into three main categories: scope divergence, measurement divergence, and weight divergence.

Scope divergence involves different attributes or indicators chosen by agencies when evaluating ESG performance. For instance, Refinitiv incorporates 282 individual indicators while S&P Global and MSCI only contain 68 and 78, highlighting a notable contrast in the attributes and indicators they employ. In addition, MSCI includes scores that measure the relevance of issues for specific companies which no other rating agencies explicitly measures. Measurement divergence occurs since each rating agency may adopt distinct metrics or methodologies when evaluating identical attributes. When analysing the average scores assigned by different ESG rating agencies to the identical categories of companies, a clear discrepancy in the ratings is noticeable. This is particularly evident in the opposing scores that Sustainalytics and Moody's award for corporate lobbying efforts as well as the divergent ratings Sustainalytics and Refinitiv give on matters concerning indigenous peoples' rights. The evidence suggests that rating agencies can draw completely opposite conclusions about the same category, showing that they are divided on these ESG issues. Weight divergence is the result of rating agencies assigning different levels of importance， or weight, to various attributes. For example, the top three categories of the MSCI KLD 400 Social Index (the first U.S. index used to rate social and environmental issues) are focused on managing climate risk, ensuring product safety, and overseeing compensation. These priorities stand in stark contrast to those of Moody's, which places diversity, environmental policy, and labour practices at the forefront of its evaluation criteria.

Bronzini (2023) applied the LLaMA-based "WizardLM" natural language processing model (NLP) to analyse 6,456 sustainability reports from 4,222 companies, revealing that differences in ESG ratings often reflect specific country and industry concerns. For example, water resource issues are more prominent in water-intensive industries, while packaging concerns are emphasised in consumer goods companies, underscoring ESG rating heterogeneity. Moreover, the study also integrates large language models, context learning, and Retrieval-Augmented Generation techniques and finds a significant correlation between reporting transparency and higher ESG scores. Similar to Bronzini (2023), Hiechl (2023) employs NLPand text mining approach to explore ESG rating divergence among firms in four EU countries（France, Italy, Germany, and Sweden) and state that these country- and sector-specific rating variations may be policy-driven issues. The study specifically points out that in Italy, the "G (Governance)” aspect has the highest weight in ESG ratings due to the mandatory audit of non-financial statements as per Italian law. Conversely, it appears to be less crucial in the French market, possibly owing to the sectoral makeup of its economy. Beyond from a policy-driven perspective, Hiechl (2023) highlights the critical impact of country-specific cultural attributes and regulatory frameworks on ESG rating practices and emphasise the necessity to incorporate these elements when comparing ESG ratings across different countries. Renneboog (2017) states that the differentiated implementation of CSR (Corporation Social Responsibility) disclosure regulations across countries leads to heterogeneity in ESG ratings. Mandatory CSR disclosure law increases corporate administrative costs while higher administrative expenses can diminish its financial performance, ultimately affecting its "Governance" score. In common law countries where shareholder interests are prioritised, such administrative costs may further negatively impact ESG rating outcomes. Conversely, in countries centred on stakeholder interests, the persistence and stickiness of administrative costs may reflect a long-term commitment to "S (Social)", which could show a positive influence on ESG scores. Liang (2017) also suggests that companies in common law countries have lower CSR than companies in civil law countries, while civil law companies in Scandinavia have the highest voluntary CSR ratings. This characteristic demonstrates the heterogeneity of ESG ratings under different regulatory frameworks. Based on this study, the latest researchers believe that the application and interpretation of ESG can vary across regions and jurisdictions. Shen (2023) proposes that ESG research in China stress firms' ESG disclosure and performance, as well as ESG investments. According to the data, in 2021, 30% of listed companies published reports related to ESG, of which 77% were social responsibility reports, 13% were ESG reports and 5% were sustainability reports. These figures show that CSR reports dominate ESG reports in China, while the proportion of ESG reports and sustainability reports is relatively low (Shen et al., 2023).

## Empirical Methodologies

This section outlines the empirical methodologies and data sources that I intend to use to address my research questions.

**Research question 1:** Assessing the impact of country-specific regulatory frameworks and cultural practices on ESG ratings divergence.

