CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	LLM-powered question answering for Environmental, Social, and Governance (ESG) reports
Authors	Tran, Vinh-Khanh;Tran, Khanh-Tung;Pham, Quoc-Viet;Guha, Krishnendu;Nguyen, Hoang D.
Publication date	2025
Original Citation	Tran, V.-K., Tran, K.-T., Pham, Q.-V., Guha, K. and Nguyen, H.D. (2025) 'LLM-powered question answering for Environmental, Social, and Governance (ESG) reports', AIQAM '25: Proceedings of the 2nd ACM Workshop in AI-powered Question & Answering Systems, pp. 49-53. https://doi.org/10.1145/3746274.3760395
Type of publication	Conference item
Link to publisher's version	10.1145/3746274.3760395
Rights	© 2025, the Authors. This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License. - https://creativecommons.org/licenses/by-nd/4.0/
Download date	2026-11-15 18:48:28
Item downloaded from	https://hdl.handle.net/10468/18364



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

LLM-Powered Question Answering for Environmental, Social, and Governance (ESG) Reports

Vinh-Khanh Tran
125117729@umail.ucc.ie
University College Cork
Ireland

Khanh-Tung Tran
123128577@umail.ucc.ie
University College Cork
Ireland

Quoc-Viet Pham
viet.pham@tcd.ie
Trinity College Dublin
Ireland

Krishnendu Guha
KGuha@ucc.ie
University College Cork
Ireland

Hoang D. Nguyen
hn@cs.ucc.ie
University College Cork
Ireland

Abstract

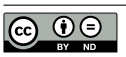
Sustainability reporting has become a crucial tool for companies to demonstrate commitment and performance on Environmental, Social, and Governance (ESG) matters. However, implementing ESG reporting presents many challenges, including data standardization as well as conforming to various reporting frameworks and standards such as Global Reporting Initiative, Sustainability Accounting Standards Board, Task Force on Climate-related Financial Disclosures, and European Sustainability Reporting Board. These challenges are more prevalent, especially in small-to-medium enterprises. To address these issues, this paper proposes a question answering framework leveraging Retrieval-Augmented Generation combined with large language models to support ESG management in business environments. The framework aims to ease the required resources and efforts for ESG analysis, improve the accuracy and relevance of retrieved information, and aid organizations in analyzing their sustainability reports. Additionally, we explore different methods for handling heterogeneous data types in ESG reports, including text and tabular data retrieval strategies. Our implemented system, tested on data obtained from leading companies in Australia, showcases promising results in accommodating a wide variety of use cases in the ESG domain. This demonstrates the potential of the proposed framework as well as other Artificial Intelligence-driven tools in assisting organizations in effectively managing ESG reporting and improving their sustainability performance.

CCS Concepts

• **Computing methodologies** → **Artificial intelligence**.

Keywords

Sustainability, ESG reporting, Artificial Intelligence, Retrieval-Augmented Generation, Heterogeneous document, Conversational



This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.
AIQAM '25, Dublin, Ireland
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2056-7/2025/10
https://doi.org/10.1145/3746274.3760395

ACM Reference Format:
Vinh-Khanh Tran, Khanh-Tung Tran, Quoc-Viet Pham, Krishnendu Guha, and Hoang D. Nguyen. 2025. LLM-Powered Question Answering for Environmental, Social, and Governance (ESG) Reports. In *Proceedings of the 2nd ACM Workshop in AI-powered Question & Answering Systems (AIQAM '25), October 27–28, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3746274.3760395

1 Introduction

More and more companies have adopted sustainability reporting as a tool to demonstrate to stakeholders their commitment as well as their performance regarding Environmental, Social, and Governance (ESG) matters [5]. According to KPMG's Survey of Sustainability Reporting, 96% of the world's biggest 250 companies and 79% of the 5,800 studied companies have issued sustainability reports in the year 2024 [6]. However, implementing ESG reporting may pose many challenges, ranging from data-based challenges, which involve acquiring, validating, and standardizing diverse data across business units, to the contextual challenges requiring organizations to comply with and navigate through complex and evolving standards and frameworks, including the Global Reporting Initiative (GRI), the Sustainability Accounting Standards Board (SASB), the Task Force on Climate-related Financial Disclosures (TCFD), and the European Sustainability Reporting Standards (ESRS) [1]. This is especially true for Small and Medium-sized Enterprises (SMEs), which face many challenges from a lack of expertise and limited resources hindering them from successfully implementing these complicated standards and frameworks [11].

