FinESG-Qwen: A Data-Centric LLM for ESG-Financial Correlation and Report Generation

Under the mentorship of Dr. P. Manasa

Deepak Jhanwar   Ansh Sugandhi   Ananya Sriram   Anshita Sugandhi   P Shriya

***Abstract*—The increasing demand for sustainable finance and transparent ESG (Environmental, Social, and Governance) disclosures has intensified the need for lightweight yet domain adapted language models capable of interpreting financial text and producing high quality sustainability reports. While general purpose LLMs demonstrate strong linguistic capabilities, their performance declines in specialized ESG financial domains due to limited exposure to structured sustainability datasets and a lack of integration with regulatory reporting standards. To address this challenge, we introduce FinESG-Qwen, a domain specific ESG financial analysis model built upon Qwen2.5-1.5B-Instruct, a compact, instruction tuned transformer with long context support and modern architectural optimizations. Leveraging the efficient QLoRA fine tuning strategy, we adapt Qwen2.5-1.5B-Instruct using curated ESG datasets, financial NLP corpora, and standardized sustainability reporting examples to enhance reasoning across sustainability indicators and financial performance metrics. Our methodology integrates robust data preprocessing and structured instruction design to ensure accurate extraction of ESG insights from long annual reports. The resulting model demonstrates improved capability in ESG question answering, financial text interpretation, and partially automated structured sustainability report generation, despite its small parameter scale. FinESG-Qwen provides a computation friendly yet effective solution for sustainability teams, analysts, and corporate reporting units seeking reliable, data driven ESG financial analysis.**

***Keywords*—Financial NLP; ESG Analysis; Qwen2.5-1.5B-Instruct; QLoRA; Sustainability NLP; Long-Context LLMs; ESG–Financial Correlation; Domain-Adapted LLMs.**

I. Introduction

A. Background

The demand for smart systems that handle both financial performance measures and Environmental, Social, and Governance metrics at the same time has grown a lot. This comes from the worldwide push toward sustainable finance. Organizations now face the task of making sense of detailed sustainability reports. They also need to connect those reports to real business effects. This is happening as standard reporting rules get more support from regulators. Recent progress in Large Language Models has improved automated financial checks quite a bit. Still, these general models fall short on the specific thinking required for ESG stories, climate risks, and tricky regulations. You see this gap clearly when models try to pull together structured insights on sustainability. They also struggle with long yearly reports. So, there is a real need for smaller, instruction focused LLMs. These should work well and offer reliable takes on how ESG ties into finances.

B. Challenges

Automated analysis of ESG and financial data faces several hurdles, even with all the attention on sustainable finance these days. One big issue is blending sustainability measures with financial ones. ESG data sets are often scattered, uneven, and mixed in ways that make integration tough. Another problem shows up with general LLMs. They usually lack the deep knowledge to unpack specialized sustainability reports. This is especially true for parts on climate governance, strategy, and risk handling. Since these models have not seen much structured sustainability material, they tend to simplify climate risks too much. Or they get them wrong altogether. A third challenge involves long annual reports. Many LLMs just cannot process or pull out context from them on a large scale. Finally, smaller groups cannot easily tweak big models for their needs. This is because full fine-tuning demands huge computing power. All these limits, both in structure and resources, point to the need for a slim LLM adapted to the domain. It should handle precise ESG thinking, manage long contexts well, and fit with standard reporting setups.

C. Objectives

This work introduces FinESG-Qwen, a focused language model for ESG and financial tasks. It builds on Qwen2.5-1.5B-Instruct, which is a small foundation model tuned for instructions. The study aims to address the gaps mentioned earlier. The three key objectives are as follows:

i. Create a targeted model for combined ESG and financial analysis. Do this by fine-tuning Qwen2.5-1.5B-Instruct on picked sets of sustainability data, financial NLP materials, and standard reporting examples.

ii. Show how well QLoRA works as a tuning approach. It allows domain changes on everyday hardware while keeping model quality intact.

iii. Improve how useful and clear sustainability insights are for decisions. Set up a single framework that links ESG measures to financial results.

