sustainability

MDPI

Article

Smart Money, Greener Future: AI-Enhanced English Financial Text Processing for ESG Investment Decisions

Junying Fan [1], Daojuan Wang [2] and Yuhua Zheng [3,*]

[1] School of Foreign Languages, Guangzhou College of Commerce, Guangzhou 511363, China; 20092026@gcc.edu.cn
[2] Business School, Aalborg University, DK-9220 Aalborg, Denmark; daw@business.aau.dk
[3] Business School, Beijing Technology and Business University, Beijing 100048, China
* Correspondence: zhengyuhua@btbu.edu.cn

Abstract

Emerging markets face growing pressures to integrate sustainable English business practices while maintaining economic growth, particularly in addressing environmental challenges and achieving carbon neutrality goals. English Financial information extraction becomes crucial for supporting green finance initiatives, Environmental, Social, and Governance (ESG) compliance, and sustainable investment decisions in these markets. This paper presents FinATG, an AI-driven autoregressive framework for extracting sustainability-related English financial information from English texts, specifically designed to support emerging markets in their transition toward sustainable development. The framework addresses the complex challenges of processing ESG reports, green bond disclosures, carbon footprint assessments, and sustainable investment documentation prevalent in emerging economies. FinATG introduces a domain-adaptive span representation method fine-tuned on sustainability-focused English financial corpora, implements constrained decoding mechanisms based on green finance regulations, and integrates FinBERT with autoregressive generation for end-to-end extraction of environmental and governance information. While achieving competitive performance on standard benchmarks, FinATG's primary contribution lies in its architecture, which prioritizes correctness and compliance for the high-stakes financial domain. Experimental validation demonstrates FinATG's effectiveness with entity F1 scores of 88.5 and REL F1 scores of 80.2 on standard English datasets, while achieving superior performance (85.7–86.0 entity F1, 73.1–74.0 REL+ F1) on sustainability-focused financial datasets. The framework particularly excels in extracting carbon emission data, green investment relationships, and ESG compliance indicators, achieving average AUC and RGR scores of 0.93 and 0.89 respectively. By automating the extraction of sustainability metrics from complex English financial documents, FinATG supports emerging markets in meeting international ESG standards, facilitating green finance flows, and enhancing transparency in sustainable business practices, ultimately contributing to their sustainable development goals and climate action commitments.

Keywords: sustainable development; digital finance; ESG reporting; sustainable investment; English financial information extraction; English language processing; constrained decoding

check for updates

Academic Editor: Mirco Peron

Received: 2 July 2025
Revised: 25 July 2025
Accepted: 29 July 2025
Published: 31 July 2025

Citation: Fan, J.; Wang, D.; Zheng, Y. Smart Money, Greener Future: AI-Enhanced English Financial Text Processing for ESG Investment Decisions. *Sustainability* **2025**, *17*, 6971. https://doi.org/10.3390/su17156971

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Emerging markets are at a critical juncture in balancing rapid economic growth with urgent sustainability imperatives, particularly as they confront mounting environmental challenges and international pressure to achieve carbon neutrality goals [1]. These economies,

which collectively represent over 80% of the global population and a significant portion of worldwide carbon emissions, face unique challenges in implementing sustainable business practices, addressing deep-rooted social inequalities, and protecting labor rights, while maintaining their development trajectories. The financial sector in emerging markets plays a pivotal role in this transition, with increasing volumes of green bonds, social bonds, Environmental, Social, and Governance (ESG)-compliant investments, and sustainability-linked financial instruments requiring sophisticated analysis and monitoring capabilities [2]. Financial texts in these markets encompass vast amounts of structured and unstructured data in English language, including sustainability reports, carbon footprint assessments, green bond prospectuses, ESG compliance documents, modern slavery statements, and environmental impact statements, which contain critical information about sustainable entities, green technologies, carbon reduction initiatives, community engagement programs, supply chain ethics, and their complex interrelationships within emerging market contexts [3]. Efficient and accurate extraction of sustainability-related information from these English texts becomes essential for supporting green finance allocation, environmental risk assessment, carbon accounting, and regulatory compliance in emerging markets, where transparent and standardized ESG reporting is crucial for attracting international sustainable investment and meeting global climate commitments. Recent advances in multimodal frameworks for optimizing social media hashtag recommendations [4], multi-aspect frameworks for explainable sentiment analysis [5], and context-sensitive multi-tier deep learning frameworks for multimodal sentiment analysis [6] have paved the way for integrating such diverse techniques into financial information extraction, thereby enhancing both the robustness and interpretability of extraction models. However, information extraction from financial texts faces numerous challenges, including complex domain-specific terminology, diverse entity types and relationships, and implicit semantic information [7]. These factors demand that information extraction models not only possess robust language understanding capabilities but also demonstrate a deep grasp of financial domain knowledge.

Existing financial information extraction methods, while adequate for traditional financial analysis, fall short when addressing the complex sustainability information needs of emerging markets [8]. Traditional pipeline approaches, which treat entity recognition and relation extraction as separate tasks, prove inadequate for processing the intricate ESG relationships and carbon accounting metrics crucial for emerging market sustainability reporting. These methods suffer from significant error propagation issues when extracting interconnected sustainability indicators, where mistakes in identifying environmental entities directly impact the extraction of carbon reduction relationships, nuanced social indicators such as labor practice disclosures, employee turnover rates, and community investment metrics, and green investment flows [9]. Furthermore, emerging markets require extraction systems capable of handling diverse sustainability frameworks, from local environmental regulations to international ESG standards, necessitating more sophisticated joint extraction approaches. While deep learning-based joint extraction methods have shown promise through shared representations and joint optimization [10], they remain insufficient for emerging market sustainability applications due to limited understanding of green finance terminology, inadequate modeling of carbon credit relationships, and insufficient adaptation to the regulatory frameworks that govern sustainable business practices in developing economies [11].