The first research question aims to investigate how country-specific regulatory structures and cultural practices contribute to the variation in ESG ratings, beyond the methodological differences found among ESG rating agencies. To do so, conditional ordered logistic regression with fixed effects will be considered as one of the potential models for empirical analysis (See Equation 1). In Equation 1, $Y_{ijt}$ represents the ordinal dependent variable which is the ESG rating for $Firm_i$, in $Country_j$ at $Time_t$. The ratings are categorised in AAA, AA, A, BBB, BB, B and CCC where ESG performance is getting worst from AAA to CCC. The implication of the logit model allows to transform the probability of the ESG rating being at or below a certain level $k$ into a continuous variable that can be estimated via linear regression model techniques. The $\alpha_k$ represents the threshold coefficients or cut point for each category k of the dependent variable.

$$logit\left[Pr\left(Y_{ijt} \leq k\right)\right] = \alpha_k + \beta_1 Law_{jt} + \beta_2 Macro_{jt} + \beta_3 Culture_{jt} + \gamma FirmControls_{ijt} + \mu_j + \lambda_t + \varepsilon_{ijt} \tag{1}$$

$Law_{jt}$ is an index that reflects the level of a country's legal framework. $Macro_{jt}$ represents a vector of country-specific macroeconomic factors, such as GDP, market capitalization, number of listed companies, number of companies with ESG disclosures, and the number of government subsidies or investments in sustainable businesses. $Culture_{jt}$ represents an index that reflects the country-specific cultural practices. $\mu_j$ *and* $\lambda_t$ represents individual country fixed effects accounts for time effects, accordingly. $\varepsilon_{ijt}$ refers to other unobserved factors affecting the ESG rating for $Firm_i$ in $Country_j$ at $Time_t$ that are not included in the model. Besides from the above variables of interest, the model also controls firm-specific characteristics $FirmControls_{ijt}$, such as firm size, profitability ratios, liquidity ratios, gearing ratios, industry, age of the company, market performance, R&D expenditures, etc.

For robustness check, the ordered probit model will be performed as an alternative mode. Unlike the original model, the ordered probit model adopts a normal distribution while the ordered logit model is based on the logistic distribution. If the results of these two models are consistent, it indicates that our findings are not dependent on specific distributional assumptions, hence augmenting the veracity of the conclusions. The function of the probit model shows as below:

$$\begin{aligned} Probit\left[Pr\left(Y_{ijt} \leq k\right)\right] &= \Phi^{-1}\left[Pr\left(Y_{jit} \leq k\right)\right] \\ &= \alpha_k + \beta_1 Law_{jt} + \beta_2 Macro_{jt} + \beta_3 Culture_{jt} \\ &\quad + \gamma FirmControls_{ijt} + \mu_j + \lambda_t \end{aligned} \quad (2)$$

where $\Phi^{-1}$ is the inverse cumulative distribution function. Information criteria such as AIC and BIC will be used to assess the overall performance between the two models. Moreover, a K-folds cross-validation analysis will be used to evaluate the overall predictive accuracy of the model. In particular, the given sample data will be randomly divided into K subsets (folds). In each $K$ iterations, a subset is selected as the test set, and the remaining $K-1$ subsets are used as the training set. In each iteration, the training set is used to fit our original ordered logit model. At this stage, the parameter $\beta_{1,2,3}$ for $Law_{jt}$, $Macro_{jt}$, and $Culture_{jt}$, and the parameter $\gamma$ for $FirmControl_{ijt}$ will be estimated on the training set. Then, use a test set to evaluate the model's accuracy, recall, or ability to distinguish different rating categories. Finally, the performance of all K iterations is summed, and the mean and standard deviation are calculated to assess the overall robustness of the model.

