Kalpa Publications in Computing

Volume 22, 2025, Pages 612–622

Proceedings of The Sixth International Conference on Civil and Building Engineering Informatics

Sustainability Management Method for Construction Projects Data Using Large Language Model

Xingbo Gong[1], Yuqing Xu[1], Helen H.L. Kwok[3], Xingyu Tao[2], and Jack C.P. Cheng[1]

[1] Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Hong Kong, S.A.R.

[2] Department of Building and Real Estate, Hong Kong Polytechnic University, Hong Kong S.A.R.

xgongai@connect.ust.hk

Abstract

As sustainable development is gaining more and more attention, the construction industry continues to explore this aspect. Both Sustainable Development Goals (SDGs) and Environmental, Social and Governance (ESG) provide management goals for corporate sustainable development and assessment. However, due to the complexity of construction events and multiple data sources, sustainable development management in the construction industry is still hindered by the need for a large amount of labor costs. Therefore, this paper proposes an LLM-based sustainable development data processing framework for construction, which achieves three goals: (1) identifying indicators of SDGs and ESG assessment frameworks for construction projects, (2) mapping sustainable development indicators to construction events and data, and (3) developing an LLM-based localized data processing framework for construction sustainability. The proposed method can achieve rapid data processing of construction projects and provide information and information sources related to sustainable development goals. It realizes automated report generation or correlation traceability of sustainable development in construction projects.

1 Introduction

As a major consumer of resources and energy, the construction industry has always been one of the core industries for sustainable development. Since 2015, when governments around the world agreed to implement a set of sustainable development goals (SDGs), it has provided guidance for the construction industry to promote sustainable development from environmental, social and economic perspectives (Pradhan et al., n.d.). The United Nations 2030 Agenda covers many environmental,

I. Hkust, J. Cheng and Y. Yantao (eds.), ICCBEI 2025 (Kalpa Publications in Computing, vol. 22), pp. 612–622

economic and social development issues, such as health and well-being, poverty, hunger, quality education, gender equality, climate action, water, sanitation, energy and environment, and peace and social justice. The 17 Sustainable Development Goals and 169 related targets contained in the SDGs proposed on the agenda provide guiding recommendations for the management of construction industry (Fei et al., 2021). The 2030 Agenda and its goals and targets also represent a long-term political framework for companies to contribute to sustainable development. However, although many companies have proposed their own sustainable development goals, the design of meaningful goals, sustainable strategies or key performance indicators has not been widely proposed. Corporate participation in the SDGs is limited and largely symbolic. This is partly due to the market's vague setting of performance evaluation indicators and scope related to corporate sustainability or SDGs (Johnsson et al., 2020).

Now, all large companies should understand their environmental and social footprint and disclose, and report related activities. This is achieved through a series of standards, frameworks and indicators related to the so-called ESG (Environmental, society and governance). Although development and environmental protection may be considered a task for the government, business plays an increasingly active role in achieving sustainable development goals (Chaitanya et al., 2016). These efforts and recognition come from legal and regulatory measures and market behaviors such as ESG investment. Sustainable business models have been proven to bring advantages to companies in the market and resist possible risks in the future (Banihashemi et al., 2017; De Franco, 2020). As an evaluation framework for corporate behavior, ESG can provide detailed indicators for corporate managers in a more detailed manner, so that companies can provide a basis for considering relevant factors such as the environment in management and operations. By exploring the correspondence between SDGs and ESG, it will help companies collect and share relevant data and provide clear quantitative indicators for qualitative SDGs goals.

In the assessment of the level of sustainable development of enterprises, ESG reports require a lot of data support. In the construction industry, these data come from the company's construction projects, the company's own operations and the environment of the construction area (Gong et al., 2024). At the same time, a large amount of information throughout the life cycle of each project needs to be exhaustively counted and processed to meet the requirements of ESG evaluation and sustainable development reporting (Mishra, 2023). The huge workload requires a lot of labor to process this information. This hinders the enthusiasm of enterprises to evaluate and manage the ESG performance of construction projects. So far, large language models (LLMs) have made significant breakthroughs in many automated learning tasks by using generative pre-trained transformers (GPTs) to perform unsupervised training on large corpora and datasets (Pu et al., 2024). These tasks include machine translation, text classification, intelligent question answering, etc. LLMs can accurately extract entities and entity relationships through the ability of language understanding. These characteristics have also led to some expansions and attempts in the construction industry, such as building energy analysis and urban information management.