To address these challenges, this paper proposes a framework that leverages the Retrieval-Augmented Generation (RAG) [7] paradigm to support sustainability management in business environments, with the goal of aiding companies in achieving the United Nations Sustainable Development Goals (SDGs) [13]. Specifically, the proposed framework aims to reduce the resources needed to carry out ESG analysis efficiently, ensures that information retrieved on sustainability is accurate and relevant, and gives insights into ESG reports. The many use cases of the system include assessing a company's ESG report to see if it complies with relevant reporting standards, comparing ESG reports between various enterprises to gain insight into ESG metrics, and guiding a corporation in drafting and improving ESG reports.

In addition, this paper surveys various methods for handling different types of data, including texts and tabular data, which are

some of the defining characteristics of ESG reports, and evaluates the performance of these approaches. Specifically, two different methods will be considered for handling tabular data: the naive method of treating the table as plain text, relying on keyword search, semantic search, and reranking; and the method demonstrated in [2], which leverages Large Language Models (LLMs) to summarize the columns and row samples of the tables, combined with lexical search and semantic search to retrieve relevant tables for a user's query.

The main contributions of this research include the design of a framework tailored to ESG reporting that leverages the power of LLMs. As a demonstration of the framework, we implement a system using ESG reporting data collected from top companies in Australia. Finally, an assessment of the effectiveness of the demonstrated system is conducted, specifically focusing on its ability to handle a wide range of ESG-related use cases in business environments and evaluating how different tabular data handling strategies impact the accuracy and reliability of information retrieval from ESG reports.

2 Related Works

2.1 RAG and Multimodal RAG

RAG [7] has become a widely adopted technique in AI to address the limitations of pre-trained models when it comes to data recency and correctness. However, traditional RAG approaches focus mainly on textual data, many real-world use cases often require handling non-textual data. For instance, traditional RAG approaches can be less effective in domains such as ESG reporting where non-textual data including structured and non-structured tables, figures, and charts are common. To address this issue, researchers have begun to focus on multimodal RAG approaches, which incorporate multimodal data into the RAG pipeline. For instance, the paradigm TAG proposed in [3] enhances LLMs capability to answer natural language queries over tabular data leveraging text2sql and the RAG method. Other notable efforts have been made in RAG multimodality including the research in [8] which introduces MultiRag, a multimodal RAG system designed to assist humans in information-intensive situations by reasoning over multi-source information, including video, audio, and text.

2.2 AI in the Sustainability Domain

As sustainability report gains more and more adoption, extensive research has been done in leveraging AI in the field of sustainability. For example, FinGPT [15], an open-source LLM for the financial sector, can be used to evaluate companies' ESG scores through the analysis of public reports and news articles. Another example is ESGBert [9], a language model based on the Bert [4] architecture, fine-tuned with ESG-specific corpus to accurately classify ESG-specific texts. These models excel at classification and detection tasks within the ESG domains, but they are limited based on the data they were trained on; in contrast, the framework proposed in this paper provides users with grounded, up-to-date information through RAG techniques.

Researchers have begun to incorporate the RAG paradigm into their work to address issues regarding data recency and truthfulness, which are important aspects in the ever-changing field of sustainability, particularly for ESG reports. Notably, ESGReveal [16] is an LLM-based framework combined with RAG techniques to extract information and analyze corporate ESG reports. While ESGReveal used many techniques such as Microsoft's Table-Transformer [12] and LORE-TSR [14] in the extraction phase, its retrieval phase handles tabular data and textual data identically, and primarily relies on vector similarity. In this paper, the heterogeneous nature of ESG data is addressed by surveying and evaluating different tabular data handling methods within the retrieval phase.

3 QA-ESG: An LLM-Powered Question Answering Framework for ESG Data

In this section, we will present the overall architecture and key components of the proposed framework. The discussion is divided into two subsections. The first subsection goes into the design of the proposed framework, examining all of the key modules, explaining their roles and contributions to the overall framework's operation. Following the design of the framework, the second subsection focuses specifically on how the proposed framework handles heterogeneous ESG data and compares two deployed retrieval techniques.