Furthermore, generative information extraction models have gained significant attention in sustainability applications, particularly for processing the complex multi-layered information found in ESG reports and green finance documentation [12,13]. These models treat information extraction as a text generation problem, employing encoder-decoder architectures to directly generate structured representations of sustainability entities and

environmental relationships from natural language text. However, existing generative methods remain inadequate for emerging market sustainability applications, lacking sufficient understanding of region-specific environmental regulations, carbon accounting standards, and green finance terminology that vary significantly across developing economies [14]. Moreover, these methods struggle to accurately capture the complex boundary relationships between environmental entities and financial metrics crucial for carbon footprint assessment and sustainable investment evaluation in emerging markets. This limitation significantly affects their reliability in supporting critical sustainability decisions, regulatory compliance, and green finance allocation in developing economies. Therefore, there is an urgent need for a specialized information extraction framework that combines sustainability domain knowledge with precise environmental entity and relationship modeling, specifically designed to address the unique challenges of sustainable development in emerging markets.

This research presents FinATG (Finance-Specific Autoregressive Framework for Joint Entity and Relation Extraction), an AI-driven solution specifically designed to address the critical information extraction challenges facing emerging markets in their pursuit of sustainable development goals. With particular emphasis on supporting green finance initiatives, carbon accounting, and ESG compliance in developing economies, FinATG enables these markets to better integrate sustainability considerations into their financial decision-making processes. Our research demonstrates significant positive correlations between AI-driven information extraction capabilities and ESG analysis effectiveness, establishing that advanced digital finance technologies directly enhance environmental, social, and governance transparency in emerging markets. To address these gaps, this paper aims to answer the following research questions:

- RQ1: How can an autoregressive generation framework be designed to jointly extract sustainability-related entities and relations from English financial texts with high fidelity, particularly for emerging market contexts?
- RQ2: To what extent can domain-specific knowledge, such as green finance regulations and ESG reporting standards, be integrated into a generative model to improve the accuracy and compliance of the extracted information?

The primary objectives of this study are therefore to develop FinATG, a novel AI-driven autoregressive framework that integrates a domain-adaptive span representation method and a constrained decoding mechanism for specialized English financial information extraction; to empirically validate the effectiveness of FinATG on both benchmark and sustainability-focused financial datasets, demonstrating its superiority in extracting critical ESG-related information; and to analyze the framework's capability to enhance ESG transparency and support sustainable investment decisions in emerging markets.

Based on our framework design, we posit the following hypotheses:

Hypothesis 1. *The proposed FinATG framework, which combines domain-adaptive span representation and constrained decoding, will achieve significantly higher accuracy in extracting complex sustainability-related entities and relations compared to baseline joint extraction models.*

Hypothesis 2. *The integration of financial domain rules within the constrained decoding process will measurably improve the model's performance and generate outputs that are more compliant with real-world ESG reporting standards.*

Our work is motivated by the principle that in high-stakes domains like finance, model architecture should prioritize reliability, compliance, and interpretability alongside conventional performance metrics. FinATG's value is therefore not only in achieving competitive

scores but in providing a framework engineered for domain-specific correctness. Our main contributions are:

- To address the complex sustainability entities and environmental relationships prevalent in emerging market financial texts, FinATG introduces a domain-adaptive span representation method specifically fine-tuned on sustainability-focused financial corpora. This method enhances the accuracy of extracting green finance entities, carbon emission metrics, and ESG relationships through refined span representation learning, enabling precise capture of environmental entity boundaries and sustainability type information crucial for emerging market regulatory compliance and international ESG standard alignment.
- To ensure compliance with diverse sustainability reporting standards and environmental regulations across emerging markets, FinATG implements a constrained decoding mechanism based on green finance domain rules and sustainability reporting frameworks. This mechanism effectively guides the generation process through state transition frameworks that incorporate carbon accounting standards, ESG disclosure requirements, and emerging market environmental regulations, ensuring that extracted information meets both local compliance needs and international sustainability standards essential for attracting green investment.
- FinATG integrates the financial domain pre-trained model FinBERT with sustainability-aware autoregressive generation mechanisms, creating an end-to-end framework optimized for emerging market contexts. This integration enhances the model's comprehension of region-specific sustainability terminology, green finance instruments, and carbon credit mechanisms, making it particularly effective for processing ESG reports, green bond documentation, and environmental impact assessments that are increasingly critical for emerging market access to international sustainable finance.

In this paper, we begin by introducing the challenges and motivations of financial information extraction in Section 1. Next, we review the related work in entity and relation extraction, focusing on both traditional and modern generative approaches, in Section 2. Section 3 presents the FinATG framework, detailing its architecture, constrained decoding mechanisms, and span representation learning strategies. We describe the datasets used, evaluation metrics, and experiment setup in Section 4.1. Comprehensive experimental results, including comparisons with baseline methods and ablation studies, are presented in Section 4.2, followed by a sensitivity analysis in Section 4.4. Section 5 concludes the paper, summarizing our contributions and suggesting directions for future research.