**Variables of Interest**

Main research variables

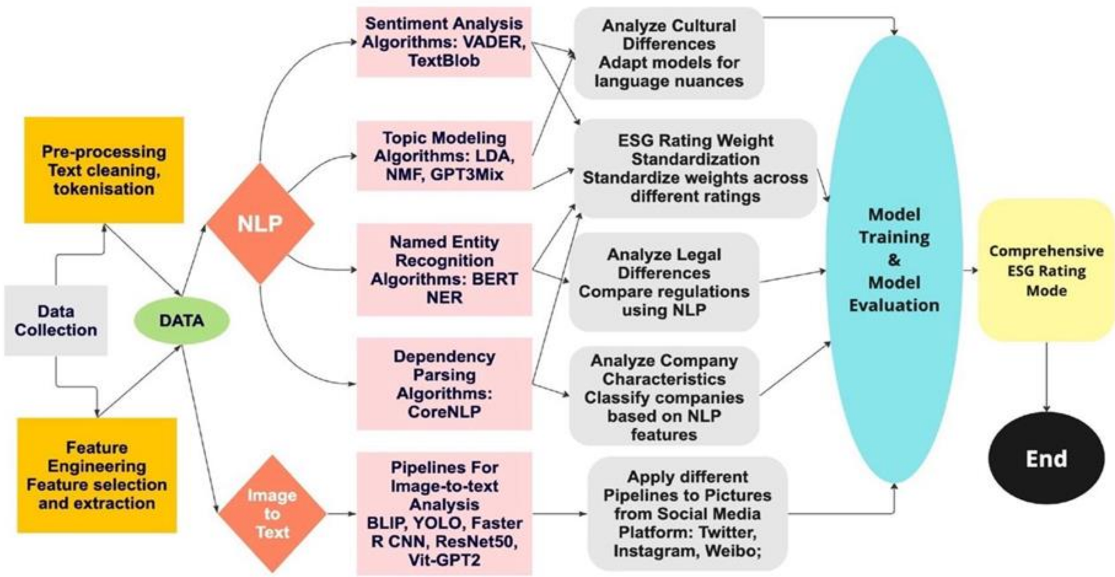
1. ESG Ratings
    - Sources: MSCI, Sustainalytics, Bloomberg, RobecoSAM, Dow Jones
    - Time Period: 2012 to 2022

Explanatory Variables

1. Legal Frameworks
    - Sources: Worldwide Governance Indicators (WGI), World Justice Project (WJP)
    - Coverage: Broad understanding of rule of law and governance quality across countries
2. Macroeconomic Factors
    - Variables may include:
        - Total value of country's stock price
        - Number of listed companies and disclosure of ESG reports
        - Government subsidies/investments in sustainable businesses
        - Inflation Rate, GDP growth Rate, Total National Debt
    - Sources: Bloomberg, International Monetary Fund (IMF), World Bank, OECD
3. Cultural Influences
    - Data Sources: GLOBE Study, World Values Survey (WVS)
    - Focus: Quantifying cultural impacts on ESG ratings

**Research question 2:** Developing a new rating mechanism that uses natural language processing (NLP) and textual analysis to address ESG rating heterogeneity.

Figure 1 provides a general overview of the machine learning framework that this research considers using for its NLP-based textual analysis, with an aim at addressing the issues of ESG rating heterogeneity.




**Data pre-processing**

This procedure involves removing irrelevant content (such as URLs, HTML tags, and non-textual elements), tokenising the text into words or phrases, normalising words to their base form using the snowball package and staff, and then eliminating common stop-words to reduce noise. Once the data is obtained, NLP analysis can be applied partially to the data set.

**Standardising ESG rating**

Firstly, classification task is performed by using encoder models such as RoBERTa, DeBERTa, and FinBERT to capture the complexity of human language and align the augmented data with the desired label semantics. Since ESG rating is a globalisation content, the study will train a multilingual model to extend its applicability in different linguistic contexts. The task involves several processes: using the MSCI ESG rating guidelines to classify multilingual news articles into 35 key ESG issues; using Latent Dirichlet Allocation (LDA) to categorise report content for focus extraction; creating word clouds for visual frequency insights; identifying key ESG terms with TF-IDF and Word2Vec; classifying through K-means and Hierarchical clustering; using GPT3Mix (a technique that generates synthetic text samples using large-scale language models) to augment the data; using Zero-Shot Classification (a process that allows a model to recognise and classify new data types without training on a specific type of data) to handle uncommon or emerging ESG issues. Finally, using tools like PANDAS for frequency analysis to weigh each standard and formulate a new unified rating benchmark.