However, the current LLM can only process limited information. A large amount of information on construction projects has little to do with sustainable development, and this additional information will affect the operation of LLM (Giudici et al., 2023). Which data and construction events are related to SDGs and ESG, and which sustainable development goals are related to these construction events need to be further considered and analyzed. Therefore, to overcome the challenges of processing sustainable development-related data in construction projects and improve its automation in project management, this paper proposes a data processing framework for sustainable development goals of construction projects based on LLM. Through automated text information processing, LLM demonstrates efficiency advantages in corporate SDGs and ESG management.

2 Method

Figure 1 shows an overview of the proposed LLM-based approach for assessing the level of sustainable development in construction projects. This approach has three modules. The first step is the analysis of SDGs and ESG evaluation indicators. This will screen out sustainable development goals and ESG indicators related to the construction industry and construction projects. The second step is to establish the correspondence between events and data in construction projects and these indicators. The established data map will describe the relationship between sustainable development goals and construction processes and provide a basis for further data processing. The third step is to establish a localized database so that LLM can obtain construction data. And further structure the text information so that the automatically processed content can provide a reference for corporate sustainable development.

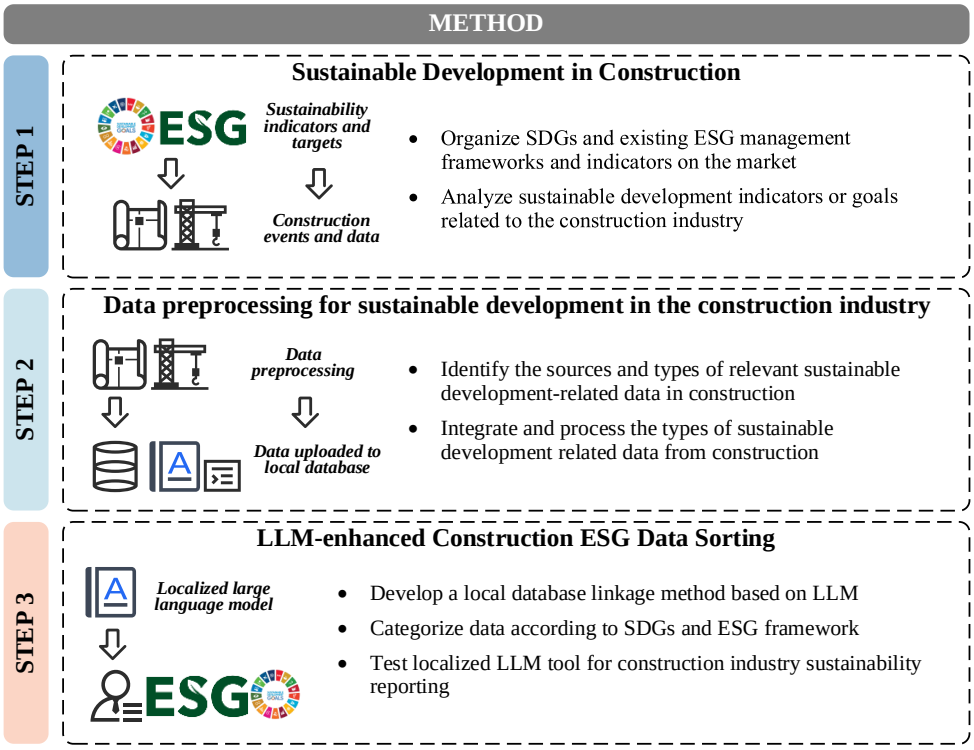

Figure 1: Method overview

2.1 Construction Sustainable Development Index

SDGs provide a new opportunity for the construction industry to expand its focus from the environmental aspect of sustainability. The behavior of enterprises and society is closely related to the SDGs during the design, construction, delivery and operation of construction projects. For example, resource utilization during construction, low-carbon model design during the operation phase, and pollution and recycling during building demolition activities. According to the BDG, the construction industry is related to each of the 17 SDGs. Among them, the more representative is that the building materials used in construction projects can directly affect the achievement of the SDGs, especially SDG