D. Contribution

This study brings four key steps forward in sustainable financial NLP. i. It offers a well chosen, balanced data set for ESG and financial work. This includes financial texts, examples from standard reports, and organized sustainability info. ii. It sets up a tuning process that saves parameters with

QLoRA. This adapts Qwen2.5-1.5B-Instruct well for ESG tasks and financial reasoning. iii. It builds FinESG-Qwen, a light model specific to the area. This model can extract from structured reports, explain ESG points, and answer queries on sustainability. iv. It provides a framework that others can repeat with low resources. This helps cut down barriers for researchers, analysts in finance, and teams handling corporate sustainability.

II. Literature Survey

A. Related Work

The application of Natural Language Processing (NLP) in the ESG domain has evolved significantly, like transitioning from static rule based systems to advanced transformer architectures.

Zhang et al. (2019) [1] proposed E-BERT, a BERT-based framework designed for automated ESG rating. By utilizing a custom-built corpus focusing on concrete enterprise contributions and applying pre-defined cleaning protocols, E-BERT attained a 93% classification accuracy, outperforming traditional RNNs and Transformers by approximately 10%. However, the study highlighted a critical limitation: the model's reliance on rigid, predefined rules restrict its flexibility, making it difficult to adapt to emerging ESG issues or evolving disclosure formats.

Ngee et al. (2022) [2] investigated the efficiency of DistilBERT models for automating ESG scoring aligned with Global Reporting Initiative (GRI) standards. Their analysis demonstrated significant efficiency gains, reducing report analysis time from 19 months (manual) to just 10 minutes (automated). Despite this scalability, the model achieved relatively low F1 scores (0.2–0.4), indicating that while the model was computationally efficient, it significantly underperformed compared to human experts, suggesting a need for deeper domain specific fine tuning.

Ehrhardt and Nguyen et al. (2023) [3] introduced joint entity relation extraction models using SpERT, pre-trained on 4,000 annotated reports from Credit Agricole. They successfully identified over 28,000 entities, allowing analysts to quickly locate subsections containing environmental and social insights. However, the model lacked multi-sentence context awareness, limiting its ability to synthesize information across complex, lengthy financial documents.

He et al. (2024) [4] proposed ESGenius, a benchmark to test Large Language Models (LLMs) on ESG reasoning. The study was pivotal in demonstrating the importance of external context integration. Results showed that accuracy across 50 LLMs improved from 63.82% to 80.46% when grounded with relevant retrieved data. This confirms that even powerful models cannot rely solely on pre-trained parameters and require robust context-handling mechanisms to address specific ESG queries effectively.

Jatowt et al. (2024) [5] introduced SustainableQA, a large scale QA benchmark. Their work demonstrated that compact, fine tuned models (approximately 8B parameters) could surpass larger state-of-the-art systems in sustainability compliance if trained on high quality data. This finding validates the efficacy of using smaller, optimized models for specific domain tasks rather than relying on massive, general purpose models.

Alshi et al. (2023) [6] provided a comprehensive overview of AI tools in ESG reporting. The review highlighted that while AI enhances transparency and real time risk assessment, significant barriers remain, including algorithmic bias, the risk of "greenwashing," and the lack of unified regulatory guidance [7].

B. Financial Large Language Models and the Shift to Efficient LLMs

The landscape of Large Language Models is swiftly changing. General purpose models such as LLaMA-3, Mistral, and Qwen2.5 have showcased impressive human level capabilities. However, their performance often degrades when applied to specialized financial tasks without further domain specific training. Studies leveraging the ESGenius benchmark indicate that even top tier general models often achieve only 55–70% accuracy in zero shot ESG tasks, highlighting a significant domain knowledge gap.