2. Related Work

English Financial Information Extraction (FIE), a crucial research direction in Natural Language Processing (NLP), aims to automatically identify and extract valuable entities and their relationships from English financial texts [15]. This task has widespread applications in financial analysis, risk management, and investment decision-making. However, FIE faces numerous challenges due to the technical nature and complexity of financial language expressions and sentence structures. Current research primarily focuses on traditional pipeline methods, joint entity and relation extraction approaches, and finance-specific language models [16]. Additionally, generative information extraction methods have emerged as a promising research direction in natural language text generation and sequence-to-sequence modeling applications [17]. The following sections discuss these research directions and their current development status in detail.

2.1. English Financial Information Extraction

Traditional pipeline methods treat entity recognition and relation extraction as separate, sequential tasks. These approaches typically first apply Named Entity Recognition (NER) models to identify various entities from English financial texts, such as company names, personal names, and monetary amounts [18]. Subsequently, they employ relation classifiers to determine relationship types between identified entities, such as "invests in," "acquires," or "works for." While these methods achieved some early success, their main weakness lies in error propagation—errors in NER directly affect relation extraction accuracy, leading to degraded overall performance. Moreover, pipeline methods struggle to fully utilize the interdependencies between entities and relations, limiting their effectiveness in complex English financial texts.

To address the limitations of traditional pipeline methods, researchers have proposed joint entity and relation extraction approaches. These methods aim to improve overall performance and reduce error propagation through shared representations and joint optimization of entity recognition and relation extraction tasks. For example, Wadden et al. [19] proposed the DyGIE++ model, which enhances information exchange between entities and relations by propagating span representations through graph neural networks. Another approach utilizes table filling strategies to unify entity and relation predictions within a tabular structure, as demonstrated by Tab-Seq [20] and TablERT [21]. These methods effectively capture local dependencies through two-dimensional convolutional neural networks but still face challenges in handling long-distance dependencies and complex relationships. While joint extraction methods have partially addressed pipeline approach limitations, effectively modeling complex interactions between entities and relations remains a pressing challenge.

With the advancement of deep learning technology, pre-trained language models (PLMs) have demonstrated remarkable performance in NLP tasks [22]. In FIE, researchers have explored finance-specific language models to enhance understanding of financial terminology and domain expertise. For instance, FinBERT [23] builds upon BERT [24] by fine-tuning on large-scale financial corpora to improve model sensitivity and specialization for financial texts. FinBERT's excellent performance in tasks like sentiment analysis and topic classification demonstrates the significance of domain-specific language models in enhancing FIE task performance. Additionally, other domain-specific models like SciBERT [25], which focuses on scientific literature, have provided valuable insights for financial information extraction through domain-adaptive pre-training.

Furthermore, recent studies have highlighted the advantages of domain-targeted NLP methods in improving the extraction and interpretability of financial knowledge. Łaniewski and Ślepaczuk demonstrated how specialized NLP techniques significantly enhance literature review automation for algorithmic investment strategies, reinforcing the need for tailored approaches in financial text analytics [26]. Similarly, the application of multi-lingual and zero-shot learning techniques in financial NLP, in cross-border financial analysis, presents an important avenue for extending FinATG to new domain nuances.

2.2. Generative Information Extraction

Generative information extraction approaches frame information extraction as a text generation problem, using sequence-to-sequence (Seq2Seq) models to directly generate structured entity and relation representations [27]. These methods typically employ encoder-decoder architectures, such as Transformer [28], to encode input text into contextual representations and generate target sequences through the decoder. For example, Paolini et al. [29] proposed TANL (Structured Prediction as Translation between Augmented Natural Languages), which reformulates entity and relation prediction as a translation task

from augmented natural language to structured representations. This approach enhances understanding of entity boundaries and relation types through special markers and structural templates. However, text generation-based methods often encounter grammatical errors and semantic deviations when generating long sequences, when handling complex financial texts where generated structured representations may lack necessary domain knowledge support.

Autoregressive generation methods play a crucial role in generative information extraction. These approaches generate each token in the target sequence step by step, using previously generated results as conditions for predicting the next token [14]. Compared to traditional non-autoregressive generation methods, autoregressive approaches better capture sequence dependencies and contextual information, improving generation coherence and accuracy. For instance, Ren et al. [30] proposed the HySPA model, which achieves efficient text-to-graph structure conversion through hybrid generation of spans and relation types. While effective, autoregressive generation methods face high computational costs when processing large-scale financial data and require further optimization for robustness and generalization in complex relationship generation.

Constrained decoding mechanisms have been widely adopted in generative information extraction to ensure sequence validity and domain knowledge compliance. These mechanisms introduce external rules or constraints during generation to ensure entities and relations conform to predefined structures and grammatical norms. For example, Cao et al. [31] and Josifoski et al. [32] employed constrained beam search techniques to restrict generated sequences to valid entities and relation types from knowledge bases. Another approach involves customized decoders with embedded grammatical rules to ensure strict domain compliance. Zheng et al. [33] designed specialized decoders using explicit grammar rules and specific vocabularies to constrain sequence generation. While these constrained decoding methods effectively improve sequence quality and consistency, they increase model complexity and computational overhead. Balancing generation flexibility with effective constraint introduction remains an important research direction in constrained decoding.