**Cross-country Analysis**

In the Cross-country Law part, the study will employ text classification models like

BERT and RoBERTa to analyse and predict the legal system basis (Common Law System, Civil Law System) of various reports and texts. The method involves encoding these documents and generating a probability distribution to determine their alignment with different legal systems. The goal is to understand how different legal origins impact corporate ESG reporting and practices. Additionally, this process will involve extracting key legal terms and phrases from the texts, providing outputs as relevance or weight scores, which aids in understanding the influence of specific legal concepts in each legal system. We plan to utilise LDA to analyse cultural texts in terms of Cross-country Culture aspect. The texts will be transformed into a format suitable for LDA analysis by TF-IDF, which involves training the model to learn the distribution of topics across documents and assigning meaning to each topic based on word frequency and distribution. The analysis will be extended to compare thematic distributions across texts from different cultures to discern how various cultural backgrounds influence these themes. Sentiment analysis will also be applied to the texts using models like BERT, GPT, or potentially custom transformer-based codes, alongside LSTM (although for LSTM retraining might be required). This will help classify the general attitude towards ESG across cultural background as positive, neutral, or negative.

**Firm Specific**

When examining Firm Specifics, the methodology incorporates a BERT-based Named Entity Recognition (NER) system to identify key characteristics of companies. This advanced technique allows to extract precise information pertinent to each firm, such as industry classification, corporate structure, and operational attributes. The relationship between a company and its ESG initiatives will be done by using Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs). This approach facilitates a deeper analysis of the connection between company-specific factors and their ESG performance. BERT-NER will be used again to analyse ESG-related press releases from company websites, news outlets, and industry publications.

**Image-to-text Analysis for social media**

Social media platform provides the public attitude toward sustainability and corporate ESG performance. To analyse this, the paper needs to define broad categories of visual features that cover the multi-dimensional characteristics of ESG. Firstly, the collected image data will be annotated manually to build and continuously expand the training dataset, thereby training and improving the accuracy of the model. The annotated data will be trained subsequently by using deep learning models such as BLIP, ResNet50 and YOLO, and the optimal model will be selected based on performance in the training and validation sets. Then, extracting features to identify the visual characteristics that best represent ESG and validating the model on an independent test set. Evaluating the model using metrics such as confusion matrix, accuracy, recall, and F1 score, and iteratively optimise based on these results. The optimised model will be able to automatically identify and categorise newly uploaded images related to ESG on social media.

Following the flowchart of Figure 1, the final model will encompass all the aforementioned sub-models based on NLP algorithms. It will be capable of rating corporate ESG using algorithms such as SVM, gradient boosting models, or serialised data processing architectures based on CNN or RNN. Additionally, Cross-Validation and Grid Search will be employed to optimise the final model.

**Data sources**

1. News and articles
   - Bloomberg: Comprehensive company news, financial data and market analysis
   - The Wall Street Journal: Financial news and information on ESG
   - Google News: Access to various news stories via API authentic

2. Legal documents and reports
   - Westlaw/LexisNexis: Legal Research Database
   - Country Based Website/institution Country based websites/institutions:
     - China Judicial Instruments Website
     - EUR-Lex: EU legal documentation, including directives and regulations on environmental and social governance.
     - Legislation.gov.uk
     - ASEAN Law Association
     - J-PlatPat and Korea Legislation Research Institute

3. Industry-specific databases
   - Carbon Disclosure Project: Provide corporate GHG emissions and water management data.
   - Global Reporting Initiative (GRI): Provide corporate sustainability reports and data.
   - RobecoSAM: Specialises in sustainable investing and provides annual sustainability assessments of companies.
   - PwC, Deloitte, EY, KPMG (The Big Four): These audit and consulting firms provide industry reports that often include ESG-related insights and analysis.

4. Corporate reports and data
   - MSCI ESG Research, Sustainalytics, FTSE Russell
   - Regional and national databases:
     - Amadeus Database by Bureau van Dijk, Bundesanzeiger, Infogreffe
     - Information platform disclosed by the China Securities Regulatory Commission, Hong Kong Exchanges and Clearing Limited
     - KISLINE, Japan Exchange Group
     - Electronic Data-Gathering, Analysis, and Retrieval system
     - ASIC (Australian Securities and Investments Commission), Companies Office of New Zealand
     - Tadawul: Saudi Arabian stock market with data on listed companies.
     - Dubai Financial Market: Information on listed companies in Dubai.

5. Transparency and disclosure platforms
   - Transparency International: Data related to corruption, which is important for the "governance" aspect of ESG assessments.
   - The World Bank and IMF: Provides data on governance and policies at the country level, which can influence a company's ESG score.