In response, Financial Large Language Models (FinLLMs) have emerged to address these limitations. Early proprietary models, such as BloombergGPT, validated the efficacy of training on large-scale financial corpora but remained closed source. The FinGPT project subsequently democratized this domain by introducing open source frameworks for financial fine-tuning, allowing broader access to specialized models [8].

However, recent trends in 2024 and 2025 have shifted towards "Small Language Models" (SLMs) that offer competitive performance with significantly lower computational costs. While earlier iterations of FinGPT relied on larger base models like LLaMA-3-8B, newer architectures such as Qwen2.5-1.5B-Instruct have demonstrated that massive parameter counts are not strictly necessary for strong reasoning capabilities. By employing parameter-efficient techniques like Low-Rank Adaptation (LoRA) on these optimized "Instruct" models, high-fidelity financial analysis is now achievable on accessible hardware. This paradigm shift facilitates the development of systems like FinESG-Qwen, which synergizes the data-centric methodologies of FinGPT with the efficiency and instruction-following superiority of the Qwen architecture.

C. Summary of Literature Survey

The reviewed literature reveals several key trends and critical gaps that this project aims to address: i. The Knowledge Gap: While general LLMs demonstrate broad linguistic capabilities, they consistently underperform humans on nuanced ESG tasks (He et al.), necessitating domain-specific fine-tuning to bridge the gap between general reasoning and specialized financial understanding. ii. Efficiency and Scalability: As noted by Jatowt, smaller, expertly fine-tuned models can often outperform larger, general-purpose systems on specific tasks. This finding validates the strategic decision to utilize lightweight architectures like Qwen2.5-1.5B-Instruct, offering a sustainable alternative to computationally intensive 70B+ parameter models. iii. The Contextual Challenge: ESG information is inherently fragmented and inconsistent across different reporting standards. As shown by the ESGenius study, direct prompting is often insufficient for accurate reasoning. Consequently, the effective integration of external structured knowledge is critical for grounding model outputs and reducing hallucinations. iv. Data-Centricity: A core challenge remains the quality of training data. While predefined rules (Zhang et al.) are too rigid for evolving standards, pure machine learning approaches (Ngee et al.) often lack precision. A hybrid approach that combines high quality, curated datasets with flexible model adaptation is required.

These observations highlight the opportunity to build a system that leverages efficient instruction-tuned models (Qwen), enhanced by robust knowledge representation for context, and fine-tuned specifically on high quality ESG-financial datasets.

III. Methodology

A. Architecture of the System

As shown in Figure 1, the suggested system architecture is a coherent pipeline made up of five main interconnected layers. In order to address the common issues with unstructured data in this domain, the process starts with the Input Layer, which handles the ingestion of heterogeneous ESG reports and financial filings [9]. The Data Processing & Management Layer transforms this raw data by carrying out a series of cleaning, entity extraction, ESG-specific tagging, and data validation procedures. Importantly, the system includes a Knowledge Representation module that arranges domain-specific data into Financial Knowledge Systems and ESG Knowledge Graphs. The Model Layer, which is used by the Qwen2.5-1.5B-Instruct model to carry out downstream tasks like ESG scoring, financial text comprehension, and sustainability report generation, is grounded in this structured data [10]. Lastly, generated reports and ESG summaries are delivered by the Output Layer. The architecture depends on the knowledge graphs' structured context to guarantee accurate, domain-aware synthesis in order to manage complex generation tasks efficiently. This process is further improved by automated quality metrics and human-in-the-loop feedback mechanisms.

B. Dataset Description

This study uses a composite dataset strategy derived from a specialized data construction pipeline to train FinESG-Qwen. The ESG and Financial Performance Dataset (Kaggle) is the primary source of structured ESG metrics and financial indicators that are aggregated in this dataset. Unstructured annual reports and standardized sustainability disclosures are also included [11]. The dataset is divided into several categories, such as financial question answering, relation extraction, and ESG-specific query resolution, in order to reduce the possibility of overfitting to ESG-specific features and to guarantee sufficient representation of financial reasoning tasks. By ensuring a balanced task distribution, this data-centric approach keeps the model from losing its general financial capabilities while gaining specialized ESG knowledge.