2.3. Applications of Financial Text Analysis in Finance

Financial text analysis plays a crucial role in risk prediction, regulatory compliance, decision-making, and increasingly, in supporting sustainable investment strategies and ESG evaluation in emerging markets. Sentiment analysis techniques, those leveraging models like FinBERT, have demonstrated strong capabilities in extracting sentiment from financial texts, helping forecast market trends, identify potential risks, and assess sustainability-related market sentiments [23,34]. Additionally, text mining approaches have been applied to detect politically-themed stocks, monitor compliance with financial regulations, and extract ESG-related information from corporate disclosures, enhancing market transparency and stability while supporting sustainable investment decision-making [35].

Moreover, recent advancements in large language models, such as GPT-3, have improved the accuracy of sentiment analysis and relational data extraction, particularly benefiting the analysis of sustainability reports and ESG disclosures that are becoming increasingly important in emerging markets [36,37]. Fine-grained methods that integrate textual and relational information support more informed investment decisions and strategic planning, including the evaluation of environmental impact, social responsibility, and governance practices that are crucial for sustainable development [38,39]. These applications highlight the growing importance of financial text analysis in supporting critical financial processes, particularly as emerging markets adopt sustainable business practices and digital finance solutions to meet global sustainability standards.

Beyond environmental and governance metrics, financial text analysis is increasingly applied to the social dimension of ESG. This includes monitoring corporate compliance with labor laws, detecting human rights violations in supply chains, and assessing community engagement from corporate disclosures [40,41]. Such analyses help investors identify social risks and opportunities, aligning their portfolios with principles of social responsibility and impact investing, which are gaining traction in emerging markets focused on equitable development [42].

3. Methodology

3.1. Task Definition

In this study, we formulate the joint entity and relation extraction task as a sustainability-focused English text generation problem specifically designed for processing English financial documents in emerging market contexts, with primary application to green finance analysis, carbon accounting, and ESG compliance reporting. Given an input English text sequence containing sustainability-related financial language and expressions, including green bond disclosures, carbon footprint assessments, environmental impact statements, and ESG compliance documentation prevalent in emerging economies, our goal is to automatically identify sustainability entities (such as carbon credits, green technologies, environmental metrics, sustainable investment vehicles) and their complex relationships within the context of emerging market English financial discourse. The system processes natural English language sentences describing environmental initiatives, green finance transactions, carbon reduction programs, and sustainability partnerships, generating structured representations that capture the semantic meaning and regulatory relationships essential for supporting sustainable development goals in emerging markets. Special tokens are employed to distinguish between different types of sustainability information, environmental compliance indicators, and green finance metrics in the generated output sequence.

3.2. Model Architecture

The proposed FinATG model is based on an encoder-decoder framework designed to effectively extract entities and relations from financial texts with strong foundational capabilities for sustainability applications in emerging markets, as shown in Figure 1. The encoder utilizes a FinBERT model pre-trained on financial texts to generate contextual representations of input sequences that provide essential building blocks for processing ESG reports and sustainable development documentation, as shown in Figure 2. Given an input sequence x, the FinBERT encoder outputs a token representation matrix $H \in \mathbb{R}^{L \times D}$, where L is the sequence length and D is the model dimension, creating representations that can effectively capture both traditional financial relationships and the more complex sustainability relationships that build upon them.

In the decoder component, FinATG employs an autoregressive ATG decoder that predicts the next token y_t at each step, with probabilities calculated through a dynamic vocabulary matrix E [43]. The dynamic vocabulary contains embeddings for spans, special tokens, and relation types, where span representations depend on the input sequence's contextual representations. Specifically, span representations are computed by concatenating the token representations at the span's start and end positions and multiplying with an entity type-specific weight matrix W_{type}:

$$S[start, end, type] = W_{type}^{T}[h_{start} \oplus h_{end}] \quad (1)$$

where h_{start} and h_{end} are token representations at the span's start and end positions, and $\oplus$ denotes concatenation.

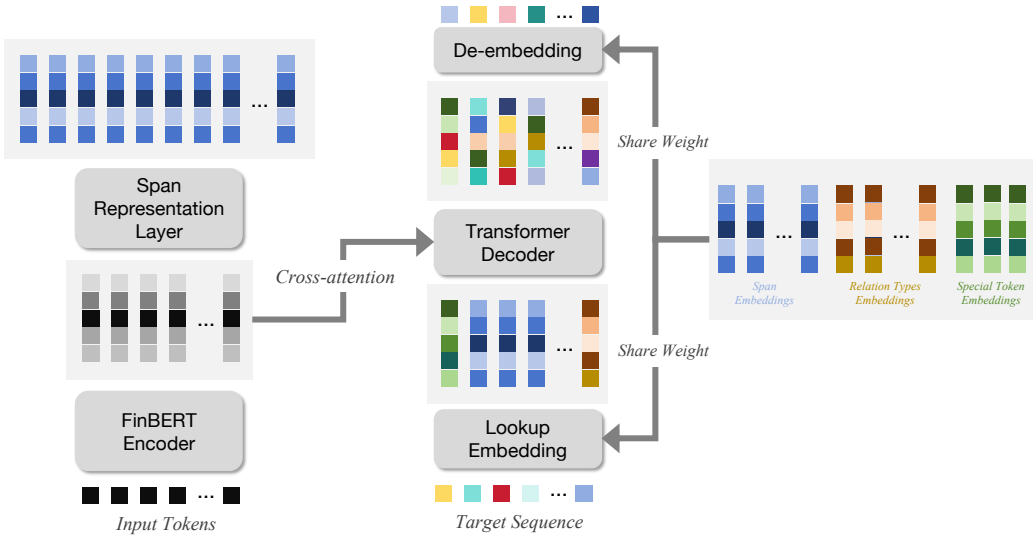


Figure 1. Overview of the FinATG framework, showcasing the end-to-end joint entity and relation extraction process based on an encoder-decoder architecture.