3.1 Framework Design

ESG data often comes in various formats, structures, and quality levels, which requires a system capable of handling heterogeneous inputs and adapting to changing reporting standards and organizational needs [1]. To accommodate the diverse and evolving nature of ESG reporting data, the proposed framework is designed with flexibility as a core principle. The framework, illustrated in Figure 1 contains multiple interconnected modules that collaborate to enable efficient ingestion, retrieval and generation of ESG related data.

The **Web Portal** serves as the interaction layer between business users and the framework. It is design to specifically meet the needs of organizations in managing ESG data. It provides a centralized interface through which users can upload and manage their ESG document corpus, submit links to external reports, and interact directly with LLMs integrated within the system. Users can upload ESG reports in various formats such as PDF, Word, Excel and CSV or alternatively provide links to online resources such as hosted sustainability reports. In addition to upload capability, the web portal allows users to browse and manage existing corpora. They can update or delete previously uploaded ESG reports. The corpora management capability of the Web Portal allows businesses to maintain a well-structured, organized and up-to-date ESG knowledge base.

In addition to handling document management, the Web Portal also serves as an interactive conversational interface to the **Generation Module**. The **Generation Module** which leverages the power of LLMs, is a core component responsible for synthesizing meaningful outputs based on the ESG data stored within the framework. The module is built to be as grounded and factually correct as possible. It leverages chunks of data retrieved from the **Document Store** and employs prompt engineering [10] to guide the model's output generation. These techniques are essential in minimizing

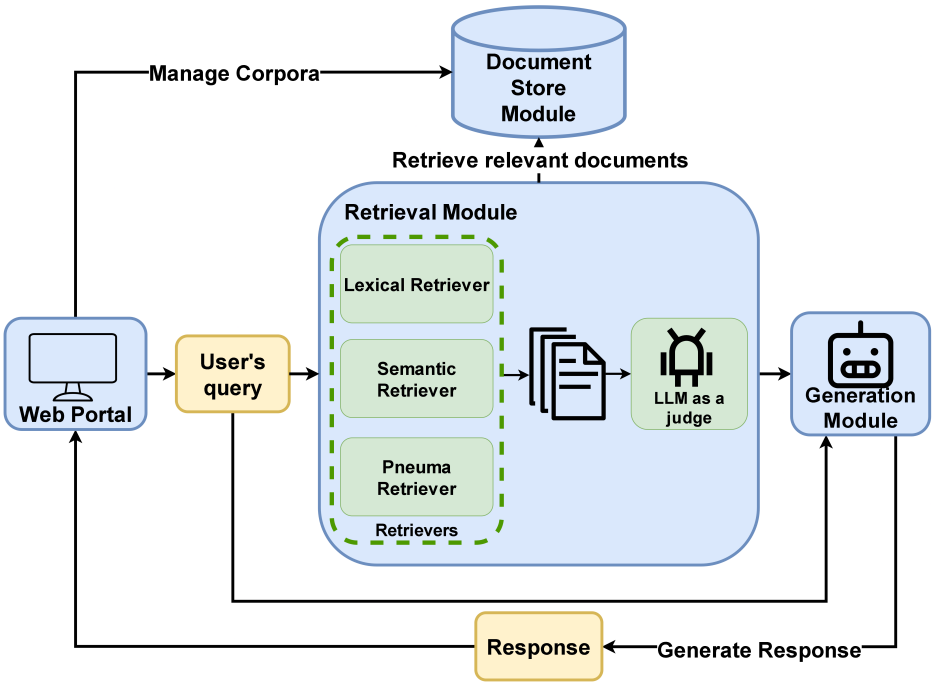


Figure 1: Overview of the proposed ESG-specific RAG framework.

hallucinations, a common phenomenon with language models, ensuring that the generated content accurately reflects the underlying ESG documents.

The **Document Store Module** serves as the central persistent storage for the framework's RAG pipeline, storing the huge amount of ESG data extracted from user-uploaded reports. This module is designed to accommodate the diverse and complex nature of ESG information, which includes unstructured textual data, structured tables, and metadata that provide essential context about each document. In our demonstrated implementation of the framework, the **Document Store** module leverages PostGreSQL - a relational database - as the storage backend for its robust support for relational data and vector search extensions. However, the framework is designed to be flexible. Some ESG systems may require working with unstructured data such as charts, figures, and infographics summarizing sustainability metrics. In such cases, different data models such as object storage systems can additionally be employed to effectively manage different types of data.