3, SDG 7, SDG 9, SDG 11, SDG 12, SDG 13 and SDG 15, and indirectly affect SDG 2, SDG 5, SDG 10 and SDG 16 (Fei et al., 2021). At the same time, ESG management of construction projects has also shown its importance today. The correlation between innovation and sustainable practices in construction projects and corporate performance has been widely studied. It can measure the performance of enterprises in these areas and serve investors and social participants. Both SDGs and ESG provide goals and foundations for the sustainable management of construction projects (Litvinenko et al., 2022). However, the general sustainable development goals and indicators are not accurate for the construction industry. Specific indicators or goals still need to be accurately identified to meet the management of relevant indicators in complex construction affairs.

Figure 2: SDGs and ESG in Construction Projects

As shown in Figure 2, this paper proposes a correspondence table between SDGs and ESG (MSCI) indicators. The selected ESG indicators are similar to the scope of SDGs, both of which are general and industry wide. However, they have different focuses in the construction industry. In this correspondence, only the most direct sustainable development goals, the relationship between ESG impacts and construction activities are considered. In terms of environmental indicators, SDG 6, SDG 7, SDG 9, SDG 11, SDG 12, SDG 13, SDG 14 and SDG 15 are all relevant. In terms of ESG indicators, they include carbon emissions, carbon footprint, water resources and waste. These goals and indicators are more relevant to construction projects. Similarly, in terms of social indicators, human resources, worker health and data security are more important in construction projects. In governance indicators, ESG management often follows the company's financial statements or annual accounting information, which is related to construction projects, but this information cannot be directly obtained in a single project. Therefore, these goals are not considered as the main consideration in the relevant affairs of construction projects.

2.2 Construction Sustainable Development Index

After determining the sustainable development indicators related to construction projects, Figure 3 shows the correspondence between these indicators and construction-related data and data sources. In the scope of construction projects, four data sources are identified, including pre-construction preparation documents, supply chain information of construction projects, construction site data and corporate reports. The sources of this information are complex and require a lot of manual processing. For example, carbon emission information has a high correlation with SDG 13, and this data comes not only from energy use (such as operational carbon), but also from the use of materials in construction projects (such as embodied carbon) (Xu et al., 2024). For other aspects such as biodiversity, land use and green technology data at the construction site, they often come from the environmental impact assessment and reports of the construction. By corresponding SDGs, ESG and construction project data, sustainable development assessment or environmental management personnel can more conveniently collect and manage relevant content.