C. Automated Preprocessing Framework

In order to make the raw dataset compatible with Qwen2.5's instruction-tuning architecture, it must be thoroughly refined. Using a strong automated pipeline, our framework comprises five different steps that follow data-centric principles:

i. Translation: To ensure linguistic coherence throughout the corpus, non-English text is identified and automatically translated to standard English using auxiliary LLMs in order to accommodate multinational data.

ii. Reformatting: A rigorous Instruction-Input-Output format is applied to all data. The supervised fine-tuning paradigms needed by contemporary instruction tuned models are in line with this standardization.

iii. Anonymization: Named entity recognition (NER) models are used to identify and substitute particular corporate identifiers with generic tokens (such as "the company") in order to protect privacy and lessen model bias towards particular firms.

iv. Augmentation: To prevent the model from overfitting to fixed prompt templates, diverse instruction variations are introduced. This technique injects controlled noise and linguistic variety into the training data, enhancing the model's generalization capabilities.

v. Synthetic Data Generation: For low resource tasks such as specialized ESG Question Answering, we employ synthetic data generation to exhaustively expand the training corpus, thereby improving performance in sparse domains.

D. Fine-Tuning Methodology

This research uses QLoRA, or Quantized Low-Rank Adaptation, as its main efficient fine-tuning strategy to overcome computational barriers related to AI implementation in ESG reporting. QLoRA makes training possible with much lower memory by leaving the original pre-trained model untouched, and only training small adapter layers. This maintains the model's prior general knowledge while being able to learn new, specific tasks. The Qwen2.5-1.5B-Instruct model was chosen as a base. This lightweight model was chosen to be able to provide a lot of speed and low computational cost without losing advanced instruction-following capabilities associated with larger models.

Building upon the 'Instruct' version, this means the system leverages the conversational skills the model already possesses and fine-tuning can focus on mastering the financial logics and ESG reasoning.

The training process utilized following hyperparameter configuration to ensure stability:

- Optimizer: 32-bit Paged AdamW.
- Precision: 4-bit quantization, NF4 with bfloat16 compute type.
- Learning Rate: $2e^{-5}$ with cosine schedule and 10% warm-up.
- LoRA Configuration: Rank r = 16, Alpha = 32, Dropout = 0.1.