Finally, the **Retrieval Module** is responsible for bridging the stored ESG data and the Generation Module by identifying and fetching relevant information based on user queries. It is designed with flexibility in mind to accommodate diverse search needs and data types, which will be further detailed in the following subsection on handling heterogeneous ESG data.

3.2 Handling Heterogeneous ESG Data

The proposed framework is designed with flexibility at its core, enabling a plug-and-play approach when it comes to the retrieval strategies in the **Retrieval Module**, particularly different retrieval strategies can be employed tailored to the complexities of ESG data. In this paper, we showcase different retrieval strategies to handle tabular data for ESG reports:

- Standard Retrieval: Standard Retrieval employs both vector search and keyword search on chunks of tabular data stored within the database to identify potentially relevant information. To enhance precision further, the retrieved results are then passed through an LLM-based reranker acting as a judge that evaluates and reorders the candidates based on their contextual relevance to the user query.
- Pneuma retrieval: The method described in [2], named Pneuma, enhances tabular retrieval by leveraging LLMs during the ingestion phase to summarize the column structure and representative row samples of tables. When tables are first processed, Pneuma uses LLMs to generate concise summaries that capture the schema and relationships within the data. These summaries are then utilized during the retrieval phase to match user queries with the most relevant tables, improving retrieval accuracy for complex tabular ESG data.

3.3 Experiment Setup

We implement the proposed framework as a demonstrative system to conduct experiments and evaluate its effectiveness in handling ESG data. For this purpose, we collect a set of publicly available ESG reports from top companies in Australia including Transurban, Coles, and Vicinity Center. These reports include data with types ranging from unstructured texts to tabular data detailing environmental, social, and governance metrics. These data are processed and stored in our implementation of the **Document Store Module**,

which uses PostgreSQL together with the pgvector extension as the backend.

The Generation Module leverages the GPT-4.1 model to generate context-aware responses grounded in the retrieved ESG data chunks through the use of prompt engineering.

For the **Retrieval Module**, we implement two versions for evaluation purposes as described in 3.2. The first version employs a standard retrieval strategy combining vector search and keyword search with an LLM-based reranker to identify and rank relevant ESG data chunks. The second builds on top of the first by integrating Pneuma [2] into the system; specifically, it uses a combination of vector search, keyword search, and Pneuma search, followed by an LLM-based reranker to further refine the relevance of retrieved results.

This setup allows us to assess the system as a whole as well as comparing the effectiveness of standard retrieval methods against a more sophisticated, structure-aware approach tailored for tabular ESG data. To evaluate the effectiveness of the system in handling ESG related use cases and to compare the effectiveness of the different retrieval strategies, we formulate a set of 20 questions designed to reflect realistic use cases based on the collected ESG reports. These questions range from simple definitions to more complex analytical queries, such as key sustainability metrics identification, comparing sustainability metrics between different reporting years, comparing sustainability reports from different firms, extracting exact reported values like natural gas consumption or total energy consumption, and high-level tasks such as drafting a sustainability report that follows recognized standards such as GRI or TCFD.

4 Results & Discussions

Testing results show that the overall system can handle queries of different complexity to accommodate a diverse range of ESG-related information needs. For simple definition-based queries such as "What is the Global Reporting Initiative?" or "What are the SDGs?, the system retrieves precise and relevant information obtained from the stored documents. Similarly, comparison queries like "Compare the Global Reporting Initiative vs the TCFD" are handled effectively with the system highlighting clear differences in purpose, scope, and structure of each framework. A common ESG use case involves comparing performance across time or between organizations. Queries such as "Compare between Vicinity center sustainability report in FY23 vs FY22" or "Compare between Coles's sustainability report and Transurban's" demonstrate the system's ability to relate and contrast ESG information across different reporting periods and organizations from multiple reports to highlight key differences.