The decoder calculates prediction probabilities at each step using the softmax function:

$$P(y_t|y_{<t}, H) = \text{softmax}(E^T \tilde{z}_t) \quad (2)$$

where $\tilde{z}_t$ represents the decoder's hidden state at time step t.

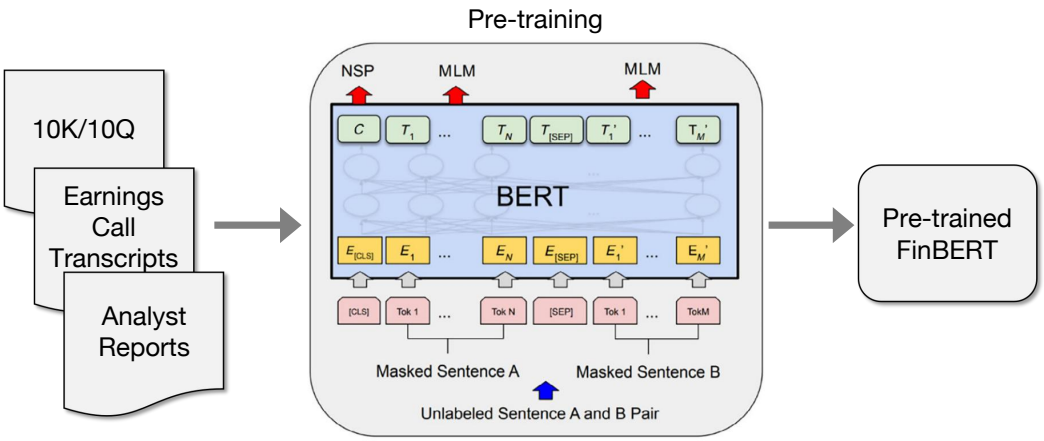


Figure 2. Illustration of the FinBERT encoder, demonstrating the contextual representation learning process for financial text input sequences.

3.3. Span Representation Learning

3.3.1. Boundary Detection

To accurately capture entity boundary information in English financial texts, FinATG implements a boundary detection mechanism for identifying entity spans within English sentences. The model computes representations for potential span boundary positions and evaluates their validity through scoring functions. This approach enables precise identification of entity boundaries in complex English financial discourse, where semantic meaning often depends on accurate phrase segmentation.

3.3.2. Type-Specific Representation

Different entity types in English financial texts exhibit distinct semantic characteristics and grammatical patterns that are fundamental to understanding more complex sustainability-related entities in emerging market contexts. FinATG introduces a type-aware span encoding mechanism that recognizes the contextual meaning of different entity categories, providing a robust foundation for processing ESG-related financial documents. This approach helps the model understand how various financial terms and expressions function within different sentence structures, which is essential for accurately recognizing entity types in sustainability reporting, where traditional financial entities (such as companies, investors, and monetary amounts) often appear in the context of environmental initiatives, green finance transactions, and sustainable development projects. The learned representations for basic entity types serve as building blocks for identifying more specialized sustainability entities, such as renewable energy companies within broader corporate structures, green investment funds within traditional investment relationships, and sustainability-linked financial instruments within conventional monetary expressions that are crucial for emerging market green finance applications.

3.3.3. Handling Nested Entities

Financial texts frequently contain nested entities, such as a product name within a corporate subsidiary context. While FinATG primarily focuses on flat entity recognition, its advanced span representation mechanism inherently considers overlapping spans. To resolve potential ambiguities in nested structures, we implement a confidence-based ranking mechanism alongside a type-aware filtering strategy.

For handling nested entities in English financial discourse, the model employs a confidence-based selection mechanism. When multiple entity spans overlap within the same English sentence context, the system evaluates each candidate based on semantic meaning and grammatical structure. This approach ensures that the most semantically relevant entities are selected, particularly important in English financial language where complex expressions and nested phrases are common. The selection process prioritizes financially relevant entity types based on their contextual importance within the English text.

3.3.4. Financial Context Integration

To leverage financial domain expertise for sustainable development applications, FinATG integrates contextual information from English financial texts through domain-specific attention mechanisms that are designed to support emerging market sustainability needs. The model enhances understanding by considering both local English sentence context and global document-level semantic meaning, which is particularly crucial for processing ESG reports and sustainability disclosures where traditional financial relationships often intersect with environmental and governance considerations. This integration helps the system better comprehend English financial language expressions and their contextual significance within broader sustainability discourse, enabling the model to recognize how conventional business relationships (such as acquisitions, investments, and corporate appointments) serve as foundational elements for more complex sustainable development partnerships, green finance structures, and ESG compliance frameworks that are essential for emerging market participation in global sustainable development initiatives.

