

# ESGReveal: An LLM-based approach for extracting structured data from ESG reports

Yi Zou <sup>a,1</sup>, Mengying Shi <sup>b,1</sup>, Zhongjie Chen <sup>a</sup>, Zhu Deng <sup>a</sup>, ZongXiong Lei <sup>a</sup>, Zihan Zeng <sup>c</sup>, Shiming Yang <sup>c</sup>, HongXiang Tong <sup>a</sup>, Lei Xiao <sup>a</sup> and Wenwen Zhou <sup>a\*</sup>

<sup>a</sup>Alibaba Cloud, Hangzhou, China

<sup>b</sup>Department of Earth System Science, Tsinghua University, Beijing, China

<sup>c</sup>Department of Environmental Science and engineering, Sun Yat-Sen University, Guangzhou, China

<sup>1</sup>These authors contribute equally.

## ARTICLE INFO

**Keywords:**

ESG

Data Extraction

Large Language Models

ChatGPT

## ABSTRACT

ESGReveal is a method introduced in this paper for the systematic extraction and analysis of Environmental, Social, and Governance data from corporate reports, designed to address the pressing need for consistent and accurate retrieval of ESG information. Using Large Language Models (LLM) combined with Retrieval Augmented Generation (RAG) techniques, ESGReveal includes an ESG metadata module for criteria queries, a report preprocessing module for building databases, and an LLM agent module for data extraction. The framework's effectiveness was evaluated using ESG reports issued by companies across 12 industries listed on the Hong Kong Stock Exchange in 2022. A carefully selected representative sample of 166 companies, based on industry distribution and market capitalization, provided a comprehensive assessment of ESGReveal's capabilities. The application of ESGReveal yielded significant findings on the current state of ESG reporting, with GPT-4 achieving accuracy rates of 76.9% in data extraction and 83.7% in disclosure analysis, outperforming baseline models. These results suggest the framework's utility in improving the accuracy of ESG data analysis. Additionally, our analysis identified the need for more robust ESG practices, with environmental data disclosure at 69.5% and social data at 57.2%, indicating room for increased corporate transparency. Recognizing the current limitations of ESGReveal, including its inability to interpret pictorial data which is a feature planned for future enhancement, the study also suggests additional research to improve and compare the differential analytical performance of various LLMs. In conclusion, ESGReveal marks a step forward in ESG data processing, providing stakeholders a tool to better assess and enhance corporate sustainability practices. Its development shows promise in advancing the transparency of corporate reporting and contributing to the wider objectives of sustainable development.

## 1. Introduction

Since the United Nations Global Compact initiated the concept of Environmental, Social, and Governance (ESG) in 2004, corporations have utilized ESG reporting to show their initiatives and commitment within these domains (Tsang et al., 2023). Across the globe, myriad ESG reporting frameworks, notably the Global Reporting Initiative (GRI) and the Sustainability Accounting Standards Board (SASB), have gained widespread acceptance. Additionally, numerous stock exchanges have enacted ESG disclosure directives to guide corporate reporting practices.

ESG disclosure quantifies corporate transparency and is crucial for assessing performance in ESG aspects, providing a basis for decision-making for investors and other stakeholders (Bui et al., 2020). While investors and analysts can access ESG reports through stock exchanges and corporate websites, the vast number and varied formats of these reports pose significant integration challenges for consolidating disclosure data at the corporate or industry level (Doe, 2021). Third-party ratings such as those from MSCI, Sustainalytics, and Bloomberg help understand corporate non-financial performance but lack the necessary transparency and detailed

disclosure metrics (Abhayawansa and Tyagi, 2021; Schieemann and Tietmeyer, 2022). Research indicates that there is currently no publicly available ESG disclosure database covering detailed metrics, which limits the depth of analysis and regulation of corporate ESG performance to some extent (Clarkson et al., 2019).

In response to these impediments, this study introduced an approach, designated as ESGReveal, based on the integration of advanced Large Language Models (LLM) and Retrieval Augmented Generation (RAG) techniques. The ESGReveal is comprised of a tripartite architecture: ESG metadata module, report preprocessing module, and LLM agent module. Furthermore, an analytical exploration utilizing ESG reports from preeminent corporations across sectors listed on the Hong Kong Stock Exchange in year 2022 was conducted. The main contributions of this research include:

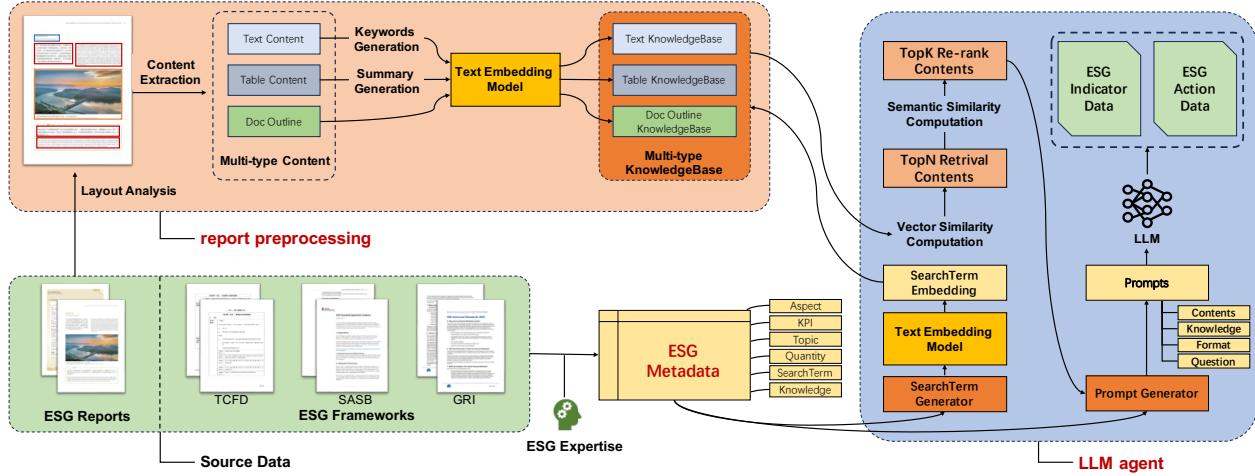
- Designing and executing ESGReveal for systematic extraction of crucial numerical and textual data from corporate ESG reports.

- Assessing the performance of different large language models in ESG information retrieval, setting a baseline for further ESG data processing and analytical studies.

\*Corresponding author.

E-mail addresses: zhoubuo.zww@alibaba-inc.com (W.Z. <sup>a</sup>)

ORCID(s):



**Figure 1:** Overall structure of ESGReveal.

- Utilizing ESGReveal to evaluate ESG reports from a representative subset of companies on the HKEx, providing industry benchmarks for ESG conduct and reporting.

## 2. Related Work

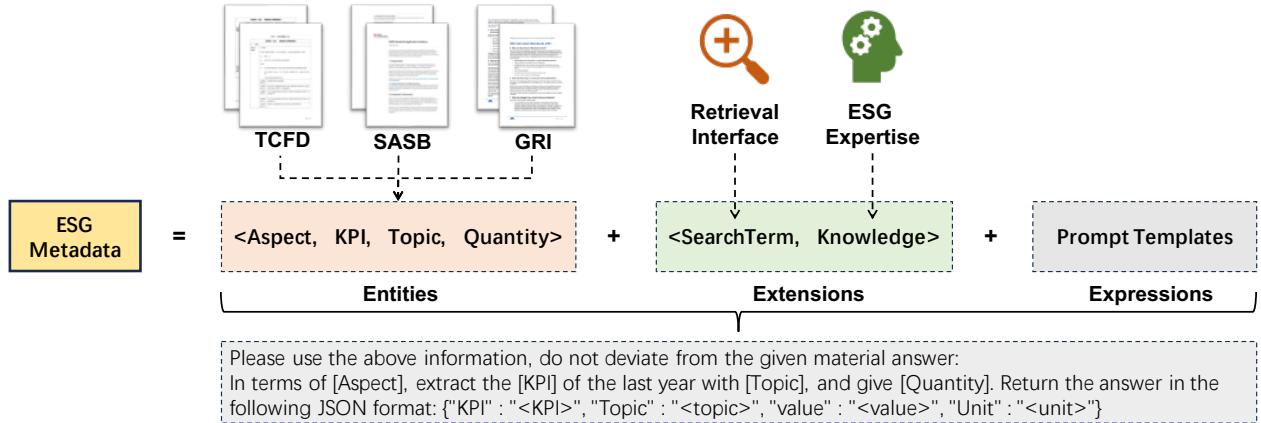
In the face of the complexity of ESG disclosure and the rapid growth in information volume, researchers have begun to explore the potential of natural language processing (NLP) techniques to enhance the precision and efficiency of ESG data extraction and analysis. Ruberg et al. (2023) utilized the BERT architecture to automatically classify content in corporate ESG reports relevant to the GRI standards, increasing the analytical efficiency for ESG assessments at the Brazilian Development Bank. Perazzoli et al. (2022) conducted an extensive analysis of 55,000 publications on ESG-related topics using NLP techniques. Luccioni et al. (2020) developed the ClimateQA model based on NLP technology to analyze information related to climate change in financial reports. Raman et al. (2020) detected ESG trends by analyzing corporate earnings call transcripts and parsing the linguistic structure within ESG texts. Fischbach et al. (2022) created a tool named ESG-Miner, which automatically extracts ESG-related information from media headlines and calculates ESG scores for companies. In these studies, advanced NLP models like BERT have made significant progress in understanding ESG contexts and semantics (Liu et al., 2023; Lee et al., 2022), yet there remain notable limitations in areas such as adaptability to the ESG sector, custom datasets, deep information mining, and multilingual and cross-cultural adaptability (Pasch and Ehnes, 2022; Fischbach et al., 2022).