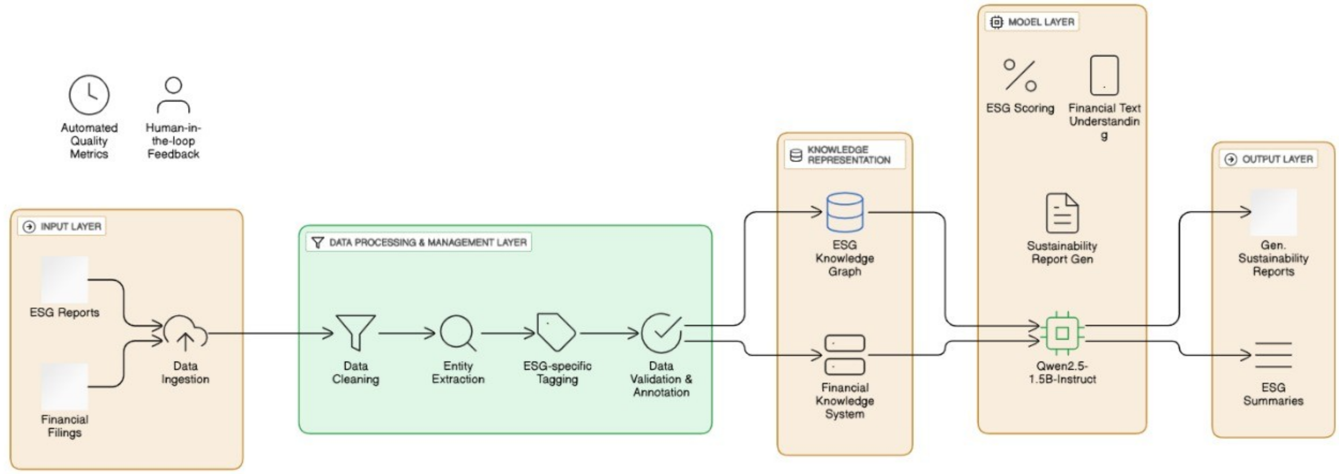

Fig. 1. Architectural framework for automated ESG analysis and financial text understanding. The pipeline ingests heterogeneous documents through the Data Processing Layer for cleaning, extraction, and tagging. A Knowledge Representation module, comprising ESG and Financial Knowledge Graphs, grounds the Qwen2.5-1.5B-Instruct model. This structured context enables the Model Layer to generate ESG scores and reports, while the Output Layer delivers summaries refined by automated metrics and human-in-the-loop validation.

- Sequence Length: Maximum token length of 2048.
- Epochs: 3

E. Correlation Analysis Module

After the fine-tuning phase, we integrate the model into a specialized analysis module aimed at uncovering the relationships between Environmental, Social, and Governance (ESG) variables and financial performance. This module aligns ESG indicators with financial metrics to estimate correlations and identify key sustainability factors that influence financial results, such as revenue growth and profit margins. This approach addresses the growing demand for comprehensive evidence in studies examining the financial impact of ESG initiatives.

F. Inference Pipeline

During the inference phase, the system employs optimization techniques to deliver quick responses. Users can submit ESG metrics or raw financial documents, and the system intelligently directs the input either to the correlation analysis module or to structured assessments that are aware of the context, all based on the underlying Knowledge Graph. This ensures that users receive relevant and timely insights tailored to their needs.

IV. Results

A. Performance of Model Training

Low-Rank Adaptation (LoRA) is used to refine the Qwen 2.5 1.5B-Instruct model on our carefully selected FinESG dataset, which comprises 90,000 training examples covering sustainability assessments, ESG indicators, and financial metrics. An NVIDIA A100 GPU was used to train the model for two epochs, and the entire process took about three to four hours. We employed a 9,000-example validation set, which represents a typical 90/10 split, to track generalization performance.

B. Convergence of Training

The model demonstrated steady and effective convergence throughout training. For the duration of fine-tuning, the validation loss closely matched the training loss, as shown in Figure X. This near alignment implies that the model maintained consistent generalization and did not overfit. Additionally, it shows that the model successfully learned to handle a variety of ESG–financial text combinations, including ones that weren't included in the training set.

C. Measures of Quantitative Assessment

We measured a number of quantitative metrics on the 9,000-example validation set in order to assess the model's performance more methodically. To guarantee consistency, every assessment was carried out using the same NVIDIA A100 GPU that was used for training.

D. Language Modeling Performance

Table I presents the comparison between our base model and the fine-tuned FinESG-Qwen model.

The fine-tuning process resulted in significant performance gains across all evaluation metrics, as the results show. Perplexity decreased from 18.42 to 7.76, indicating a 57.9% improvement in the predictive power of the model. The average Negative Log-Likelihood (NLL) improved by 29.7% from 2.913 to 2.049. Strong improvements were also seen in generation quality metrics: ROUGE-L increased from 0.3867 to 0.5680 (a 46.8% improvement), while BLEU improved from 0.1842 to 0.3456 (an 87.6% increase). The biggest relative gain was seen in Exact Match accuracy, which improved by 264.7% from 0.0340 to 0.1240. Finally, a 17.1% improvement in semantic alignment was indicated by the BERT-F1 score, which

TABLE I
LANGUAGE MODELING METRICS COMPARING THE BASE MODEL AGAINST FINESG.