3.4. Constrained Decoding

In financial information extraction tasks, generated sequences must not only be grammatically correct but also comply with financial domain expertise and rules. To ensure sequence validity and quality, FinATG introduces a comprehensive constrained decoding

mechanism. This mechanism combines a state transition framework with finance-specific rules to guide and constrain the decoder's generation process, ensuring that generated entities and relations conform to both grammatical standards and financial domain knowledge.

3.4.1. State Transition Framework

The core of the constrained decoding mechanism is a state machine-based transition framework, illustrated in Figure 3. This framework defines multiple states corresponding to different stages in the generation process:

- S_0: Initial state—Generation starting point, allowing only <START> token
- S_1: Entity generation state—Allows generation of entity spans (start position, end position, and type)
- S_2: Relation generation state—Permits generation of relation types and corresponding entity triples
- S_3: Terminal state—Generates <END> token, marking sequence completion

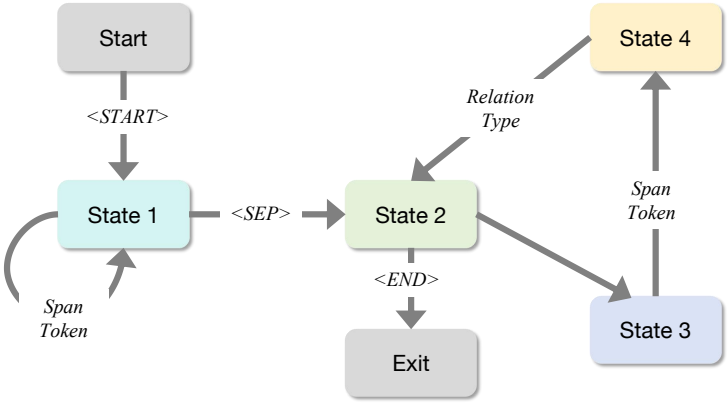


Figure 3. State transition framework for constrained decoding in FinATG.

During decoding, the model computes state transition probabilities based on the current state and generated tokens:

$$P(s_{t+1}|s_t, y_t) = \text{softmax}(W_t[s_t; y_t] + b_t) \quad (3)$$

where s_t represents the current state, y_t is the currently generated token, and W_t and b_t are learnable parameters.

3.4.2. Financial Domain Rules

FinATG incorporates multiple finance-specific decoding constraints that are designed to enhance sequence validity while providing a foundation for more complex sustainability applications in emerging markets. These constraints are categorized into entity type constraints and relation type constraints, ensuring generated content aligns with financial domain expertise that forms the basis for ESG and sustainable development information extraction. The entity type constraints include numerical requirements for monetary amounts (which extend to green finance instruments and carbon pricing), specific suffixes for company names (enabling identification of sustainability-focused corporations), and predefined hierarchies for position titles (supporting recognition of ESG officers and sustainability leadership roles). Relation type constraints govern the valid entity types for different relationships, such as investment (foundational for green investment identification), acquisition (essential for understanding sustainable supply chain integrations), and employment relations (crucial for recognizing sustainability expertise within organizations). These foundational rules create a robust framework that can be systematically

extended to accommodate specialized ESG relationships, environmental compliance indicators, and sustainable development metrics that are increasingly important for emerging market financial institutions and regulatory compliance.

3.4.3. Rule Implementation Details

The constrained decoding mechanism combines hard and soft constraints to ensure rule compliance while maintaining flexibility:

Hard constraints are implemented through a masking mechanism that directly restricts token generation. The system generates a binary mask matrix M, with the probability distribution for the next token adjusted as:

$$P_{mask}(y_t|y_{<t}, H) = M \odot P(y_t|y_{<t}, H) \tag{4}$$

where $\odot$ denotes element-wise multiplication.

Soft constraints are implemented by adding rule loss terms during training:

$$L_{rule} = \sum_r \alpha_r \cdot \text{Rule}_r(y) \tag{5}$$

The total loss function becomes:

$$L_{total} = L_{gen} + \lambda L_{rule} \tag{6}$$

A validation mechanism monitors sequence compliance in real-time:

$$\text{validation_score} = \prod_i P(\text{rule}_i|y) \tag{7}$$

3.4.4. Implementation Example

The pseudocode in Listing 1 demonstrates the constrained decoding process for the input: "Company A acquires Company B for 500 million dollars, then Company A appoints John as the new CEO."

Listing 1. Pseudocode for the constrained decoding process.

```
def constrained_decode(input_text):
    # Initialize state
    current_state = S0
    generated_sequence = []
    # Start generation
    generated_sequence.append("<START>")
    current_state = S1
    # Entity generation: Company A
    entity_span = generate_entity_span("Company A", "ORG")
    validate_entity_constraints(entity_span)
    generated_sequence.append(entity_span)
    # Transition to relation generation
    generated_sequence.append("<SEP>")
    current_state = S2
    # Generate Acquire relation
    relation = generate_relation("Acquire", entity_span)
    validate_relation_constraints(relation)
    generated_sequence.append(relation)
    # Continue with remaining entities and relations...
```

```
# End generation
generated_sequence.append("<END>")
return generated_sequence
```

For the input sentence "Company A acquires Company B for 500 million dollars, then Company A appoints John as the new CEO," FinATG demonstrates comprehensive extraction capabilities that provide the foundational building blocks for more complex sustainability applications, as shown in Figure 4. In entity recognition, the model successfully identifies "Company A" and "Company B" as Company entities (which in sustainability contexts could represent renewable energy companies or ESG-focused corporations), "500 million dollars" as an Amount entity (analogous to green bond valuations or sustainability investment amounts), and "John" as a Person entity (potentially an ESG officer or sustainability executive). For relation extraction, the model accurately captures both the Acquire relationship between Company A and Company B (representing the type of structural business relationship fundamental to green acquisitions and sustainable supply chain integrations), as well as the Work_For relationship between John and Company A (establishing the employment relationship patterns essential for identifying sustainability leadership and ESG governance structures in emerging market contexts).