In the field of natural language processing, recent developments in large language models such as GPT-3.5 (OpenAI, 2023b) and GPT-4 (OpenAI, 2023a). To address challenges with lengthy contexts and data currency in LLM, researchers have integrated RAG paradigm (Lewis et al., 2020) with LLM, which improves response precision with dynamic

information retrieval. Current research reveals the substantial potential of LLM in identifying sustainability goals and performance indicators within ESG reports (Zhang and Zhang, 2023; Burnaev et al., 2023). For instance, Kim et al. (2023) used ChatGPT to summarize the economic utility of corporate disclosures and revealed the link between information "bloat" and adverse capital market consequences. Ni et al. (2023) developed the CHATREPORT system based on LLM, which analyzed 1015 corporate sustainability reports using the Task Force on Climate-Related Financial Disclosures (TCFD) framework (<https://www.fsb-tcfd.org/>) and assessed their compliance. Bronzini et al. (2023) extracted semantically structured ESG-related information from sustainability reports using LLM, unveiling inter-corporate ESG action correlations. Moodaley and Telukdarie (2023) enhanced the ability to identify green claims and detect "green-washing" by training LLM with sustainability-related textual corpora. Yang et al. (2023)'s FinGPT showed promising application potential in recognizing information pertinent to environmental, social, and governance issues. The evolution of LLM has greatly optimized text processing methods, making the transformation of unstructured documents in ESG reports into structured data a feasible reality (Visalli et al., 2023).

Building on the foundation of existing research, this paper presents ESGReveal, a novel methodology that focuses on the extraction of data from ESG reports with a particular emphasis on the structured retrieval of numerical information. Our approach harnesses the sophisticated text processing capabilities of LLMs within RAG paradigm to systematically decompose both numerical and textual information contained in ESG reports from various standard frameworks. This systematic analysis facilitates deeper insights into ESG disclosure practices.

Furthermore, in recognition of the pivotal role played by the Hong Kong Stock Exchange (HKEx) in advancing ESG initiatives in China, this study employs ESGReveal to evaluate the ESG reports of companies listed on the HKEx.



**Figure 2:** Structure of ESG metadata module: Entities, Extensions, and Expressions.

Through this application, we aim to analyze the prevailing state of industry transparency and accountability in ESG, contributing towards a clearer understanding of the ESG landscape within the region.

### 3. Methods

In this section, we describe the architecture and components of ESGReveal. The general framework of ESGReveal is introduced in subsection 2.1. Following that, subsections 2.2, 2.3, and 2.4 elaborate on the primary modules: ESG metadata module, report preprocessing module, and LLM agent module.

#### 3.1. Overall

Figure 1 illustrates the three module design of ESGReveal. The ESG metadata module establishes a query framework that leverages ESG criteria and expertise to analyze reports. The report preprocessing module imports ESG reports and processes them to build a database for information retrieval. The LLM agent module then accesses this database to retrieve information and engages the LLM for data extraction, guided by the ESG metadata module. In the ESGReveal workflow, an ESG report is processed by the report preprocessing module and analyzed by the LLM agent module, resulting in data that adheres to the ESG specifications defined by the ESG metadata module.

#### 3.2. ESG Metadata Module

ESG metadata module is a metadata framework structured to comply with international ESG standards, such as GRI and SASB, and is segmented into three components: Entities, Extensions, and Expressions, as depicted in Figure 2. Entities outline the attributes of an indicator, detailing its category, description, subcategories, and measurement units. Extensions encompass `<Knowledge>` and `<SearchTerm>` to adapt to various ESG compilation standards. Expressions concentrate on crafting Prompt templates and formatting the output, which allow for the automatic generation of prompts to trigger the LLM. Subsequently,

this study detailed ESG metadata module's Entities and Extensions based on the *Supplementary 27 Environmental, Social and Governance Reporting Guide* (Hong Kong Exchanges and Clearing Limited, 2023b) and the *Supplementary 14 CORPORATE GOVERNANCE CODE* (Hong Kong Exchanges and Clearing Limited, 2023a) published by HKEx.

##### 3.2.1. Entities

The ESG metadata module's Entities simplify complex ESG issues into the format: `<Aspect, KPI, Topic, Quantity>`, with "KPI" denoting key performance indicators. Adhering to HKEx guidelines, the framework comprises 70 indicators, divided as 12/18/4 numerical and 14/15/7 textual indicators across the E/S/G categories, totaling 34 numerical and 36 textual indicators—detailed in Table ?? (for details, see Supplementary Table 1). Examples of Entities are as follows (for details, see Supplementary Table 2) :

Example 1: On the "A1. Emissions" issue, for querying total and intensity of waste emissions, `<Aspect, KPI, Topic, Quantity>` are "A1. Emissions", "Total waste produced (in tonnes)" and, where appropriate, "intensity", "Non-hazardous Waste, Hazardous Waste..." and "Absolute Values"

Example 2: On "A1. Emissions" issue, regarding key actions on emissions, `<Aspect, KPI, Topic, Quantity>` are "A1. Emissions", "Emissions target(s) and steps taken to achieve them", "Waste, Exhaust/Greenhouse Gases...", and "Key Actions"

##### 3.2.2. Extensions

To enhance adaptability across diverse ESG reporting standards, `<Knowledge>` and `<SearchTerm>` are integrated into ESG metadata module. The `<Knowledge>` component, containing domain knowledge curated by ESG experts and user-friendly explanations of indicators produced by sophisticated LLM, is embedded into LLM prompts via In-Context Learning to accurately carry out indicator extraction tasks. The `<SearchTerm>` component includes customized retrieval keywords gathered, for instance, from an extensive array of ESG reports based on the HKEx guidelines to

**Table 1**

Indicators of the ESG Metadata Module Based on HKEx Standards. The Full List of ESG Indicators Can Be Found in Supplementary Table 1.

Class	Number of indicators	Indicators
Environment (E)	12 quantitative indicators	Direct (Scope 1) greenhouse gas emissions (in tonnes) and, where appropriate, intensity.....
	14 quantitative indicators	Total workforce by gender,.....
Social (S)	18 quantitative indicators	Emissions target(s) and steps taken to achieve them,.....
	15 text indicators	Occupational health and safety measures adopted, and how they are implemented and monitored.....
Governance (G)	4 quantitative indicators	Number of executive and independent directors.....
	7 text indicators	Key actions on ESG governance, independent statements, diversified policies, performance evaluation procedures, and election process rules.....

improve the performance of the RAG retrieval. Illustrations of <Knowledge> and <SearchTerm> are delineated in Table ??, with comprehensive elaboration presented in Supplementary Table ??.

### 3.3. Report Preprocessing Module

To enhance the effectiveness of ESG indicator extraction, we have specially optimized the preprocessing procedures for ESG reports and the construction of the knowledge base.

#### 3.3.1. Preprocess of ESG Report

In the preprocessing of ESG reports, we initially employed advanced computer vision tools, including Microsoft's LayoutLMv3 (Huang et al., 2022) and GeoLayoutLM (Xing et al., 2023b), to extract the structural components of documents, such as headers, paragraphs, and tables. Subsequently, we conducted structural analysis using font characteristics from structural components to construct a report outline that connects to the text, thereby facilitating quicker information retrieval. Finally, recognizing the prevalence of numerical indicators in tables, we implemented algorithms like Microsoft's Table-Transformer (Smock et al., 2022) and LORE-TSR (Xing et al., 2023a) to accurately identify table cells and reconstruct table structures, ensuring the precise representation of tabular data.

#### 3.3.2. Construction of Multi-type Knowledge Base

After preprocessing, we organized ESG reports data into structured knowledge bases for textual contents, document outlines and table contents. For textual contents, we generated summaries using the model mt5 (Raffel et al., 2020) and created vector representations with the model m3e (Wang et al., 2023), storing them in vector databases like FAISS (Johnson et al., 2019) and Milvus (Wang et al., 2021). Document outlines were treated similarly, linking information with corresponding vectors for storage. For table contents, we employed a one-to-many mapping to

exclude non-essential details and highlight key information such as ESG indicator names. We then generated vector representations for these keywords, pairing each table with its vector list to establish a concise knowledge base for table data.

### 3.4. LLM Agent Module

#### 3.4.1. Retrieval of Knowledge

Following processing by the report preprocessing module, we enabled precise ESG metadata module-driven searches for ESG indicator-specific data from knowledge bases. We created vectors for queries generated by metadata using the same methods employed by the report preprocessing module and calculated their cosine similarity to those in the knowledge base. Vector retrieval helped us pinpoint the most pertinent findings from text, tables, and outlines. To improve retrieval accuracy and relevance, we applied the coROM (Zhang et al., 2022) model to assess the semantic similarity of the initial top matches, which allowed us to refine their order. Ultimately, we extracted the entries with the greatest similarity for reliable ESG indicator-specific knowledge.

#### 3.4.2. LLM Answering

After data retrieval, we created a prompt based on ESG metadata module and retrieved contents to extract indicator information. The prompt comprises the following elements: (1) Preset Information, which provides basic instructions for the LLM's behavior; (2) Reference Content, which includes the retrieved contents from the knowledge base; (3) Expert Knowledge, which incorporates ESG insights from specialists into ESG metadata module; (4) Question, which frames a targeted query for the indicator crafted from ESG metadata module, asking about the indicator's disclosure and related data; (5) Answer Format, which presents a structured request for the LLM to report data in the format of <Disclosure, KPI, Topic, Value, Unit, Target, Action>, covering disclosure status, ESG performance indicators, topics, numerical values, units, action targets, and key actions, respectively.

**Table 2**

Contents of the ESG Metadata Module. The Full List of ESG Metada Can Be Found in Supplementary Table 1.

Aspect	KPI	Topic	Quantity
A1. Emissions	Total waste produced (in tonnes) and, where appropriate, intensity (e.g. per unit of production volume, per facility).	[Non-hazardous Waste, Hazardous Waste]	[Absolute Values]
SearchTerm		Knowledge	
Construction waste, building rubbish, soil, rubble, organic waste, recyclable fertilizer, paper, plastic bottles, wood, household garbage		Hazardous waste typically includes the following categories: toxic chemical waste, electronic waste, hazardous gas waste, dangerous waste.	

For detailed prompt construction and response examples, see Supplementary Table 3

## 4. Datasets

We collected approximately 2249 ESG reports from 2022 issued by companies listed on HKEx (<https://www.hkex.com.hk/>), categorized into 12 industries according to the Hang Seng Industry Classification System (Hong Kong Exchanges and Clearing Limited, 2023b). As processing all ESG reports entailed a significant computational workload, we endeavored to select a representative sample of 166 companies, balancing industry diversity and market capitalization, for evaluating ESGReveal and performing disclosure analysis. ESG datasets for different industries are shown in Table ???. For more details about industries and companies, please refer to Supplementary Table 4.