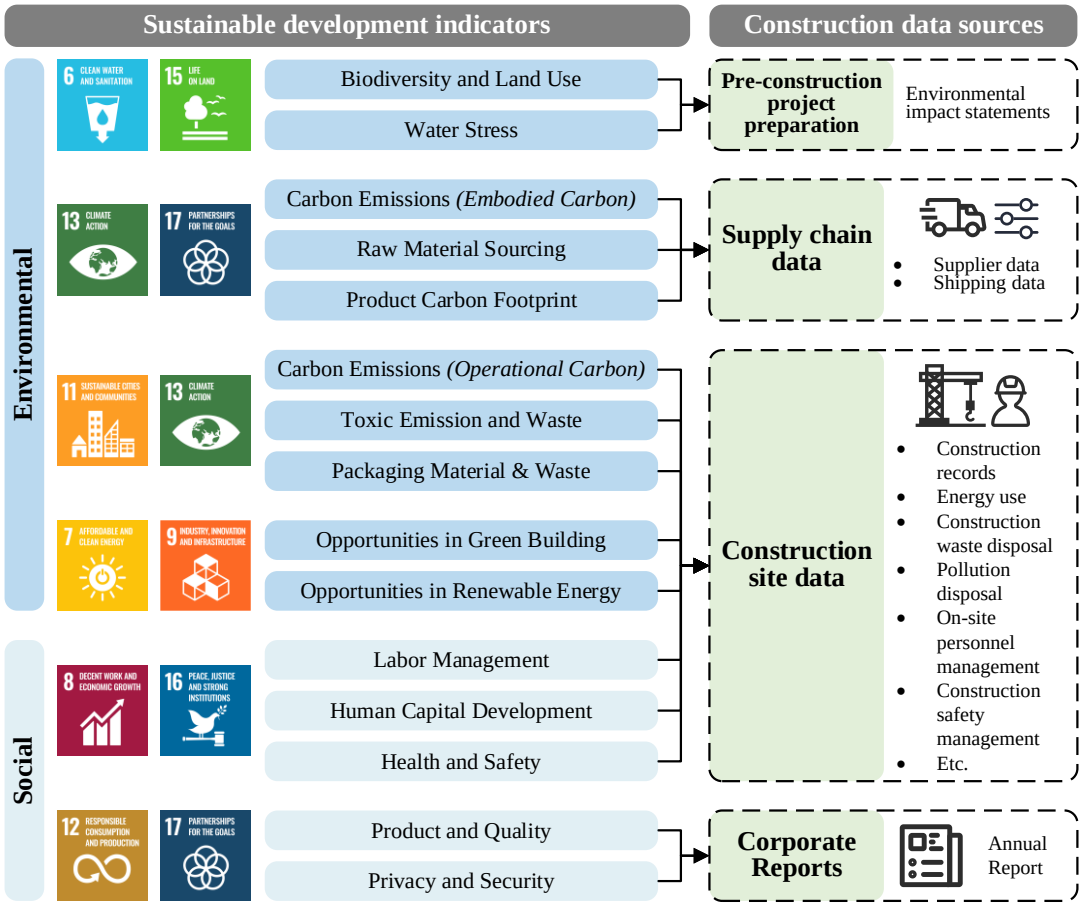

Figure 3: Data Map for SDGs and ESG indicators to Construction Data

2.3 LLM-based Data Processing of Sustainable Development of Construction

Environmental impact assessment, construction records and reports are directly related to the sustainable development management of construction projects. However, large amounts of text content require high labor costs to process. LLMs can process these contents quickly. Therefore, this paper links a local database with a localized large prediction model to achieve rapid organization and management of sustainable development affairs in construction projects. In the local database, reports and records in the construction project will be uploaded. At the same time, according to the correspondence between the goals and indicators and the construction data in Figure 3, the uploaded files will be attached with one or more indicators at the same time. Based on these data, an information collection and output program is developed for LLM. The pseudo code of the program is shown in Table 1.

Table 1. Pseudo Code for LLM and Sustainable Data Local Database Connection

Step 1: Connect to MySQL database
MySqlTool.connection = pymysql.connect(
→ host='*localhost*', database='*Sustainable data for Construction*'
Step 2: Execute query statement
cursor = connection.cursor()
Input: *Sustainable data indicator* or *File name* or *Upload time*
sql = f"SELECT * FROM ' *Sustainable data*' where indicator = {indicator} or name = '{query}' or filetime = '{query}'" → print("sql: ", sql)
cursor.execute(sql)
Step 3: Close the cursor object and database connection
cursor.close()
Step 4: Call langchain.tools import BaseTool
Call the above query statement from the Sustainable database
Step 5: Structured output
model_with_structure = model.with_structured_output(indicator name + file source)
structured_llm = llm.with_structured_output(*Sustainable*)
structured_llm.invoke("indicators related data")
Output: ResponseFormatter(answer="indicator name" + "file source".)
End

It can link the local MySQL database and LLM. This process is based on the linking mode of Langchain (Asyrofi et al., 2023). At the same time, during the LLM output process, the relevant sustainable development indicators and the data location in the construction project can be directly obtained through specific prompt words. This can help managers check the original information to ensure and correct the accuracy of the LLM output information.

3 Proof of Concept

After completing the above development, this paper verifies the designed construction project sustainable development data processing framework. The verification environment of the framework is to verify through a design scenario, in which the data in the project comes from an infrastructure construction process in Hong Kong. In this verification scenario, the environmental impact assessment report of the construction project is disassembled so that it can be directly applied to the designed construction sustainable development indicators. At the same time, the supply chain data in the construction project (including material suppliers, construction project location, transportation records and energy consumption), construction site data (human resource records, construction safety reports, material usage, energy and water usage, etc.) and the company's current operation report and strategic development report are also submitted, see Figure 4 (a). After the sustainable development indicators and documents and data are matched one by one, they are uploaded to the local database, see Figure 4 (b). After completing the above operations, the managers or evaluators of the sustainable development of the construction project can briefly obtain this information through the local LLM, as shown in Figure 4 (c). Similar to the pre-trained large language model connected to the normal network, the LLM linked to the local database can interact. At the same time, if the pre-designed prompt words are used, the model will obtain the text information of the construction project through the local database and output the content related to the sustainable development of the construction project. At the same time, the output of information sources can also be realized.