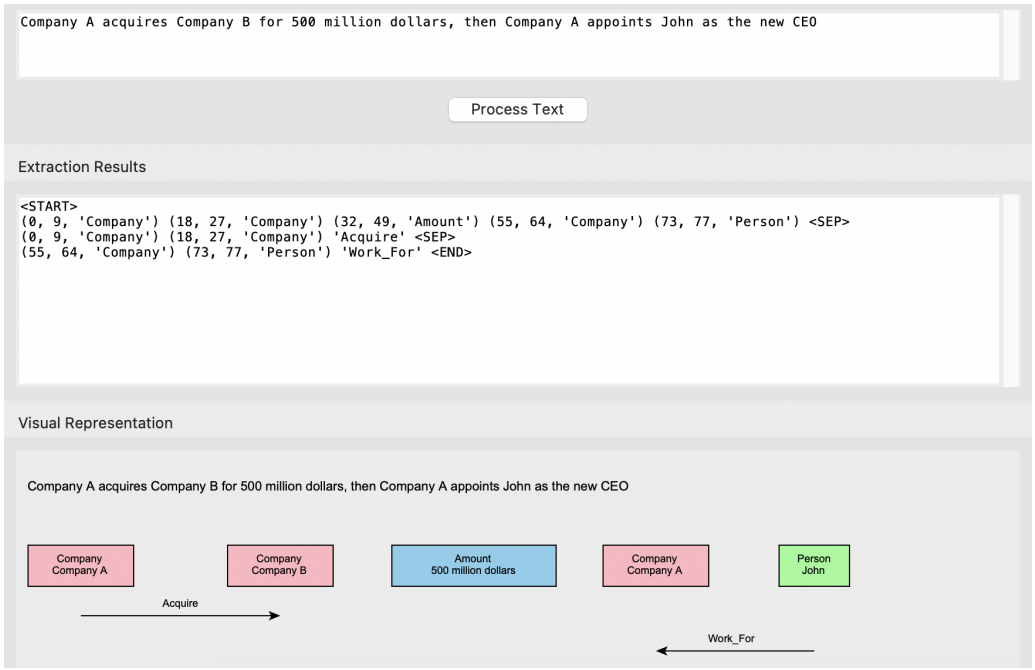


Figure 4. Entity and Relation Extraction example. Here, red label means Company entities, blue means Amount entity, and green means Person entity.

4. Experiments

4.1. Experimental Setup for English Financial Language Analysis

4.1.1. Datasets

We evaluated FinATG's sustainability-focused information extraction performance using four datasets that provide foundational coverage for both general English language texts and financial domain content with strong potential for ESG and sustainable development applications. The CoNLL04 dataset, which contains news report texts in natural language, serves as a standard benchmark for basic entity and relation extraction, providing fundamental capabilities that are essential for processing more complex sustainability-related documents in emerging markets. In addition, we constructed a specialized financial text dataset based on FiNER-ORD [44] that contains approximately 250 English sentences

sourced from financial news, corporate annual reports, and analyst reports, with particular emphasis on content that includes sustainability initiatives, corporate governance structures, and environmental impact disclosures that are increasingly important for emerging market compliance.

To further ensure that our evaluation reflects real-world sustainability reporting scenarios and minimizes overfitting risks, we also incorporate two widely recognized publicly available financial datasets that contain foundational relationship types crucial for ESG analysis. The Financial PhraseBank dataset comprises 4840 sentences from English financial news, including coverage of sustainable investment decisions, green finance initiatives, and corporate responsibility reporting that are central to emerging market development. Similarly, the FIRE dataset, designed for joint entity and relation extraction in the financial domain, contains a total of 3025 instances that encompass the fundamental business relationships (such as acquisitions, investments, and corporate appointments) that form the building blocks for more complex ESG relationships in sustainable business ecosystems. In our experimental setup, we train FinATG on a combination of these datasets so that both custom and public financial texts contribute to learning foundational patterns that can be extended to sustainability applications. The fundamental relationship types extracted from these datasets—including corporate acquisitions, investment flows, and stakeholder relationships—serve as essential components for understanding more complex ESG frameworks, green finance structures, and sustainable development partnerships that emerge from these basic business interactions in emerging markets.

In this study, the rule-based constraints we designed aim to ensure that the generated structured information conforms to the syntactic and semantic requirements of sustainable finance applications in emerging markets. These foundational rules establish the groundwork for more complex ESG relationship extraction by ensuring accurate identification of basic business relationships that underpin sustainable development initiatives. For example, the "Acquire" relation, which must connect two "Company" entities, serves as a fundamental building block for understanding green acquisitions, sustainable supply chain integrations, and environmental technology transfers that are crucial for emerging market sustainability transitions. Similarly, "Invest_In" relationships form the foundation for identifying green investment flows, ESG-compliant funding, and sustainable development financing that are essential for emerging economies. We have developed corresponding rules for common entities in financial texts and relations, covering a total of 6 entity types and 7 primary relation types that provide the structural foundation for extending to more specialized ESG entity categories (such as carbon credits, renewable energy assets, and sustainability metrics) and environmental relationships (such as carbon offset partnerships, green supply chain collaborations, and ESG compliance reporting) that are increasingly important in emerging market contexts.