## 5. Results

### 5.1. Assessing ESGReveal: Accuracy Indicators, LLM Benchmarking and Ablation Study

ESG disclosure frameworks encompass a wide range of reporting indicators, broadly categorized into quantitative (e.g., greenhouse gas emissions) and qualitative (e.g., employee diversity policies). Due to difficulties in quantifying the latter, this study focused on the precision of ESGReveal in extracting quantitative indicators as specified by ESG metadata module. We detailed the accuracy assessment method in Section 5.1.1, compared various LLMs' performances in Sections 5.1.2, and discussed ablation study results using the ESG metadata module framework in Section 5.1.3. This study utilized the ESGReveal tool to dissect selected sample reports, extracting critical information on ESG data disclosures, quantitative figures, and key initiatives, with detailed findings presented across sections 5.2.1 to 5.2.3.

#### 5.1.1. Accuracy Indicators

Utilizing ESGReveal, this study extracted the degree of indicator disclosure and their specific numerical values from sample reports under the HKEx framework. To quantitatively assess the accuracy of data extraction, we calculated

the disclosure coverage accuracy ( $Acc_{DC}$ ) and the data extraction accuracy ( $Acc_{DE}$ ). Initially, we manually annotated the numeric indicators within the sample reports to obtain the disclosure status  $D_{label}$  and specific values  $V_{label}$  for each company's indicators.

$$D_{label} = \{d_i | i \in \{1, 2, 3, \dots, N_{mq}\}, d_i \in \{1, 0\}\} \quad (1)$$

$$V_{label} = \{v_j | j \in \{1, 2, 3, \dots, N_v\}\} \quad (2)$$

Here,  $N_{mq}$  represents the total number of indicators defined in ESG metadata module, with the disclosure status of each indicator  $d_i$  being marked as "1" (disclosed) or "0" (not disclosed).  $N_v$  denotes the total number of numeric indicators that have been actually disclosed, with the specific value of each indicator  $v_j$  represented in JSON format.

Subsequently,  $Acc_{DC}$  is defined as the ratio of the number of disclosure indicators correctly identified by ESGReveal to the number of indicators actually disclosed in the reports. Similarly, data extraction accuracy  $Acc_{DE}$  is defined as the ratio of the number of numeric values correctly output by ESGReveal to the number of true numeric values in the sample reports.

$$Acc_{DC} = \frac{1}{N_{mq}} \sum_{i=1}^{N_{mq}} 1(d_i = \hat{d}_i), \quad (3)$$

where  $d_i \in D_{label}$  and  $\hat{d}_i \in D_{ESGReveal}$

$$Acc_{DE} = \frac{1}{N_v} \sum_{j=1}^{N_v} 1(v_j = \hat{v}_j), \quad (4)$$

where  $v_j \in V_{label}$  and  $\hat{v}_j \in V_{ESGReveal}$

Here,  $D_{ESGReveal}$  and  $V_{ESGReveal}$  represent the disclosure status of indicators and specific values extracted by ESGReveal, respectively.

These accuracy metrics are calculated as the average across the studied samples in order to evaluate the overall performance of ESGReveal.

**Table 3**

Sectors and firms of ESG reports. The Full List of ESG Reports Can Be Found in Supplementary Table 4.

Sector	Firms	Number of reports
Consumer Staples	Budweiser Brewing Company APAC Limited,.....	14
Properties & Construction	Country Garden Holdings Company Limited,.....	15
Telecommunications	China Mobile Limited,.....	13
Consumer Discretionarys	ANTA Sports Products Limited,.....	18
Industrials	JD Logistics, Inc.,.....	13
Utilities	Kunlun Energy Company Limited,.....	12
Financials	China Construction Bank Corporation,.....	11
Energy	China National Offshore Oil Corporation,.....	16
Healthcare	Sihuan Pharmaceutical Holdings Group Ltd.,.....	14
Materials	China Hongqiao Group Limited,.....	16
Information technology industry	Meituan Corporation-W,.....	14
Conglomerates	Fosun International Limited,.....	10
Total	China National Building Material Company Limited,.....	166

### 5.1.2. LLM Benchmarking

Table 4 shows the performance of ESGReveal when extracting ESG data using different LLMs. For comparative analysis, four advanced models were chosen: GPT-3.5, GPT-4, ChatGLM (Du et al., 2022), and the QWEN (Bai et al., 2023) model.

The experimental results indicate that compared to  $Acc_{DC}$ , all LLMs generally showed a decline of 4.8% to 7.7% in  $Acc_{DE}$ . This consistent performance gap may reflect the inherent complexity of the data extraction task, which requires not only identifying the presence of disclosed information but also accurately extracting specific values and attributes. Particularly when dealing with the extraction of data on topics such as "Employee turnover rate by gender, age group and geographical region" the task involves detailed values for multiple age brackets. This requires that the LLM be capable of accurately processing and distinguishing between multiple data points.

In the performance comparison of different LLMs, GPT-4 achieved an accuracy of 76.9% in  $Acc_{DE}$  and 83.7% in  $Acc_{DC}$ , leading other models. QWEN followed with accuracy of 54.9% and 61.4%, respectively. Lastly, GPT-3.5 and ChatGLM performed similarly, with GPT-3.5 achieving 47.1%/51.9% and ChatGLM achieving 46.2%/53.9% accuracy rates. The superiority of GPT-4 in the analysis of ESG disclosures and the extraction of relevant indicators can be attributed to its enhanced model capacity and superior comprehension capabilities. GPT-3.5, QWEN and ChatGLM have equivalent capabilities and are inferior to GPT-4. These findings underscore the pivotal role that a model's capacity and interpretive proficiency play in augmenting its

effectiveness when applied to specialized domain-specific tasks.

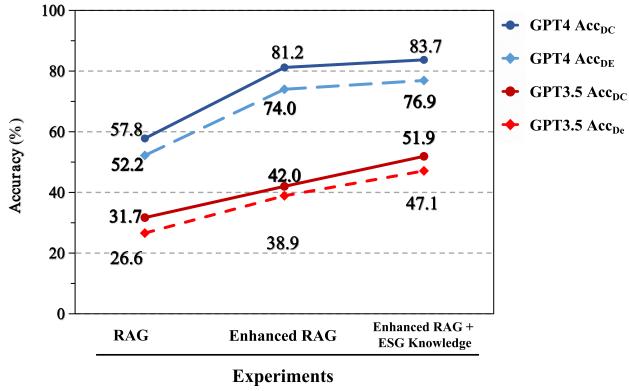
### 5.1.3. Ablation Study

Based on the experimental performance in Section 5.1.2, we selected GPT-3.5 and GPT-4 as representatives to conduct a series of ablation experiments to evaluate the effectiveness of each module in ESGReveal proposed in this study. We used a basic RAG implementation as the Benchmark (LangChain, Inc., 2023). As illustrated in Figure 3, under the Benchmark implementation, GPT-4 achieved accuracy of 57.8% in disclosure analysis and 52.2% in data extraction, while GPT-3.5 achieved 31.7% and 26.6%, respectively.

#### 1) Effectiveness of the Enhanced RAG

Building upon the Benchmark, we utilized the improved the report preprocessing module and LLM agent module, which significantly enhanced the performance of ESGReveal through better document preprocessing and content retrieval methods. As shown in Figure 3, marked as Enhanced-RAG, GPT-4 saw notable improvements in performance for both types of tasks, achieving 81.2% (+23.4%) for disclosure analysis and 74.0% (+21.8%) for data extraction. GPT-3.5 also brings improvement of 10.3% and 12.3%, reaching a performance of 42.0% and 38.9%, respectively. These results underscore the effectiveness of report preprocessing and LLM agent modules proposed in this study. Better structuring of documents and refined retrieval content significantly improve the performance of ESG disclosure analysis and data extraction under the RAG framework.

#### 2) Effectiveness Analysis of ESG Knowledge



**Figure 3:** Ablation Study of ESGReveal

With the introduction of <Knowledge> of ESG metadata module, both GPT-4 and GPT-3.5 achieved varying degrees of improvement. GPT-4’s accuracy increased to 83.7% (+2.5%) for disclosure analysis and 76.9% (+2.9%) for data extraction, whereas GPT-3.5 saw substantial increases of 9.9% and 8.2%, with accuracy reaching 51.9% and 47.1%, respectively. This indicates that by incorporating supplementary knowledge from the ESG domain, LLMs can experience different levels of enhancement in their parsing abilities. Especially for the LLMs with weaker model capacity and comprehension abilities (e.g., GPT-3.5), the accuracy gains are much greater than those for LLMs with stronger capacities and comprehension (e.g., GPT-4). This implies that the addition of ESG domain knowledge in ESG Metadata can greatly enhance the performance of less capable LLMs in ESG disclosure analysis and data extraction tasks.

## 5.2. Dissecting ESG Reports: Disclosure, Metrics, and Key Actions

This study applied ESGReveal to parse selected sample reports, thereby obtaining information on ESG data disclosure, related quantitative data, and key actions within the reports. Subsequently, in Sections 5.2.1 to 5.2.3, we described and conducted a preliminary analysis of the extracted data, in order to understand the specific performance of the sample reports in terms of ESG aspects.