When focusing exactly on exact data point extraction with queries such as "Citylink total energy consumption in FY23" or "Lane Cove Tunnel renewable electricity purchases in FY2023," the standard retrieval method falls short, failing to pinpoint the exact data point in the documents as shown in Figure 2. One possible reason is that due to chunking, the relationship between a numerical value and its corresponding column header may be lost, making it difficult for the retrieval process to correctly retrieve the relevant figure for the user's query. The methods described in [2] named Pneuma were able to retrieve the relevant tabular documents and output relevant

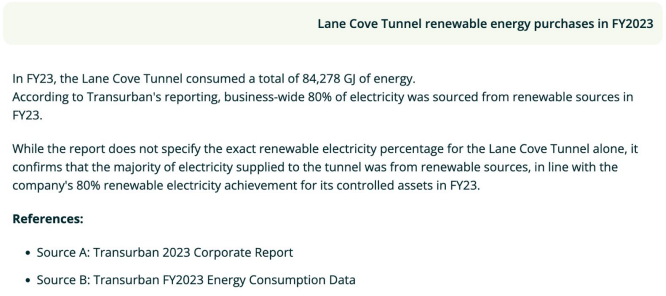


Figure 2: Answers from the model with the standard retrieval for the query "Lane Cove Tunnel renewable electricity purchases in FY2023."

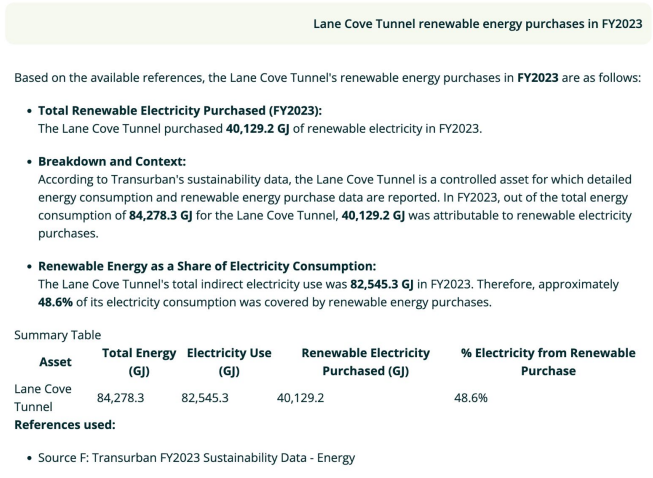


Figure 3: Answers from the model with Pneuma retrieval for the query "Lane Cove Tunnel renewable electricity purchases in FY2023."

information regarding the user's query as shown in Figure 3. However, a notable limitation of this approach is that it requires passing the entire document into the LLM, which raises issues regarding the context length of LLMs, highlighting that though this approach is more effective in retrieving tabular data, it is not scalable when it comes to downstream reasoning tasks.

We demonstrated our question answering system to several expert groups, including business users and government environmental units, in December 2024 in China. Qualitative feedback using Australian datasets shows great potential for the deployment of QA-ESG in businesses.

5 Conclusion and Future Works

As sustainability reporting becomes more essential and more complex, the ability to navigate and interpret ESG data fragmented across unstructured text and intricate tables has become a critical

need. The application of AI in the sustainability domain, particularly in the ESG reporting field, offers promising opportunities to improve sustainability practices. As demonstrated in our proposed framework, AI specifically LLMs combined with retrieval-augmented techniques, can play a crucial role in addressing the challenges of ESG reporting. Through our experiment of implementing the framework with data gathered from the top companies in Australia, our system shows robust capabilities in supporting a variety of query types, including definition-based queries, comparison-based queries, task-based queries, and data extraction queries, though with some limitations.

While our studies demonstrate the potential of leveraging AI in sustainability reporting, there are still significant rooms for improvement. Future work should experiment with different LLMs and incorporate different retrieval techniques to accommodate for a diverse set of ESG use cases. Further testing with a more diverse set of business users would better capture real-world usage scenarios, expectations, and challenges. In addition, incorporating multimodal RAG to interpret prevalent visual elements in ESG reports such as charts and figures, would enable support for additional use cases such as extracting insights from graphs and comparing visualized figures across reports.

To conclude, this research demonstrates the strong potential of AI to enhance ESG reporting in businesses. While there are still rooms for advancement, the research sets the foundation for promising future innovations that can alter how organizations manage and communicate their sustainability efforts.