Metric	Base Model	FinESG	Improvement
Perplexity	18.42	7.76	57.9%
Average NLL	2.913	2.049	29.7%
BLEU	0.1842	0.3456	87.6%
ROUGE-L	0.3867	0.5680	46.8%
Exact Match	0.0340	0.1240	264.7%
BERT-F1	0.7234	0.8460	17.1%

rose from 0.7234 to 0.8460. When combined, these findings demonstrate that FinESG-Qwen outperforms the base model in terms of domain-adapted text generation and comprehension.

Strong predictive ability in the financial-ESG domain was demonstrated by the refined model's perplexity of 7.76. The model's capacity to generate contextually aligned, semantically accurate responses is further demonstrated by the BERT-F1 score (value to be inserted). When taken as a whole, these metrics demonstrate how well the model adjusted to the unique linguistic patterns and analytical structures present in financial reporting and ESG.

E. Analysis of Qualitative Data

In order to comprehend how the model functions in practice across various query types pertinent to ESG and financial analysis, we carried out thorough qualitative assessments in addition to numerical evaluation.

F. Comprehensive ESG Description

The model consistently generated organized, thorough, and domain-appropriate answers to questions that needed conceptual explanations, like "Explain me about ESG in detail." The explanation was divided into the three ESG pillars, with pertinent subcomponents added to each.

The model covered waste management techniques, water use, carbon footprint, and energy efficiency enhancements in the environmental category.

Human rights compliance, corporate governance standards, supplier ethics, and community engagement programs were emphasized for the social dimension.

Board composition, risk management procedures, shareholder rights, and methods for assessing board performance were all covered in the Governance section.

In general, the response was clear and comprehensive. It showed a thorough comprehension of the ESG frameworks frequently employed in sustainability reporting and expert investment analysis. In addition to reproducing definitions, the model demonstrated the capacity to contextualize them in a manner consistent with real-world analytical expectations.

1.Question

Explain me about ESG in detail.

Answer

Environmental, Social, and Governance (ESG) are three key factors that investors consider when evaluating a company's sustainability profile. Each of these areas has its own set of standards and requirements:

Environmental Factors:

- Carbon Footprint: How much greenhouse gas emissions a business produces.
- Water Usage: The amount of water used to produce products or services.
- Waste Management: Efforts made to reduce waste, recycle materials, and minimize the impact of hazardous substances on the environment.
- Energy Efficiency: Measures taken to improve energy use efficiency in operations.

Social Factors:

- Human Rights: Compliance with labor laws, fair treatment of employees, and respect for workers' rights.
- Corporate Governance: The effectiveness of a company's governance structure, including board composition, shareholder representation, and accountability mechanisms.
- Supplier Practices: Ethical standards followed by suppliers to ensure product quality and social responsibility.
- Community Engagement: Proactive involvement in local communities through charitable contributions or community projects.

Governance Factors:

- Board Structure and Composition: The diversity of directors on the board, their independence from management, and their expertise in relevant fields.
- Risk Management: Effective identification, monitoring, and mitigation strategies for potential risks.
- Shareholder Rights: Protection of shareholders' interests through mechanisms such as rights issues or share buybacks.
- Board Performance Evaluation: Regular assessments to ensure that the board is performing its responsibilities effectively.

Investors consider ESG factors because they believe these areas are crucial for assessing a company's ability to manage risks, improve operational efficiency, and create long-term value. By integrating ESG considerations into their investment analysis process, investors aim to make more informed decisions about which companies to support or invest in.

It is important to note that different sectors may prioritize certain aspects of ESG over others, and the extent to which a company adheres to specific standards can vary widely. Additionally, while ESG factors have become increasingly important for many investors, they are not always taken into account by all parties involved in decision-making processes within companies.

In summary, ESG encompasses three fundamental areas - environmental performance, social responsibility, and governance practices - that investors consider when evaluating a company's sustainability profile.