### 5.2.1. Disclosure

Guided by the HKEx framework, this study designed 34 numerical indicators, specifically, the environmental dimension includes 12 indicators, while the social dimension involves 18 indicators. In the detailed analysis of corporate ESG reports, we also observed that the vast majority of companies tend to disclose data primarily on these two dimensions. For visual clarity, we display the disclosure of environmental, social, and overall numerical indicators across various industries in Figure 4. Following the practice of professional ESG rating agencies like S&P Global and MSCI, which categorize ESG rating results into high, medium, and low tiers based on score ranges (S&P Dow

**Table 4**  
 $Acc_{DE}$  and  $Acc_{DC}$  Across LLMs.

Models	GPT-4	QWEN	GPT-3.5	ChatGLM
$Acc_{DE}$	76.9%	54.9%	47.1%	46.2%
$Acc_{DC}$	83.7%	61.4%	51.9%	53.9%

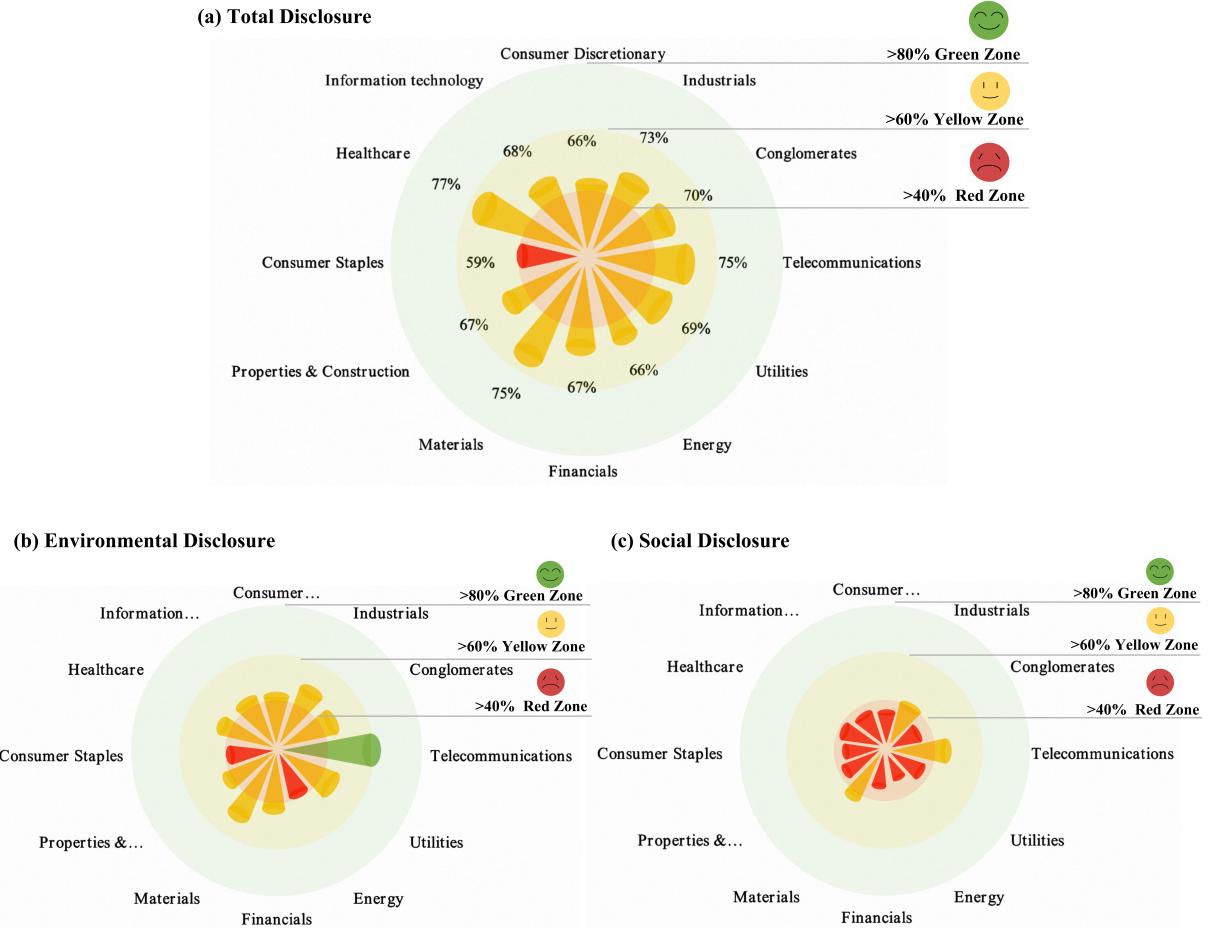
Jones Indices, 2023), and in conjunction with the actual disclosure practices of various industries, this study adopted a three-tier assessment system to evaluate industry disclosure levels, namely: excellent (over 80%), moderate (over 60%), and poor (over 40%).

The statistical results show that the average environmental disclosure rate is 69.5%, while the average social disclosure rate is 57.2%, indicating that, in general, environmental disclosure is superior to social disclosure. Moreover, even among the representative subset of companies by market capitalization listed on the HKEx, the average industry disclosure level has not exceeded 80%. This finding suggests that there is an urgent need to strengthen ESG disclosure practices within the industry. Upon further examination of the healthcare industry, which possesses the highest disclosure level among all sectors, for industry analysis (see Supplementary Figure 1). Within the healthcare sector, there are 5 companies that show good disclosure performance, while the remaining 6 are all rated as moderate. From the perspective of industry distribution, the telecommunications sector, such as China Tower Corporation Ltd., China Mobile Ltd., and China Unicom (Hong Kong) Ltd., healthcare sector, like China Biologic Products Holdings Inc., Alibaba Health Information Technology Ltd., and Materials sector, such as Northern Mining Ltd. stand out with higher disclosure rates, while the Consumer Staples, such as China Starch Holdings Ltd., Want Want China Holdings Ltd. and the energy sector, like China National Offshore Oil Corporation Ltd., China Petroleum & Chemical Corporation Ltd. show weaker performance.

### 5.2.2. Metrics

In the process of parsing ESG reports with ESGReveal, we not only examined the disclosure status but more importantly, extracted quantitative data from the reports. Taking the data on Scope 1 emissions (direct emissions) and Scope 2 emissions (indirect emissions) as an example, to mitigate the impact of industry differences and company sizes on the magnitude of the data, this study normalized the company market values, calculating the greenhouse gas emission intensity per unit of million Hong Kong dollar market value, thereby rendering the emission data of different companies comparable.

According to the data displayed in Figure 5, we observed a general trend: in most industries, the average Scope 2 emissions are higher than Scope 1 emissions. However, there are some exceptions, for example, in the Utilities, Energy, Healthcare and Materials, the level of Scope 1 emissions



**Figure 4:** Comparative Analysis of Disclosure Levels by Industry: (a) Overall Disclosure, (b) Environmental Disclosure, and (c) Social Disclosure.

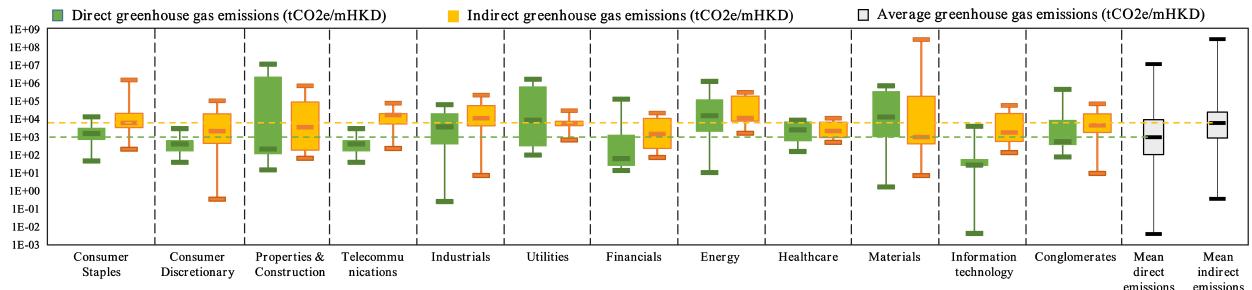
exceeded that of Scope 2. The Utilities sector, in particular, is prominent in Scope 1 emission data due to its coverage of energy-intensive large enterprises such as natural gas, coal gas, and electricity. In contrast, the Financials sector, as a typical labor-intensive industry, is more prominent in Scope 2 emission data. This trend, corroborated by the industry's data on electricity consumption, highlights the significant reliance of the Financials industry on electricity.

### 5.2.3. Key Actions

Based on the analysis of key action data extracted from ESG reports, we further summarized and generalized this data to reveal core information within the text. Supplementary Figure 2 displays the top five high-frequency words that we extracted from cross-industry environmental issue discussions. For instance, within the theme of greenhouse gas emission management, "Greenhouse gas reduction" is the most frequently mentioned phrase across industries, with a word frequency of 24.3%. For water resource management, "Adopt water-saving practices" and "Water-saving training for staff" are equally frequent, each surpassing 10%. Discussions related to the use of fossil fuels indicate a gradual

shift from fossil fuels toward green, renewable energy, as evidenced by terms like "Renewable energy" and "Green transition". Dialogues on hazardous and non-hazardous waste management highlight "Pollution source control" and "Compliant emissions". Additionally, the topics of electricity and energy usage emphasize "Green office practices" and "Smart energy control". Overall, through a comprehensive analysis of high-frequency vocabulary, we find that expressions such as "ESG management" and "Green transition" frequently appear in ESG reports from different industries. These results reflect common actions across industries to reduce environmental impact.

Supplementary Table 2 lists key action words that are closely associated with specific industries, reflecting industry-specific action tendencies and dependencies. For instance, in Properties & Construction sector, including companies such as China National Building Material Company Ltd. and China Resources Cement Holdings Ltd., there is an emphasis on controlling dust and volatile organic compound emissions, implementing water-saving systems, and reusing building materials. In Telecommunications sector, companies including China Tower Corporation Ltd. and China Mobile Ltd. focus on the research and development



**Figure 5:** Comparative Analysis of Greenhouse Gas Emission Intensity for Direct and Indirect Emissions by Industry.

of environmentally friendly products, advancing the recycling and reuse of batteries, and enhancing environmental monitoring and pollution control related to communication towers. As for Healthcare sector, companies such as China Biologic Products Holdings, Inc. and Alibaba Health Information Technology Ltd. are concentrating their ESG efforts on promoting the development of digital medical services. Financials sector plays a leading role in the field of green finance, utilizing financial instruments to support the sustainable development of industries.

## 6. Discussion

Although this study has yielded some initial findings, we have identified several factors that suggest avenues for further research. First, we have noted substantial differences in the capabilities of various LLMs when it comes to ESG analysis tasks. These differences underscore the reliance of LLMs on their intrinsic computational abilities and the depth of their ESG domain knowledge when performing specialized analyses. Second, the structural parsing and retrieval accuracy of ESG reporting documents are of paramount importance when LLMs are employed for ESG disclosure analysis. Enhanced document structuring and more effective information retrieval can notably improve the performance of the RAG framework in terms of data extraction and analysis. Lastly, given that key information in ESG reports is sometimes presented in pictorial forms, our current ESGReveal approach falls short in extracting such data. We aim to rectify this in our future work by refining and optimizing our approach to accommodate these kinds of information.