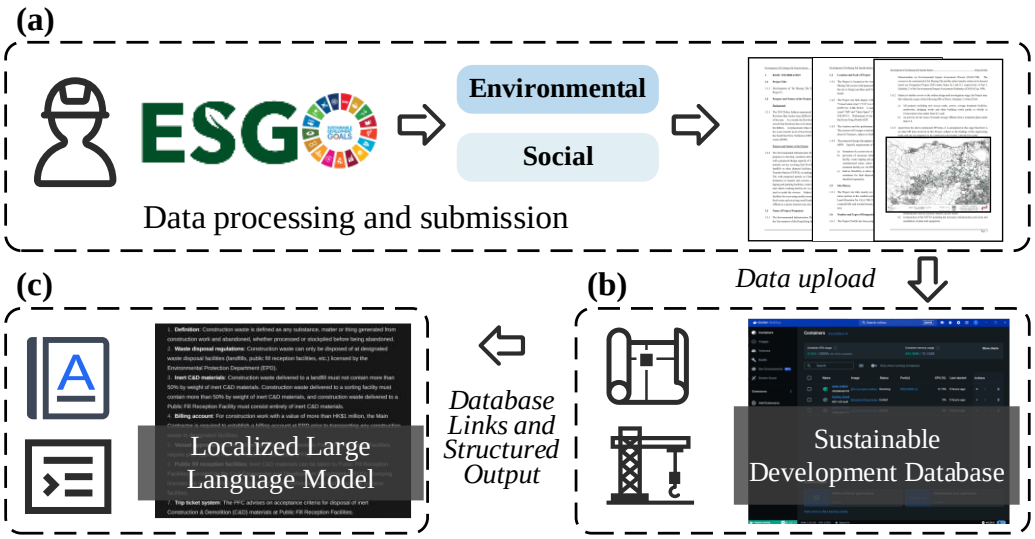

Figure 4: Verification scenarios: (a) Sustainable data collation of construction projects; (b) Data upload; (c) Data acquisition based on LLM

Through the above validation steps, the framework has been proven to be able to process and manage sustainable development related content in construction projects. For some sustainable development indicators, LLM can also organize and classify information based on the local construction database. For example, when asking about waste-related content in construction projects, the output content will include construction events of different types of waste, such as C&D, solid waste, and building waste. Through pre-designed prompt words, the model can also locate and output the document sources related to the indicators in the database. The efficiency of each output of the framework is also counted, as shown in Figure 5. The output efficiency is maintained within an acceptable range.