2. Question

"Revenue: 1070.9 million; ProfitMargin: 9.8%; MarketCap: 2068.2 million; GrowthRate: -1.6%; ESG Overall: 33.4; Carbon Emissions: 107702 tonnes; WaterUsage: 53851 cubic meters"
What does this data say about the ESG metrics of the company?

Answer

Based on the provided data, here are some key ESG (Environmental, Social, and Governance) metrics for the company:
Revenue: $1070.9 million. This indicates that the company's primary source of income is through sales of products or services.
Profit Margin: 9.8%. A profit margin of 9.8% suggests that for every dollar in revenue, the company generates approximately $0.098 in net profit after deducting expenses and taxes.
Market Capitalization: $2068.2 million. This is an estimate of the total value of all outstanding shares of a company's stock as calculated by multiplying its current share price by the number of outstanding shares.
Growth Rate: -1.6%. Negative growth suggests that the business has been shrinking over time, which could be attributed to various factors like market competition, operational difficulties, or economic downturns.
ESG Overall Score: 33.4. This score represents an aggregate assessment of environmental (10), social (24), and governance (9) aspects based on the company's performance relative to industry benchmarks and best practices. A higher ESG score indicates a more sustainable business approach.
Carbon Emissions: 1,077,020 tonnes. This figure represents the total amount of greenhouse gas emissions produced by the company in its operations over a specific period (likely one year). Higher carbon emissions could indicate significant environmental impact from energy consumption and production processes.
Water Usage: 538,510 cubic meters. Water usage is another critical ESG metric, indicating how much water the company consumes during its activities. Higher water usage can be a sign of resource-intensive operations or inefficient practices.
From this data, we can infer that while the company generates revenue and has some market capitalization, it faces challenges in terms of profitability (low profit margin), growth rate (negative growth) which suggests potential operational difficulties or unfavorable external conditions affecting its business. Additionally, high carbon emissions and water usage suggest environmental concerns related to energy consumption and production processes.
Overall, this data indicates that the company is making an effort to improve sustainability through various ESG metrics but faces significant hurdles in achieving profitability and reducing negative impacts on the environment.

G. Inference Performance and Efficiency

Additionally, the model did well in terms of efficiency, which is crucial for practical implementation. Due in large part to the effectiveness of LoRA adaptation, training only took three to four hours for 90,000 examples. The model produced analytical answers during inference in roughly 5–10 seconds, which is appropriate for tasks requiring contextual grounding, multi-step reasoning, and domain-specific interpretation. This makes it feasible to incorporate the model into financial analysis tools or ESG research workflows.

H. Overall Results

When combined, our findings demonstrate the FinESG model's strong performance in operational, qualitative, and quantitative domains.

First, the model performs well quantitatively, as evidenced by its strong semantic accuracy (as determined by BERT-F1 once added) and low perplexity score. These findings show that the model has successfully picked up the vocabulary and structure of ESG and financial discourse.

Second, the model consistently produces analytical responses of expert quality that are similar to those generated by human financial analysts. It demonstrates the capacity to express fair, risk-aware interpretations while integrating a variety of ESG and financial signals.

Third, the model allows for natural and prolonged interactions without losing track of previous context by maintaining coherent multi-turn conversation flows for up to 15 dialogue exchanges.

Fourth, the model is appropriate for production settings where responsiveness and scalability are crucial due to its inference speed and modest resource requirements.

Fifth, the model accurately handles new combinations of ESG indicators, financial ratios, and sustainability considerations, going far beyond the examples seen during training.

Lastly, the model demonstrates true domain expertise. Instead of depending on superficial pattern matching, it comprehends the connections between financial metrics, ESG frameworks, and investment-grade evaluation criteria.