Looking to the future, research in this domain can evolve in several promising directions. One key direction is the meticulous development of more granular, industry-specific ESG datasets. Such datasets can drive highly accurate data analyses, significantly boosting the transparency and credibility of corporate ESG disclosures. Moreover, the automated extraction techniques pioneered in this study are anticipated to broaden their reach, with potential applications extending to the analysis of pivotal documents such as those from the Intergovernmental Panel on Climate Change. This expansion would leverage digital technology to make a meaningful impact on critical global conversations around climate change mitigation and sustainable development.

In summary, the continued refinement of technological tools and frameworks, along with the expansion of the database's scope of application, promises to not only furnish businesses with invaluable insights but also to play a part in steering entire industries towards a path of enhanced transparency and sustainability.

## 7. Conclusion

In the domain of ESG reporting analysis, this study introduced a method for data extraction, ESGReveal, suited to a multitude of ESG compilation standards. The method is anchored in the RAG paradigm and leverages LLM to accomplish structured extraction of disclosed data within ESG reports. Furthermore, this study utilized the 2022 ESG reports from a representative subset of companies listed on the HKEx to build a structured database for each industry and conducted a quantitative analysis. The pivotal findings of this study are as follows:

(1) The significance of ESG metadata module in ESG Analysis.

The ESG metadata module, rooted in LLM technology, enables the conversion of various ESG reporting standards into structured data extraction commands. By defining ESG indicator attributes and providing an extension mechanism, it bolsters adaptability in a multi-standard landscape. The module significantly enhances LLM performance in ESG analytical tasks, evidenced by gains of +9.9% in GPT-3.5 and +2.5% in GPT-4.

(2) The Applicability of ESGReveal in ESG Analysis Tasks.

ESGReveal, built upon LLM and the RAG paradigm, enables efficient extraction of key indicator data and essential actions from ESG reports. On GPT-4, the accuracy rates for data extraction and disclosure analysis tasks reached 76.9% and 83.7%, respectively, representing an increase of over 20% compared to baseline tests. These results amply demonstrate the potent potential and practical application value of ESGReveal in ESG analysis tasks.

(3) Cross-Industry ESG Data Disclosure and Key Action Analysis at HKEx.

Upon analyzing ESG data disclosures of representative companies under the HKEx framework, the findings reveal that the average disclosure rate for environmental data stands

at 69.5%, while social data disclosure averages at 57.2%. The overall disclosure rate does not exceed 80%, indicating a need for enhanced ESG practices across all sectors. Furthermore, an analysis of key action terms extracted has identified common ESG actions implemented across industries, as well as unique actions that are characteristic of specific sectors.

## References

- Abhayawansa, S., Tyagi, S., 2021. Sustainable investing: The black box of environmental, social, and governance (esg) ratings. *The Journal of Wealth Management*.
- Bai, J., Bai, S., Chu, Y., et al., 2023. Qwen technical report. arXiv preprint arXiv:2309.16609.
- Bronzini, M., et al., 2023. Glitter or gold? deriving structured insights from sustainability reports via large language models. arXiv preprint arXiv:2310.05628.
- Bui, B., Houqe, M.N., Zaman, M., 2020. Climate governance effects on carbon disclosure and performance. *The British Accounting Review* 52, 100880.
- Burnaev, E., et al., 2023. Practical ai cases for solving esg challenges. *Sustainability* 15, 12731.
- Clarkson, P., Li, Y., Richardson, G., Tsang, A., 2019. Causes and consequences of voluntary assurance of csr reports. *Accounting, Auditing & Accountability Journal* 32, 2451–2474.
- Doe, J., 2021. "glitter or gold? deriving structured insights from sustainability reports via large language models". *Journal of Sustainable Reporting* 10, 100–120.
- Du, Z., et al., 2022. Glm: General language model pretraining with autoregressive blank infilling, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 320–335.
- Fischbach, J., et al., 2022. Automatic esg assessment of companies by mining and evaluating media coverage data: Nlp approach and tool.
- Hong Kong Exchanges and Clearing Limited, 2023a. Corporate Governance Code. URL: [https://en-rules.hkex.com.hk/sites/default/files/net\\_file\\_store/HKEX4476\\_3828\\_VER23680.pdf](https://en-rules.hkex.com.hk/sites/default/files/net_file_store/HKEX4476_3828_VER23680.pdf).
- Hong Kong Exchanges and Clearing Limited, 2023b. ESG Reporting Guide. URL: [https://en-rules.hkex.com.hk/sites/default/files/net\\_file\\_store/HKEX4476\\_3841\\_VER18584.pdf](https://en-rules.hkex.com.hk/sites/default/files/net_file_store/HKEX4476_3841_VER18584.pdf).
- Huang, Y., et al., 2022. Layoutlmv3: Pre-training for document ai with unified text and image masking, in: Proceedings of the 30th ACM International Conference on Multimedia.
- Johnson, J., Douze, M., Jégou, H., 2019. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data* 7, 535–547.
- Kim, A.G., Muhn, M., Nikolaev, V.V., 2023. Bloated Disclosures: Can ChatGPT Help Investors Process Information? Technical Report 23-07. Chicago Booth Research Paper.
- LangChain, Inc., 2023. Langchain: Building applications with llms through composability. <https://python.langchain.com/>.
- Lee, O., Joo, H., Choi, H., Cheon, M., 2022. Proposing an integrated approach to analyzing esg data via machine learning and deep learning algorithms. *Sustainability* 14, 8745.
- Lewis, P., Perez, E., Piktus, A., et al., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks, in: Advances in Neural Information Processing Systems, pp. 9459–9474.
- Liu, M., Luo, X., Lu, W.Z., 2023. Public perceptions of environmental, social, and governance (esg) based on social media data: Evidence from china. *Journal of Cleaner Production* 387, 135840.
- Luccioni, A., Baylor, E., Duchene, N., 2020. Analyzing sustainability reports using natural language processing. arXiv preprint arXiv:2011.08073 .
- Moodaley, W., Telukdarie, A., 2023. A conceptual framework for sub-domain specific pre-training of large language models for green claim detection. *European Journal of Sustainable Development* 12, 319.
- Ni, J., et al., 2023. Chatreport: Democratizing sustainability disclosure analysis through llm-based tools. arXiv arXiv:2307.15770.
- OpenAI, 2023a. Gpt-4 technical report. Technical Report.
- OpenAI, 2023b. Instructgpt: Ai for generating instructions. <https://openai.com/research/instructgpt/>.
- Pasch, S., Ehnes, D., 2022. Nlp for responsible finance: Fine-tuning transformer-based models for esg, pp. 3532–3536.
- Perazzoli, S., et al., 2022. Evaluating environmental, social, and governance (esg) from a systemic perspective: An analysis supported by natural language processing. *SSRN Electronic Journal* .
- Raffel, C., et al., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research* 21, 5485–5551.
- Raman, N., Bang, G., Nourbakhsh, A., 2020. Mapping esg trends by distant supervision of neural language models.
- Ruberg, N., Pereira, R., Stein, M., 2023. Greenai – an nlp approach to esg financing.
- Schiemann, F., Tietmeyer, R., 2022. Esg controversies, esg disclosure and analyst forecast accuracy. *International Review of Financial Analysis* .
- Smock, B., Pesala, R., Abraham, R., 2022. Pubtables-1m: Towards comprehensive table extraction from unstructured documents, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4634–4642.
- S&P Dow Jones Indices, 2023. ESG Score Methodology. URL: <http://www.spdji.com/esg-score-methodology>.
- Tsang, A., Frost, T., Cao, H., 2023. Environmental, social, and governance (esg) disclosure: A literature review. *The British Accounting Review* 55, 101149.
- Visalli, F., et al., 2023. Esg data collection with adaptive ai, in: Proceedings of the [Conference Name], pp. 468–475.
- Wang, J., et al., 2021. Milvus: A purpose-built vector data management system, in: Proceedings of the 2021 International Conference on Management of Data, pp. 2614–2627.
- Wang, Y., Sun, Q., He, S., 2023. M3e: Moka massive mixed embedding model.
- Xing, H., Gao, F., Long, R., et al., 2023a. Lore: Logical location regression network for table structure recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press. doi:10.1609/aaai.v37i3.25402.
- Xing, H., et al., 2023b. Lore: Logical location regression network for table structure recognition, in: Proceedings of the AAAI Conference on Artificial Intelligence.
- Yang, H., Liu, X.Y., Wang, C.D., 2023. Fingpt: Open-source financials large language models. arXiv arXiv:2306.06031.
- Zhang, A.Y., Zhang, J.H., 2023. Renovation in environmental, social and governance (esg) research: the application of machine learning. *Asian Review of Accounting* .
- Zhang, Y., et al., 2022. Hlatr: Enhance multi-stage text retrieval with hybrid list aware transformer reranking. arXiv preprint arXiv:2205.10569.

## Supplementary Material

**Table 1**

Indicators of the ESG Metadata Module Based on HKEx Standards.