6 Acknowledgements

This publication has emanated from research supported in part by grants from Research Ireland under Grant [12-RC-2289-P2] and [18/CRT/6223] which is co-funded under the European Regional Development Fund. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

[1] Noor Hidayah Ab Aziz and Sajead Mowafaq Alshdaifat. 2024. *ESG Reporting: Impacts, Benefits and Challenges.* Springer Nature Singapore, Singapore, 69–76. doi:10.1007/978-981-97-2981-4_5

[2] Muhammad Imam Luthfi Balaka, David Alexander, Qiming Wang, Yue Gong, Adila Krisnadhi, and Raul Castro Fernandez. 2025. Pneuma Leveraging LLMs for Tabular Data Representation and Retrieval in an End-to-End System. *Proceedings of the ACM on Management of Data* 3, 3 (June 2025), 1–28. doi:10.1145/3725337

[3] Asim Biswal, Liana Patel, Siddarth Jha, Amog Kamsetty, Shu Liu, Joseph E. Gonzalez, Carlos Guestrin, and Matei Zaharia. 2024. Text2SQL is Not Enough: Unifying AI and Databases with TAG. arXiv:2408.14717 [cs.DB] https://arxiv.org/abs/2408.14717

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL] https://arxiv.org/abs/1810.04805

[5] Renzo Mori Junior, Peter J. Best, and Julie Cotter. 2014. Sustainability Reporting and Assurance: A Historical Analysis on a World-Wide Phenomenon. *Journal of Business Ethics* 120, 1 (2014), 1–11. doi:10.1007/s10551-013-1637-y

[6] KPMG International. 2024. *The Move to Mandatory Reporting: Survey of Sustainability Reporting 2024.* Technical Report. KPMG International. https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2024/11/the-move-to-mandatory-reporting-web-copy.pdf.coredownload.inline.pdf Accessed: 2025-07-23.

[7] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401 [cs.CL] https://arxiv.org/abs/2005.11401

[8] Mingyang Mao, Mariela M. Perez-Cabarcas, Utteja Kallakuri, Nicholas R. Waytowich, Xiaomin Lin, and Tinoosh Mohsenin. 2025. Multi-RAG: A Multimodal Retrieval-Augmented Generation System for Adaptive Video Understanding. arXiv:2505.23990v2 [cs.AI] https://arxiv.org/abs/2505.23990

[9] Srishti Mehra, Robert Louka, and Yixun Zhang. 2022. ESGBERT: Language Model to Help with Classification Tasks Related to Companies' Environmental, Social, and Governance Practices. In *Embedded Systems and Applications (EMSA 2022).* Academy and Industry Research Collaboration Center (AIRCC), 183–190. doi:10.5121/csit.2022.120616

[10] Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Tali Bers, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M. Rush. 2022. Multitask Prompted Training Enables Zero-Shot Task Generalization. arXiv:2110.08207 [cs.LG] https://arxiv.org/abs/2110.08207

[11] Santi Setyaningsih, Rosita Widjojo, and Peter Kelle. 2024. Challenges and opportunities in sustainability reporting: a focus on small and medium enterprises (SMEs). *Cogent Business & Management* 11, 1 (2024), 2298215. arXiv:https://doi.org/10.1080/23311975.2023.2298215 doi:10.1080/23311975.2023.2298215

[12] Brandon Smock, Rohith Pesala, and Robin Abraham. 2021. PubTables-1M: Towards comprehensive table extraction from unstructured documents. arXiv:2110.00061 [cs.LG] https://arxiv.org/abs/2110.00061

[13] United Nations. n.d.. Sustainable Development Goals. https://sdgs.un.org/goals. Accessed: 2025-07-23.

[14] Hangdi Xing, Feiyu Gao, Rujiao Long, Jiajun Bu, Qi Zheng, Liangcheng Li, Cong Yao, and Zhi Yu. 2023. LORE: Logical Location Regression Network for Table Structure Recognition. arXiv:2303.03730 [cs.CV] https://arxiv.org/abs/2303.03730

[15] Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. 2023. FinGPT: Open-Source Financial Large Language Models. arXiv:arXiv:2306.06031v1 [q-fin.ST] https://arxiv.org/abs/2306.06031

[16] Yi Zou, Mengying Shi, Zhongjie Chen, Zhu Deng, ZongXiong Lei, Zihan Zeng, Shiming Yang, HongXiang Tong, Lei Xiao, and Wenwen Zhou. 2023. ESGReveal: An LLM-based approach for extracting structured data from ESG reports. arXiv:2312.17264 [cs.CL] https://arxiv.org/abs/2312.17264