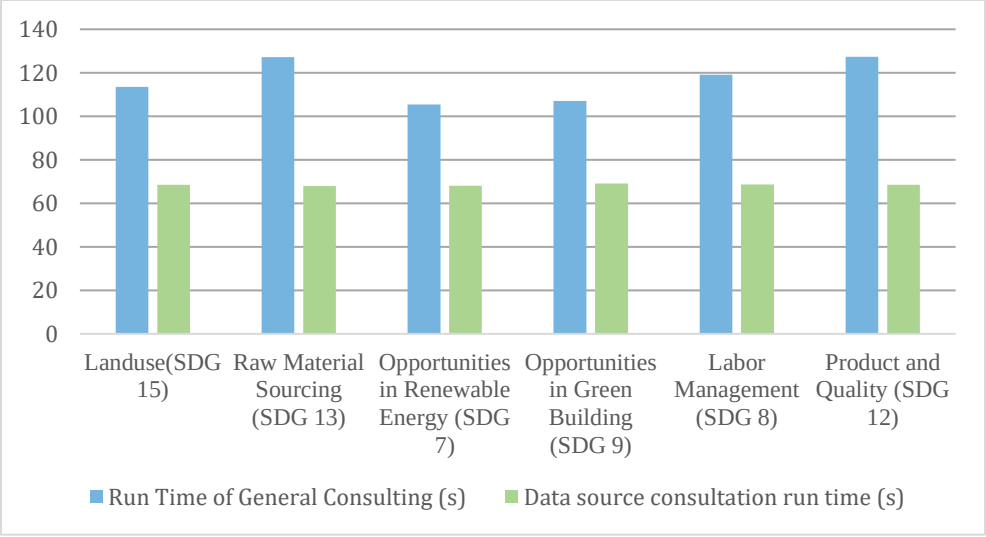

Figure 5: Average Runtime of Framework Validation

4 Discussion

In recent years, various industries have realized the importance of sustainable development on their impact, including the construction industry. This shift comes from the attention of investors, regulators and governments to related matters. The Sustainable Development Goals provide a new perspective for the construction industry. From this perspective, the global vision for sustainable development can be transformed into business solutions. In building management and operations, construction companies should emphasize collaborative ESG activities to gain a good public business reputation, such as reducing carbon dioxide emissions and construction waste and improving safety and health (Jain et al., 2020). The refinement of SDG goals and ESG assessment frameworks can help the construction industry deal with these issues in a more accurate way. At the same time, accurate management of corporate development also requires the promotion of technological development to reduce management costs and improve efficiency. LLM can help managers achieve this goal in the face of large and complex construction data.

The sustainable development data management framework for construction projects developed in this paper based on the above background can achieve rapid processing of complex text data to a certain extent. This relies on the ability of the pre-trained large language model to understand and deconstruct the text. At the same time, it provides a mitigation method for the situation that sustainable development management relies on a large number of professional people to carry out work. It effectively reduces the management cost of relevant data in construction projects. At the same time, the localized database and locally deployed LLM can avoid the risk of commercial data leakage. It provides support for the secure data management of construction projects.

However, the existing framework has only been validated in designed scenarios. For more complex and multi-party actual engineering projects and business operations, the method is not capable enough. It is necessary to develop a specialized training LLM model for the construction industry. This will not only enhance the efficiency of data exchange in the construction industry, but also help the industry achieve sustainable development goals. At the same time, for actual engineering projects, data in various formats are related to sustainable development assessments. For example, building information modeling (BIM) data, engineering drawing data, and even sensor data from construction sites and building operations are related to comprehensive sustainable development assessments of projects. The completer and more comprehensive the data, the more realistic it can show the situation and expected development of the construction project. This is helpful for dynamic management of sustainable development. Current methods often target a single data source. Further management models, such as the development of management frameworks and methods based on digital twins and machine learning, are necessary.

5 Conclusion

The SDGs proposed in the United Nations 2030 Agenda for Sustainable Development and the ESG assessment methods widely used in the market have put forward new goals and directions for environmental protection, social support and economic development of enterprises. The construction industry is one of the key industries for sustainable development because of its own pollution and high social relevance. To solve the difficulties caused by the large amount of complex data in construction projects for sustainable development management, this paper proposes a sustainable development data management framework for construction projects using LLM. It achieves three goals. First, this paper identified sustainable development goals for construction projects. This is conducive to accurately grasping the project's development focus on environmental protection and social development for the construction industry. At the same time, it mapped the relationship between sustainable development

indicators and data sources in the project to help managers more quickly identify the relationship between different indicators and construction events. Finally, a data management framework based on LLM was developed. After verification, the developed framework can realize the extraction and organization of sustainable development information of construction projects based on local databases.

However, the current methods still have limitations. First, for the existing data management model, the complex format in the construction project is not fully considered. In addition to text, structured data such as building information models, image data, etc. are also related to sustainable development assessment. How this multimodal data can be linked and interacted with LLM is still difficult. In addition, the current LLM model is usually universal across the industry, and it is often difficult to meet the specialized needs of the construction industry. Therefore, for sustainable development-related matters, the development and adjustment of LLM is in demand for the construction industry. In the future, some specific optimization methods such as vector database design and specialized report output will be studied.