<b>Indicator Type</b>	<b>Environment (E)</b>	<b>Social (S)</b>	<b>Governance (G)</b>
<b>Quantitative Indicators</b>	<b>12 indicators</b>	<b>18 indicators</b>	<b>4 indicators</b>
	<p>Direct (Scope 1) greenhouse gas emissions (in tonnes) and, where appropriate, intensityEnergy indirect (Scope 2) greenhouse gas emissions (in tonnes) and, where appropriate, intensityTotal hazardous waste produced (in tonnes) and, where appropriate, intensityTotal Non-hazardous waste produced (in tonnes) and, where appropriate, intensityTotal Nitrogen oxide emissions Total Sulfur oxide emissionsTotal particulate emissionsDirect and/or indirect energy consumption in total and intensityDirect and/or indirect gas or oil consumption in total and intensityWater consumption in total and intensityTotal paper consumption and total consumption densityTotal packaging material used for finished products (in tonnes) and, if applicable, with reference to per unit produced</p>	<p>1.Total workforce by gender2.Total workforce by employment type3.Total workforce by age group 4.Total workforce by geographical region5.Employee turnover rate by gender6.Employee turnover rate by age group 7.Employee turnover rate by region8.Employee turnover rate by employment type9.Number work-related fatalities occurred in each of the past three years including the reporting year10.Rate of work-related fatalities occurred in each of the past three years including the reporting year11.Lost days due to work injury12.The percentage of employees trained by gender13.The percentage of employees trained by employee category12.The average training hours completed per employee by gender13.The average training hours completed per employee by employee category14.Number of suppliers by geographical region15.Percentage of total products sold or shipped subject to recalls for safety and health reasons16.Number of products and service related complaints received and how they are dealt with17.Number of concluded legal cases regarding corrupt practices brought against the issuer or its employees during the reporting period and the outcomes of the cases18.Resources contributed (e.g. money or time) to the focus area</p>	<p>1.Number of executive and independent directors2.Number and proportion of female directors3.Antitrust fines, settlement amount and quantity4.Green bond issuance amount</p>
<b>Text indicators</b>	<b>14 indicators</b>	<b>15 indicators</b>	<b>7 indicators</b>
	<p>1. Emissions target(s) and steps taken to achieve them2. Hazardous or Non-hazardous wastes reduction target(s) and steps taken to achieve them3. Greenhouse gases reduction target(s) and steps taken to achieve them4. Waste water reduction target(s) and steps taken to achieve them5. Dusts reduction target(s) and steps taken to achieve them6. Noises reduction target(s) and steps taken to achieve them7. Energy use efficiency target(s) and steps taken to achieve them8. Water efficiency target(s) and steps taken to achieve them9. Packaging materials target(s) and steps taken to achieve them10. Building materials target(s) and steps taken to achieve them11. Significant impacts of activities on soil resources and the actions taken to manage them12. Significant impacts of activities on water resources and the actions taken to manage them13. Policies on identification and mitigation of physical risks of climate change14. Policies on identification and mitigation of Transformational risks of climate change.</p>	<p>1. Occupational health and safety measures adopted, and how they are implemented and monitored2.Measures to review employment practices to avoid child and forced labour3.Steps taken to eliminate such practices when discovered4.Practices relating to engaging suppliers, number of suppliers where the practices are being implemented, and how they are implemented and monitored5.Practices used to identify environmental and social risks along the supply chain, and how they are implemented and monitored6.Practices used to promote environmentally preferable products and services when selecting suppliers, and how they are implemented and monitored7.Products and service related complaints received and how they are dealt with8.Practices relating to observing and protecting intellectual property rights9.Quality assurance process and recall procedures10.Consumer data protection and privacy policies, and how they are implemented and monitored11.Consumer data protection and privacy policies, and how they are implemented and monitored12.Anti-corruption training provided to directors and staff13.Education contribution14.Environmental concerns contribution15.Labour needs contribution</p>	<p>1.Key actions on ESG governance, independent statements, diversified policies, performance evaluation procedures, and election process rules2.Financial expert of the audit committee3.Member of the Remuneration Committee4.Business ethics and anti-corruption management structure, reporting procedures, risk assessment, incident disclosure5.Anti-monopoly rectification measures6.Policies, management structures, action plans for scientific and technological ethics and risk management7.Investment projects in the ESG field</p>

**Table 2**  
Examples of the ESG Metadata Module Based on HKEx Standards.

Aspect	KPI	Topic	Quantity	SearchTerm	Knowledge
<b>Environment (E)</b>					
<b>A1. Emissions</b>	Direct (Scope 1) and indirect (Scope 2) greenhouse gas emissions (in tonnes) and, where appropriate, intensity (e.g. per unit of production volume, per facility).	[Direct (Scope 1), Energy Indirect (Scope 2)]	[Absolute Values]	Greenhouse gases, carbon dioxide, CO <sub>2</sub> , carbon dioxide equivalents	Greenhouse gases include carbon dioxide, methane, nitrous oxide, hydrofluorocarbons, perfluorocarbons and sulphur hexafluoride. Scope 1 (direct emissions) refers to the direct greenhouse gas emissions from fuel combustion activities and physical and chemical production processes that a company directly controls. Typical Scope 1 emissions include those from coal-fired power generation, the use of company-owned vehicles, chemical material processing, and equipment. Scope 2 (indirect emissions) encompasses greenhouse gas emissions from purchased energy sources external to the enterprise, including electricity, heating, steam, and cooling.
	Total waste produced (in tonnes) and, where appropriate, intensity (e.g. per unit of production volume, per facility).	[Non-hazardous Waste, Hazardous Waste]	[Absolute Values]	Construction waste, organic waste, recyclable fertilizers, paper, plastic bottles, wood, municipal waste	Non-hazardous waste includes broken bricks, waste cement, crushed stone materials, municipal waste, etc. Hazardous waste includes: (i) Household waste: waste printer toner, waste ribbons, waste disks, waste calculators, waste fluorescent lamps, etc. (ii) Construction waste: waste paint buckets, waste coatings, etc.
	The types of emissions and respective emissions data.	[Nitrogen oxides (NO <sub>x</sub> ), Sulfur Oxides (SO <sub>x</sub> ), Particulate Matter (PM)]	[Absolute Values]	Nitrogen oxides, NO <sub>x</sub> , sulfur oxides, SO <sub>x</sub> , sulfur dioxide, particulate matter, PM, inhalable suspended particles	Nitrogen oxides (NO <sub>x</sub> ) refer to compounds consisting only of nitrogen and oxygen. Sulfur oxides (SO <sub>x</sub> ) primarily comprise sulfur dioxide and sulfur trioxide, both of which are acidic gases. Particulate matter (PM), also known as dust, refers to the various solid or liquid particles uniformly dispersed within an aerosol system.
	Emissions target(s) and steps taken to achieve them.	[Waste, Exhaust and Greenhouse Gases, Wastewater Dust, Noise]	[Key Actions]	Exhaust and Greenhouse Gases: Gas emissions, greenhouse gases, carbon dioxide. Wastewater: Wastewater, sewage, water consumption. Dust: Dust emission, dust. Noise: Noise pollution	Waste: Waste can be categorized into three physical forms: waste gas, waste liquid, and waste slag, collectively referred to as "the three wastes". Greenhouse gases: Greenhouse gases refer to gases in the atmosphere that can absorb terrestrial long-wave radiation reflected from the ground and re-emit radiation, such as water vapor, carbon dioxide, most refrigerants, etc. Wastewater: Wastewater is water discharged from domestic and production sources that has been contaminated to a certain degree.
.....					
<b>Social (S)</b>					
<b>B1. Employment</b>	Total workforce by gender, employment type (for example, full-time or part-time), age group and geographical region.	[Gender, Age, Region, Type of employment]	[Absolute Values]	Gender: Male, Female. Age: 21-30 years, 31-40 years, 41-50 years, 51-60 years, over 60 years. Type of employment: Full-time, Part-time. Region: Northeast China, North China, South China, East China, West China, Central China, etc. Employee category: Senior management, Mid-level management, General staff	The total number of employees refers to the number of staff currently employed, which includes those who have established a labor relationship with the taxpayer and have legally signed a labor contract or service agreement.
	Employee turnover rate by gender, age group and geographical region.	[Gender, Age, Region]	[Absolute Values]	Same as above	The employee turnover rate refers to the proportion of employees resigning during a given period relative to the average total number of employees. The appropriate turnover rate varies depending on the nature of the enterprise, with traditional businesses typically considering 2-4% as normal.
.....					
<b>Governance (G)</b>					
<b>G1. BOARD COMPOSITION AND NOMINATION</b>	Board composition, succession and evaluation	[ESG governance, Independence Declaration, Diversity Policy, Performance Evaluation Procedures, Election Process Rules]	[Key Actions]	Board of Directors, Corporate Social Responsibility, Ethical Business Conduct, Transparency, Risk Management, Director Elections, Stakeholders	The ESG governance of the Board of Directors refers to the roles and responsibilities of the board in the domains of Environmental, Social, and Corporate Governance (ESG). The Board's Independence Declaration is a statement made by board members prior to fulfilling their duties, confirming their independence and compliance with relevant independence criteria. The Board's Diversity Policy aims to ensure that board members represent a wide range of perspectives in terms of gender, race, age, background, experience, and skills.
	Board composition, succession and evaluation	[Number of Executive Directors, Number of Independent Directors, Number of Female Directors, and their Proportion]	[Absolute Values]	Board Composition, Board Members, Director Appointments	Executive Directors are those who hold full-time positions on the board. They are typically senior executives of the company, responsible for directly managing and operating the daily business activities of the company. Independent Directors are board members who serve in an independent capacity, without direct or indirect interests in the company's management, and are able to provide independent opinions, advice, and oversight.
.....					

**Table 3**

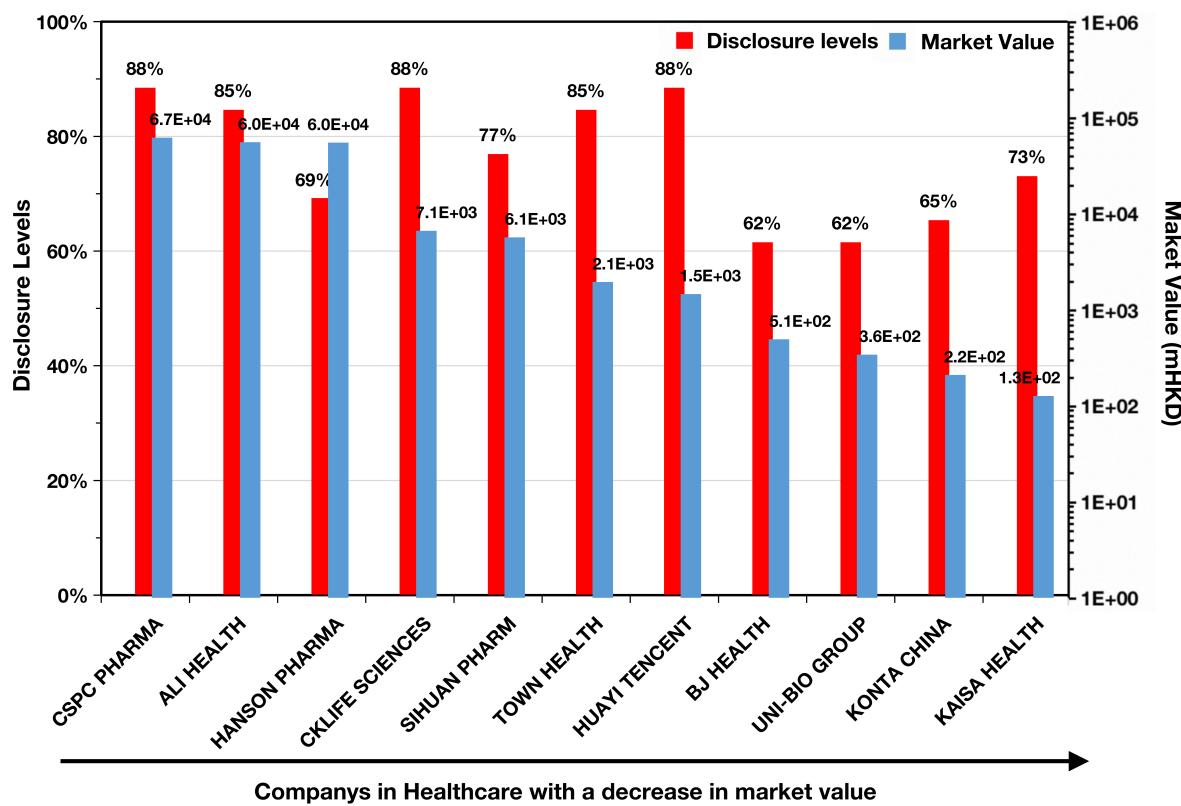
Examples of Prompts and Answers.