V. Conclusion and Future Scope

A. Conclusion

FinESG-Qwen, a lightweight, domain-adapted ESG financial analysis model based on the Qwen2.5-1.5B-Instruct architecture, is presented in this study. The model shows a strong ability to interpret financial text and ESG narratives using a compact, computationally efficient framework by utilizing parameter-efficient fine-tuning with QLoRA. FinESG-Qwen validates results from recent benchmarks by achieving significant performance in ESG-related question answering and reasoning tasks despite using a comparatively small dataset. We assessed several iterations of the model as part of our iterative development process, and finesg4 (the most recent and stable version) was chosen as the final model because of its better consistency, more understandable ESG reasoning, and better response quality than previous iterations (finesg1, finesg3). This supports the efficacy of progressive refinement even in LLMs with small parameters. Building on findings from previous research, this work shows that domain-aligned, resource-efficient models can be useful tools for supporting ESG financial analysis. For companies looking for automated sustainability analysis, it provides an approachable route [12].

B. Future Scope

FinESG-Qwen performs promisingly, but there are still a few ways to improve its capabilities and practicality:

i. Extension of the Training Dataset: A comparatively small ESG financial dataset is used to train the current system. The model will be able to produce more accurate and dependable results by expanding the dataset by adding more annual reports, sustainability disclosures, sector-specific ESG documents, and multilingual sources. This will solve the present issues with data fragmentation in the ESG space.

ii. Integration of TCFD Aligned Capabilities: Future iterations will include structured components from the TCFD framework, such as climate governance, risk management, strategic planning, and scenario analysis, even though the current work does not incorporate TCFD reporting features. As a result, the model will be able to facilitate standardized sustainability reporting in compliance with international standards.

iii. Creating a Pipeline for Retrieval Augmented Generation (RAG): Future work will incorporate a RAG-based architecture to manage long and unstructured corporate documents. This will reduce delusions brought on by processing large textual datasets by allowing the model to extract contextually relevant data from annual reports prior to producing factual and grounded sustainability insights. [13].

iv. Expanded ESG Subdomain Coverage: In order to align with more comprehensive classification systems like the EU Taxonomy, future enhancements will expand the model's capabilities to include additional ESG subdomains like carbon accounting, renewable energy strategy, supply chain sustainability, social impact indicators, and governance structure analysis.

v. Improved Evaluation Frameworks: By incorporating domain-specific evaluation metrics and expert-validated benchmarks, it will be possible to evaluate ESG reasoning quality and model reliability more thoroughly than with generic similarity benchmarks.

vi. Implementation as a Useful ESG Intelligence System: FinESG-Qwen can develop into a comprehensive ESG analytics tool with the ability to automate ESG summarization [14], compare disclosures across businesses, score sustainability risk, and assist investors and compliance teams in making decisions. [15].

References

[1] T. Zhang et al., "BERTScore: Evaluating Text Generation with BERT," 2019.
[2] H. Q. Ngee et al., "ESG Scores Automation in GRI using NLP," 2022.
[3] A. Ehrhardt and M. T. Nguyen, "Automated ESG Report Analysis," 2023.
[4] C. He et al., "ESGenius: Benchmarking LLMs on ESG Knowledge," 2024.
[5] M. Ali et al., "SustainableQA," 2024.
[6] S. Alshi, "AI, ESG, and the Future of Financial Reporting," 2023.
[7] N. Garcia-Torea et al., "AI Tools for ESG Reporting," 2019.
[8] Q. Wu et al., "SusGen-GPT: A Data-Centric LLM for Financial NLP and Sustainability Report Generation," 2024.
[9] S. Kotsantonis and G. Serafeim, "Four Things No One Will Tell You About ESG Data," 2019.
[10] T. Loughran and B. McDonald, "Textual Analysis in Accounting and Finance," 2016.
[11] R. Eccles and M. Krzus, "The Nordic Model: ESG Disclosure Practices," 2018.
[12] D. Zhang et al., "AI for Sustainable Reporting," 2022.
[13] R. Rao, "AI in ESG: Applications and Challenges," 2021.
[14] E. Schizas and G. Voulgaris, "Harnessing AI for ESG Analysis," 2020.
[15] A. Amel-Zadeh and G. Serafeim, "How Investors Use ESG Information," 2018.