6. Social Media
   - Twitter API，Weibo API, Instagram API
   - Image databases
     - ImageNet: a large-scale image database that can be used to train and validate deep learning models.
     - Flickr: A large number of images can be accessed through its API, including

images that may be labelled as ESG-related.

**Robustness check**

In conducting the model’s robustness check, select representative time intervals such as 2017-2018, 2019-2020, and 2021-2022; and identical cleaning and preprocessing steps were applied to the data at each time point to ensure consistency in the analysis. The ESG ratings of companies across different industries for the selected time periods were assessed separately using historical data. Upon obtaining the rating results, the rating changes derived from the new model were compared with those provided by existing ESG rating agencies. If the trend of changes in ESG ratings for companies within the model aligns with that of the existing ESG rating agencies, this suggests that the new model is robust to temporal variations.

To further analyze the validity of these changes, the difference in the magnitude of trend strength between the old and new models for the same company over the same time interval was compared (for instance, the new model rated a company from AAA to BB, whereas the old model rated it from AAA to A). Initially, a numerical code was assigned to each rating level to convert qualitative ratings into quantitative data (e.g., AAA was coded as 1, AA as 2, and so on). Then, for each company, the change in its rating between two time points was calculated. For example, if a company's rating drops from AAA to A in the new model, the change value is -2. The change value for each company should be $\Delta$， where

$$\Delta = Label_{latter\ year} - Label_{start\ year}$$

Utilized Mann-Whitney U test to test for statistically significant differences in rating transitions between the new model and existing agencies. Using two datasets, one representing rating changes from the new model (Group X) and the other encapsulating changes as per the existing rating agency (Group Y). The collective data from both groups were ordered by the magnitude of rating change, and ranks were assigned to each data point. In instances of tied rating changes, average ranks were bestowed.

$R_x$ is sum of ranks for Group X and $R_y$is sum of rank for group Y, $n_x$ and $n_y$ is the number of observations in both groups， the Mann-Whitney statistic was then computed, with the smaller of the two U values derived from the rank sums being indicative of the test result.

$$U_x = R_x - \frac{n_x(n_x + 1)}{2}$$

$$U_y = R_y - \frac{n_y(n_y + 1)}{2}$$

Apply smaller U value to the Mann-Whitney U statistics and look up in a Mann-Whitney U distribution table to find the p-value. The ensuing p-value determined the statistical significance of the rating change disparities observed between the two models, thus serving as a robustness indicator for the new ESG rating model. Also, could check the model by future analyze the variety of internal and external factors of the company such as financial statements, changes in governance structure, changes in the market environment, etc... If it changes more closely matches the downward/ upward trend of the new ratings, it indicates that the new model has been improved and is able to provide more sensitive ratings.

## Proposed timeline

## Reference list

Berg, F., Kölbel, J.F. and Rigobon, R. (2022), "Aggregate Confusion: The Divergence of ESG Ratings", *Review of Finance*, Vol. 26 No. 6, pp. 1315–1344, doi: 10.1093/rof/rfac033.

Berger-Walliser, G. and Scott, I. (2018), "Redefining Corporate Social Responsibility in an Era of Globalization and Regulatory Hardening", *American Business Law Journal*, Vol. 55 No. 1, pp. 167–218, doi: 10.1111/ablj.12119.

Billio, M., Costola, M., Hristova, I., Latino, C. and Pelizzon, L. (2021), "Inside the ESG ratings: (Dis)agreement and performance", *Corporate Social Responsibility and Environmental Management*, Vol. 28 No. 5, pp. 1426–1445, doi: 10.1002/csr.2177.

Brandon, R.G., Krueger, P. and Schmidt, P.S. (n.d.). "ESG Rating Disagreement and Stock Returns".

Bronzini, M., Nicolini, C., Lepri, B., Passerini, A. and Staiano, J. (2023), "Glitter or Gold? Deriving Structured Insights from Sustainability Reports via Large Language Models", arXiv, 3 November.

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., *et al.* (2020), "Language Models are Few-Shot Learners", arXiv, 22 July.