Type	Prompts	Answer Examples
Quantitative Indicators	<p>You are an expert in the field of ESG (Environmental, Social, and Governance). Your task is to analyze reference content in both text and CSV table formats to answer questions, providing your responses in JSON format. Please follow these steps for your analysis:</p> <p>Begin by interpreting the meaning of the data disclosed in the table, summarizing it in brief terms.</p> <p>Then, be aware that the provided reference content may not be related to the question. Assess whether the reference content is relevant to the question. If it is, extract all the data related to the question and provide your answer.</p> <p>Your response should include: (1) Whether the reference content discloses data relevant to the question, indicated by a 'disclosure' field with a value of 0 or 1. (2) If relevant data exists, provide the disclosed data in the 'data' field.</p> <p>=====</p> <p>The reference content is as follows: """&lt;Retrievl Contents&gt;"""</p> <p>Supplementary ESG expertise is as follows: """&lt;Knowledge&gt;"""</p> <p>The question is: """Please answer based on the above information and do not strip away the given materials. In terms of &lt;Aspect&gt;, extract the &lt;Topic&gt; about &lt;KPI&gt; in 2022, and output &lt;Quantity&gt;."""</p> <p>Answer in the following format:</p> <p>"""</p> <p>Analysis: {Concise analytical process}</p> <p>Response: {Disclosure:1, "KPI": "{kpi}", "Topic": "{topic}", "Value": "{value}", "Unit": "{unit}"}</p> <p>"""</p>	<pre>{   Disclosure:1   Data:   {     "KPI": "Total waste produced (in tonnes) and, where appropriate, intensity (e.g. per unit of production volume, per facility).",     "Topic": "Non-hazardous Waste",     "Value": "77",     "Unit": "Metric Tons"   },   {     "KPI": "Total waste produced (in tonnes) and, where appropriate, intensity (e.g. per unit of production volume, per facility).",     "Topic": "Hazardous Waste",     "Value": "0.1",     "Unit": "Metric Tons"   } }</pre>
Text indicators	<p>You are an expert in the field of ESG (Environmental, Social, and Governance). Your task is to analyze reference content to answer questions and provide your responses in JSON format. Please follow these steps for your analysis:</p> <p>First, try to interpret the meaning of the content disclosed in the table, and summarize it in concise terms.</p> <p>Next, be mindful that the provided reference content may not relate to the question at hand. Assess whether the reference content is relevant to the question. If it is, extract all the content related to the question and provide your answer.</p> <p>Your response should include: (1) Whether the reference content covers text relevant to the question, indicated by a 'disclosure' field with a value of 0 or 1. (2) If it does cover the relevant text, respond with the related text content in the 'data' field.</p> <p>=====</p> <p>Reference content is as follows: """&lt;Retrievl Contents&gt;"""</p> <p>Supplementary ESG expertise is as follows: """&lt;Knowledge&gt;"""</p> <p>The question is: """Please answer based on the above information and do not strip away the given materials. In terms of &lt;Aspect&gt;, extract the &lt;Topic&gt; about &lt;KPI&gt; in 2022, and output &lt;Quantity&gt;."""</p> <p>Answer in the following format:</p> <p>"""</p> <p>Analysis: {Concise analytical process}</p> <p>Response: {"KPI": "{kpi}", "Topic": "{topic}", "Target": "{target}", "Action": "{action}"}  <p>"""</p> </p>	<pre>{   Disclosure:1   Data:   {     "KPI": " Hazardous or Non-hazardous wastes reduction target(s) and steps taken to achieve them.",     "Topic": "Hazardous or Non-hazardous Waste Management",     "Target": "Reduce carbon emission intensity by 1% annually",     "Action": [       "Conduct a comprehensive inventory and verification of carbon emissions data, manage and reserve carbon assets",       "Continuously strengthen the operational management and research support capacity building of CCUS (Carbon Capture, Utilization, and Storage) demonstration units",       "Focus on carbon offsetting efforts, promote ecological forest construction, and leverage carbon reduction and sequestration through tree planting, reforestation, and mine reclamation"     ]   } }</pre>

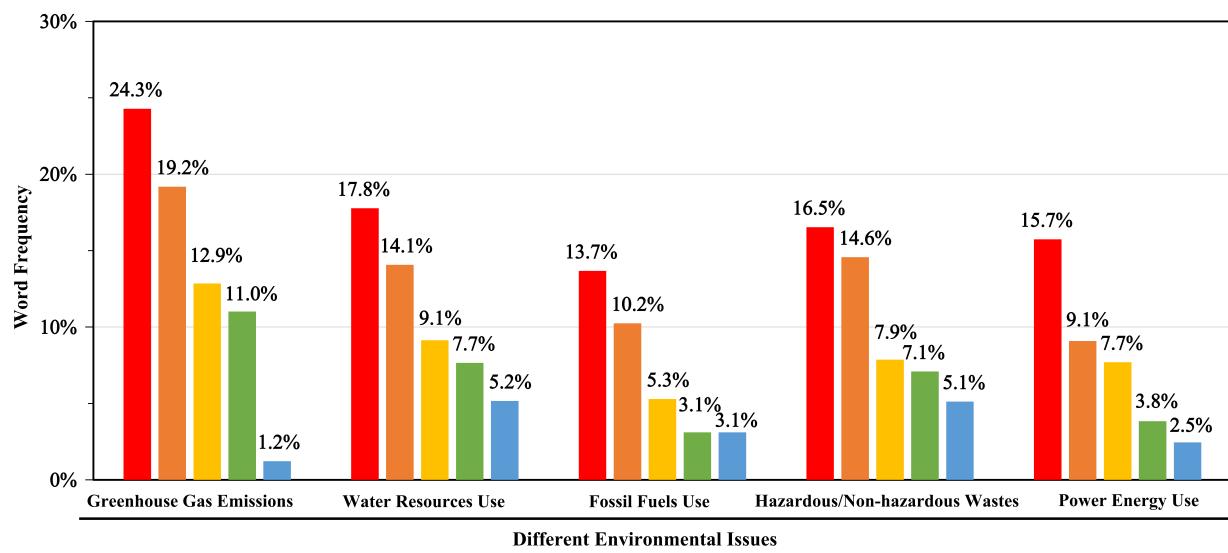
**Table 4**

Representative Companies in Each Sector According to the Hang Seng Industry Classification System.