References

Asyrofi, R., Dewi, M.R., Lutfhi, M.I., Wibowo, P., 2023. Systematic Literature Review Langchain Proposed, in: 2023 International Electronics Symposium (IES). Presented at the 2023 International Electronics Symposium (IES), pp. 533–537. https://doi.org/10.1109/IES59143.2023.10242497

Banihashemi, S., Hosseini, M.R., Golizadeh, H., Sankaran, S., 2017. Critical success factors (CSFs) for integration of sustainability into construction project management practices in developing countries. International Journal of Project Management 35, 1103–1119. https://doi.org/10.1016/j.ijproman.2017.01.014

Chaitanya, K., Aromar, R., Jessica, E., Holger, K., 2016. Getting Started with the SDGs in Cities | Environment & Urbanization. Sustainable Development Solutions Network 110.

De Franco, C., 2020. ESG Controversies and Their Impact on Performance. JOI 29, 33–45. https://doi.org/10.3905/joi.2019.1.106

Fei, W., Opoku, A., Agyekum, K., Oppon, J.A., Ahmed, V., Chen, C., Lok, K.L., 2021. The Critical Role of the Construction Industry in Achieving the Sustainable Development Goals (SDGs): Delivering Projects for the Common Good. Sustainability 13, 9112. https://doi.org/10.3390/su13169112

Giudici, M., Abbo, G.A., Belotti, O., Braccini, A., Dubini, F., Izzo, R.A., Crovari, P., Garzotto, F., 2023. Assessing LLMs Responses in the Field of Domestic Sustainability: An Exploratory Study, in: 2023 Third International Conference on Digital Data Processing (DDP). Presented at the 2023 Third International Conference on Digital Data Processing (DDP), pp. 42–48. https://doi.org/10.1109/DDP60485.2023.00019

Gong, X., Tao, X., Zhang, M., Xu, Y., Kwok, H.H.L., Dai, J., Cheng, J.C.P., 2024. Secure Environmental, Social, and Governance (ESG) Data Management for Construction Projects Using Blockchain. Sustainable Cities and Society 105582. https://doi.org/10.1016/j.scs.2024.105582

Jain, S., Singhal, S., Jain, N.K., Bhaskar, K., 2020. Construction and demolition waste recycling: Investigating the role of theory of planned behavior, institutional pressures and environmental consciousness. Journal of Cleaner Production 263, 121405. https://doi.org/10.1016/j.jclepro.2020.121405

Johnsson, F., Karlsson, I., Rootzén, J., Ahlbäck, A., Gustavsson, M., 2020. The framing of a sustainable development goals assessment in decarbonizing the construction industry – Avoiding "Greenwashing." Renewable and Sustainable Energy Reviews 131, 110029. https://doi.org/10.1016/j.rser.2020.110029

Litvinenko, V., Bowbrick, I., Naumov, I., Zaitseva, Z., 2022. Global guidelines and requirements for professional competencies of natural resource extraction engineers: Implications for ESG principles

and sustainable development goals. Journal of Cleaner Production 338, 130530. https://doi.org/10.1016/j.jclepro.2022.130530

Mishra, S., 2023. ESG Impact Type Classification: Leveraging Strategic Prompt Engineering and LLM Fine-Tuning, in: Chen, C.-C., Huang, H.-H., Takamura, H., Chen, H.-H., Sakaji, H., Izumi, K. (Eds.), Proceedings of the Sixth Workshop on Financial Technology and Natural Language Processing. Association for Computational Linguistics, Bali, Indonesia, pp. 72–78. https://doi.org/10.18653/v1/2023.finnlp-2.11

Pradhan P., Costa L., Rybski D., Lucht W., Kropp J.P., n.d. A Systematic Study of Sustainable Development Goal (SDG) Interactions. https://doi.org/10.1002/2017EF000632

Pu, H., Yang, X., Li, J., Guo, R., 2024. AutoRepo: A general framework for multimodal LLM-based automated construction reporting. Expert Systems with Applications 255, 124601. https://doi.org/10.1016/j.eswa.2024.124601

Xu, Y., Tao, X., Das, M., Kwok, H.H.L., Liu, H., Kuan, K.K.L., Lau, A.K.H., Cheng, J.C.P., 2024. A blockchain-based framework for carbon management towards construction material and product certification. Advanced Engineering Informatics 61, 102242. https://doi.org/10.1016/j.aei.2023.102242