Chatterji, A.K., Durand, R., Levine, D.I. and Touboul, S. (2016), "Do ratings of firms converge? Implications for managers, investors and strategy researchers", *Strategic Management Journal*, Vol. 37 No. 8, pp. 1597–1614, doi: 10.1002/smj.2407.

Chen, Y.-C., Hung, M. and Wang, Y. (2018), "The effect of mandatory CSR disclosure on firm profitability and social externalities: Evidence from China", *Journal of Accounting and Economics*, Vol. 65 No. 1, pp. 169–190, doi: 10.1016/j.jacceco.2017.11.009.

Christensen, D.M., Serafeim, G. and Sikochi, A. (2022), "Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings", *The Accounting Review*, Vol. 97 No. 1, pp. 147–175, doi: 10.2308/TAR-2019-0506.

Crifo, P. and Rebérioux, A. (2016), "Corporate governance and corporate social responsibility: A typology of OECD countries", *Journal of Governance and Regulation*, Vol. 5 No. 2, pp. 14–27, doi: 10.22495/jgr_v5_i2_p2.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019), "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", arXiv, 24 May.

Dobrick, J., Klein, C. and Zwergel, B. (2023), "Size bias in refinitiv ESG data", *Finance Research Letters*, Vol. 55, p. 104014, doi: 10.1016/j.frl.2023.104014.

Gibson Brandon, R., Krueger, P. and Schmidt, P.S. (2021), "ESG Rating Disagreement and Stock Returns", *Financial Analysts Journal*, Vol. 77 No. 4, pp. 104–127, doi: 10.1080/0015198X.2021.1963186.

Gu, Y., Dai, J. and Vasarhelyi, M.A. (2023), "Audit 4.0-based ESG assurance: An example of using satellite images on GHG emissions", *International Journal of Accounting Information Systems*, Vol. 50, p. 100625, doi: 10.1016/j.accinf.2023.100625.

He, K., Zhang, X., Ren, S. and Sun, J. (2015), "Deep Residual Learning for Image Recognition", arXiv, 10 December.

He, P., Liu, X., Gao, J. and Chen, W. (2021), "DeBERTa: Decoding-enhanced BERT with Disentangled Attention", arXiv, 6 October.

Heichl, V. and Hirsch, S. (2023), "Sustainable fingerprint – Using textual analysis to detect how listed EU firms report about ESG topics", *Journal of Cleaner Production*, Vol. 426, p. 138960, doi: 10.1016/j.jclepro.2023.138960.

Huang, A.H., Wang, H. and Yang, Y. (2023), "FINBERT : A Large Language Model

for Extracting Information from Financial Text*”, *Contemporary Accounting Research*, Vol. 40 No. 2, pp. 806–841, doi: 10.1111/1911-3846.12832.
Kang, J., Kchouk, M., Bellato, S., Gan, M. and Maarouf, I.E. (n.d.). “FinSim4-ESG Shared Task: Learning Semantic Similarities for the Financial Domain. Extended edition to ESG insights”.
Lee, H., Choi, J., Kwon, S. and Jung, S. (2023), “EaSyGuide : ESG Issue Identification Framework leveraging Abilities of Generative Large Language Models”, arXiv, 13 June.
Liang, H. and Renneboog, L. (2017), “On the Foundations of Corporate Social Responsibility”, *The Journal of Finance*, Vol. 72 No. 2, pp. 853–910, doi: 10.1111/jofi.12487.
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., *et al.* (2019), “RoBERTa: A Robustly Optimized BERT Pretraining Approach”, arXiv, 26 July.
Luo, D., Yan, J. and Yan, Q. (2023), “The duality of ESG: Impact of ratings and disagreement on stock crash risk in China”, *Finance Research Letters*, Vol. 58, p. 104479, doi: 10.1016/j.frl.2023.104479.
Shen, H., Lin, H., Han, W. and Wu, H. (2023), “ESG in China: A review of practice and research, and future research avenues”, *China Journal of Accounting Research*, p. 100325, doi: 10.1016/j.cjar.2023.100325.
Tsang, A., Frost, T. and Cao, H. (2023), “Environmental, Social, and Governance (ESG) disclosure: A literature review”, *The British Accounting Review*, Vol. 55 No. 1, p. 101149, doi: 10.1016/j.bar.2022.101149.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., *et al.* (2023), “Attention Is All You Need”, arXiv, 1 August.