Industries	Consumer Staples	Properties & Construction	Telecommunications	Consumer Discretionaries	Industrials	Utilities	Financials	Energy	Healthcare	Materials	Information technology industry	Conglomerates
1	BUD PAC (01876)	COUNTRY GARDEN (02007)	CHINA MOBILE (00941)	ANTA SPORTS (02020)	JD LOGISTICS (02618)	KUNLUN ENERGY (00135)	CCB (00939)	CNOOC(0 0883)	CSPC PHARMA (01093)	CHINA HONG QIAO (01378)	MEITUAN-W (03690)	FOSUN INTL (00656)
2	NONGFU SPRING (09633)	CNBM (03323)	CHINA UNICOM (00762)	CHINA RUYI (00136)	GCL TECH (03800)	HKELECTRIC-SS (02638)	HKEX (00388)	PETROCHINA (00857)	ALI HEALTH (00241)	MMG (01208)	XIAOMI-W (01810)	KINGKEY FIN INT (01468)
3	WANT WANT CHINA (00151)	CR BLDG MAT TEC (01313)	HKT-SS (06823)	ALI PICTURES (01060)	CSSC SHIPPING (03877)	CHINA POWER (02380)	ICBC (01398)	SINOPEC CORP (00386)	HANSON PHARMA (03692)	THEME INT'L (00990)	LENOVO GROUP (00992)	SHUN TAK HOLD (00242)
4	WH GROUP (00288)	GEMDAL E PPT (00535)	CHINA TELECOM M (00728)	HOPE EDU (01765)	BAY AREA DEV (00737)	EB ENVIRONMENT (00257)	BANKCOMM (03328)	CHINA SHENHUA A (01088)	CKLIFE SCIENCE S (00775)	JINCHUAN INTL (02362)	SENSETIME-EW (00020)	MEILLEURE HEALTH (02327)
5	CHINA FEIHE (06186)	WEST CHINA CEMENT (02233)	CHINA TOWER (00788)	GOME RETAIL (00493)	CORNERS TONE TEC (08391)	CONCOR DNE (00182)	ABC (01288)	UNITEDENERGY GP (00467)	SIHUAN PHARM (00460)	BROCKMAN MINING (00159)	FIT HONG TENG (06088)	WANG ON GROUP (01222)
6	COFCO JOYCOME (01610)	NANHAI (00680)	PCCW (00008)	AGTECH HOLDINGS (08279)	TIANJINPORT DEV (03382)	BJ ENERGY INTL (00686)	PSBC (01658)	SHOUGANG RES (00639)	TOWN HEALTH (03886)	CDAYENONFER (00661)	FIH (02038)	BEIDA JADE BIRD (08095)
7	CHINA STARCH (03838)	OKG TECH (01499)	HUTCHTEL HK (00215)	NET-AGO TECH (01483)	FULLSHARE (00607)	CAPITAL ENV (03989)	CPIC (02601)	KINETIC DEV (01277)	HUAYI TENCENT (00419)	NEWTIMES ENERGY (00166)	CHINA YOUZAN (08083)	ETERNITY INV (00764)
8	CH BEIDAHUANG (00039)	IB SETTLEMENT (00147)	HKBN (01310)	TAI HING GROUP (06811)	DYNAGREEN ENV (01330)	BG BLUE SKY (06828)	CITIC BANK (00998)	SINOPEC SSC (01033)	JW THERAP-B (02126)	GRANT G GOLD (08299)	KUANGCHI (00439)	CHINA INV HOLD (00132)
9	MING FAI INT'L (03828)	CHINA SANDI (00910)	SMARTONE TELE (00315)	MEIAH ENTER (00391)	CORNERS TONE TEC (08391)	TIANJIN CAPITAL (01065)	CHINA HUARONG (02799)	HONGHUA GROUP (00196)	HOSPITAL CORP (03869)	CWT INT'L (00521)	CHANGHONG JH (03991)	RICH GOLDMAN (00070)
10	HUA LIEN INT'L (00969)	UCD (01599)	SILKWAVE INC (00471)	MAGNIFICENT (00201)	TONGDA GROUP (00698)	KONG SUN HOLD (00295)	CHINA VERED FIN (00245)	CHINA ENERGY (00228)	BJ HEALTH (02389)	GLOBAL BIO-CHEM (00809)	COOLPAD GROUP (02369)	ZACD (08313)
11	BONJOUR HOLD (00653)	CH SUPPLY CHAIN (03708)	KINGWISOFT TECH (08295)	SC HOLDINGS (00413)	CON AERO TECH (00232)	CHI PEOPLE HOLD (00681)	CENTRAL WEALTH GP (00139)	HIDLINDUSTRY (01393)	UNIBIO GROUP (00690)	CSC HOLDING S (00235)	HANG TAI YUE GP (08081)	
12	BAWANG GROUP (01338)	WLS HOLDINGS (08021)	GW TERROIR (00524)	GLODSTEREAM INV (01328)	MAGNUS CONCOR DIA (01172)	C SMARTENERGY (01004)		YUANHE NG GAS (00332)	CT ENTERPRISE ISE (03839)	CHI SILVER GP (00815)	CHINA DIGITAL INFO (00250)	
13	VEEKO INT'L (01173)	CHINA OCEANWIDE (00715)	DIRECTEL (08337)	HYBRID KINETIC (01188)	PAK TAK INT'L (02668)			EPI HOLDING S (00689)	KONTA CHINA (01312)	TAUNG GOLD (00621)	SUPERROBOTICS (08176)	
14	CHAODA MODERN (00682)	SRE GROUP (01207)		Atv HOLDING S (00707)				JINTAI ENERGY H (02728)	KAISA HEALTH (00876)	NORTH MINING (00433)	MEITUAN-W (03690)	
15		YUANDA CHINA (02789)		KAI YUAN HLDGS (01215)				HUARONG ENERGY (01101)		NPE HOLDING S (02326)		
16				DETAI NEWENE RGY (00559)				GREEN LEADER (00061)		RELIANCE GLO HL (00723)		
17				ZHONGZHENG INTL (00943)								
18				SINOFOR TUNE FIN (08123)								



**Figure 1:** Disclosure Levels and Market Value Across Different Companies in Healthcare.



Key actions:

Greenhouse Gas Emissions	Water Resources Use	Fossil Fuels Use	Hazardous/Non-hazardous Wastes	Power Energy Use
■ Greenhouse gas reduction	■ Adopt water-saving practices	■ Renewable energy	■ Pollution source control	■ Green office
■ Green transformation	■ Water-saving training for staff	■ Green transformation	■ ESG management	■ Clean energy
■ Environmental goals	■ Water-saving technology	■ Energy-saving system	■ Compliant emissions	■ Smart energy control
■ Environmental protection facilities	■ Waste water reuse	■ Comply with regulations	■ Improve awareness	■ Optimize resource use
■ ESG management	■ Water facilities maintenance	■ Achieve zero carbon	■ Green packaging	■ LED lighting

**Figure 2:** Key Actions and Word Frequencies Under Different Environmental Issues.

**Table 5**

Partial Examples of Key Actions by Industries for Different Environmental Issues.

Main Actions	Greenhouse Gases Management	Hazardous or Non-hazardous Wastes Management	Water Resources Use Management	Power Energy Use Management	Fossil Fuels Use Management
<b>Properties &amp; Construction</b>	[Provide a guarantee for the completion of project development] [Build an energy-saving office space] [Replacement of first-class energy-efficient equipment or energy-efficient equipment]	[Control the source of potentially volatile organic compounds] [Use sprinkler devices to suppress dust] [Entrust qualified third-party units to dispose] [procure and digest industrial waste]	[Regular inspection, maintenance and leakage testing of the water supply system] [water-saving irrigation method promotion]	[New Energy System Solutions] [Preparation of Central Air Conditioning Operation Plan] [Installation of Efficient Equipment to Manage Fuel and Electricity Use]	[Friendly and sustainable housing] [Control the impact of construction environment] [Promote collaborative disposal projects] [Reuse of building materials]
<b>Telecommunications</b>	[Design security and environmental protection products] [Improving the green level of communication network infrastructure] [Artificial intelligence and machine learning energy control]	[SmarTone encourages battery recycling] [Use red bags to process sensitive information]	[Full coverage of water conservancy communication and water conservancy video surveillance construction application] [High-level monitoring service pollution prevention and control of iron tower] [Shared communication tower construction video monitoring points and radar monitoring points]	[Installation of sensors in different locations] [Widely used in power generation business] [Indoor air conditioning temperature in summer is maintained at 25°C]	[Cross-departmental procurement of office supplies] [Optimize the use of standby generators] [Build a green supply chain] [Reduce the printing of customer communications and enterprise information]
<b>Healthcare</b>	[Providing high-quality online medical and health services] [DMUlish a hazardous waste management plan] [Reduce the number of round trips of patients] [Use digital medical services to help save energy and reduce emissions]	[Re-examination of the production workflow] [Chemical residues produced by empty salts are mixed, air dried and transported to compost manufacturers] [Closed collection of waste gas, exhaust gas absorption device]	[Implement water-saving measures] [Free medication guidance] [Implement water-saving measures]	[Refrigeration unit with frequency conversion control] [Transformation of power distribution facilities] [Artificial intelligence management system (software and hardware) + Internet of Things technology + energy-saving transformation of computer room equipment]	[Implementation of solar power generation project] [Use natural gas as fuel] [Installation of solar photovoltaic] [Eliminate seriously polluting production lines and related facilities]
<b>Energy</b>	[Strengthen the control of total energy] [Update equipment gas source and technology] [Equipment of natural gas collection device at the wellhead] [Oil well maintenance equipment changed to gas boiler] [Decarbonization in the power generation industry]	[Start the pilot construction of "waste-free group"] [Classified treatment of general industrial solid waste] [Protection and control of oil fences and linoleum] [Facilities to prevent seepage and protect the environment]	[Sewer circulation system and recycling pool management] [Installation of solar reverse osmosis device] [Establishment of drilling fluid transfer station] [Circulation and waste reduction] [Implementation of solar collector system on the roof of the plant]	[Close idle electrical appliances] [Eliminate backward motors] [Examine and evaluate risk management policies] [Continue to promote the modernization of corporate governance system]	[Establish a energy consumption and resource assessment system] [Control combustion emissions] [Vehicle emission reduction measures] [Vigorously develop natural gas business] [Scientific and reasonable adjustment of fresh water and natural gas]
<b>Materials</b>	[Establishment of Quality, Health, Safety and Environment Committee] [Adopting Forest Certification System] [Regular Monitoring of Air Quality] [Early Discovery and Repair of Emission Sources]	[Develop new technologies with R&D capabilities] [Digital management of environmental protection facilities] [Avoid oil leakage and implement remedial action plans] [Adoption blasting mining and recycling production processes]	[Water recycling in the mine] [Monitoring business water consumption] [Strengthen pollutant discharge monitoring] [Mining solid waste classification treatment and recycling]	[Reconstruction into a green energy center] [Set up a monitoring system to track fuel consumption] [Technological innovation and transformation and lean management] [Attract hydrogen energy and other industries]	[Management and control of climate change risks] [Redevelopment and transformation of abandoned factories] [Pre-planning routes to improve fuel consumption efficiency]
<b>Information technology</b>	[Supervision of the implementation of energy-saving measures] [Promotion of natural lighting and energy-saving electrical appliances] [Use risk-based standard methods to assess risks] [Improved three categories of input data] [Pass ISO14001 environmental management system certification]	[Waste toner cartridges and waste ink cartridges are handed over to suppliers for unified treatment] [Optimized disposal method] [Implementation of dynamic induction switch] [Hazardous waste conversion energy]	[Routine inspection of sewage treatment system] [some workplaces, warehouses and service stations] [focus on inspection and regular inspection and maintenance equipment]	[Implementation of Intelligent Management System] [Quantification and Conversion of Energy Usage Indicators] [Regular Audit and Inspection of Resource Usage] [Energy Graded Managed Methods] [Purchase Renewable Energy Credit Limit]	[Realizing data center energy saving] [Optimizing the use of resources] [Improving the operation and management strategy of air conditioning] [Regular evaluation of the number of lamps] [Regular evaluation of material consumption]
<b>Financials</b>	[Green Loan] [Carbon Check and Carbon Reduction Plan] [Provide Financial Living Water] [Develop Sustainable Finance] [Listening to Reports and Review Environmental Risk Management Effectiveness]	[Establishment of energy and resource consumption monitoring and diagnosis system] [Underwriting green bonds] [Specially used for the upgrading and transformation of electrolytic aluminum environmental protection and energy-saving technology]	[Help industrial upgrading] [Serving the protection of the Yellow River ecological basin] [Serving the development of the Yangtze River economic belt] [Support the improvement of saline and alkaline land]	[Obtained Green Power Certificate] [Green Finance Business] [Introduction of Carbon Cost Analysis] [Paperless Office] [Disposal of High Fuel Consumption Official Vehicles]	[Establish key performance indicators for environmental protection] [Priority arrangement of energy-saving transformation projects] [Strengthen cooperation between banks, government, banks and enterprises] [Strengthen carbon footprint identification and management]