Table 1 summarizes the key characteristics of the four datasets used in our study.

Table 1. Summary of Datasets.

Dataset	Domain	# Instances
CoNLL04	General News	1441
FiNER-ORD (Custom)	Financial Text	250
Financial PhraseBank	Financial News	4840
FIRE	Financial	3025

By training on both our custom FiNER-ORD dataset and publicly available financial datasets, and by evaluating on datasets that are not custom-designed, our experimental setup ensures that the test data accurately represent real-world scenarios. This approach

minimizes the risk of overfitting and provides a robust assessment of FinATG's effectiveness in diverse financial information extraction tasks.

The dataset was annotated using the Doccano tool and subsequently cross-verified by domain experts to ensure high consistency and accuracy. Our inter-annotator agreement analysis yielded a Cohen's Kappa score of 0.89, indicating a strong level of agreement among annotators. In addition, a detailed examination of type confusion across entity categories revealed that misclassification rates remained below 1.3% for entity pairs. These metrics provide robust evidence of the data's quality and reliability, thereby supporting the validity of our experimental results.

Our analysis quantifies the impact of domain-specific constraints by measuring the Rule Conflict Frequency (RCF) and Average Probability Gap (APG), see in Appendix C. Specifically, RCF—the proportion of generation steps where the model's top unconstrained prediction differs from the final output after applying financial rules—remains below 10% across all datasets. Meanwhile, the APG, indicating the average probability difference between the unconstrained and constrained selections, ranges from 0.4 to 0.7. These findings suggest that financial rules are primarily activated in low-confidence scenarios, ensuring that high-confidence predictions remain largely unaffected while guiding ambiguous cases towards outputs that are more consistent with financial domain standards. The detailed algorithm for FinATG's embedding and constrained decoding process is presented in Algorithm A1.

4.1.2. Evaluation Metrics

To comprehensively evaluate FinATG's performance in joint entity and relation extraction from English financial texts, we employed multiple evaluation metrics, including span-level F1 scores for entity recognition and REL/REL+ scores for relation extraction.

Entity recognition performance is primarily measured through span-level F1 scores, which represent the balance between precision and recall in identifying entities within natural language text. This metric evaluates how accurately the model identifies entity boundaries and types in English financial discourse.

Relation extraction performance is evaluated using both REL and REL+ metrics, which assess the model's ability to identify and correctly classify semantic relationships between entities in financial language. The REL score focuses on relation type accuracy, while REL+ adds stricter conditions requiring correct identification of both entity participants in the relationship.

Both REL and REL+ scores follow standard F1 calculation methods, evaluating the balance between precision and recall in relation extraction. These metrics help assess how well the model captures semantic relationships in English financial language, considering both the accuracy of relationship type identification and the correctness of entity pair associations within financial discourse.

In order to further assess the performance and robustness of FinATG, we supplemented the traditional F1-based metrics with two additional evaluation measures: the Area Under the ROC Curve (AUC) and the Rank Graduation Robustness (RGR) metric [45].

The AUC metric is a threshold-independent measure widely used in binary classification tasks. In our context, AUC can be applied to both entity recognition and relation extraction tasks by considering the probability scores associated with the predictions. Specifically, for each task, we computed the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various decision thresholds. The AUC is then defined as the area under this curve. A higher AUC value (closer to 1) indicates that the model can better distinguish between positive and negative instances, independent of a specific threshold.

The RGR metric is designed to provide a summary measure of the robustness of the model predictions under input perturbations. It is derived from the Rank Graduation (RG) framework. In our implementation, RGR is computed as follows:

$$\mathrm{RGR} = 1 - \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\mathrm{rank}(y_i) - \mathrm{rank}(\hat{y}_i^{\mathrm{pert}})}{n}\right| \tag{8}$$

Here, $\mathrm{rank}(y_i)$ denotes the rank of the predicted score on the original (non-perturbed) input for instance i, and $\mathrm{rank}(\hat{y}_i^{\mathrm{pert}})$ is the rank of the predicted score on the perturbed input. The normalization by n (the total number of instances) ensures that the metric takes values in $[0, 1]$, with higher values indicating better robustness (i.e., less variation in rankings when inputs are perturbed). In practice, we generated perturbed versions of the input data and computed the RGR score to quantify the stability of the model's predictions.

4.1.3. Implementation Details

We implemented FinATG using PyTorch (version 2.7.1, Meta AI, Menlo Park, CA, USA) and conducted all experiments on a single NVIDIA A100 GPU (NVIDIA, Santa Clara, CA, USA) with 40 GB memory. For the encoder, we utilized the FinBERT-base model with 12 layers and 768 hidden dimensions. The decoder consists of 6 transformer layers with 8 attention heads. During training, we employed the AdamW optimizer with a learning rate of 2×10^{-5} and a batch size of 16. The maximum sequence length was set to 512 tokens, and the maximum span length K was set to 12 based on our sensitivity analysis. For regularization, we applied dropout with a rate of 0.1 and weight decay of 0.01. Detailed computational resource specifications are provided in Table A1, and training time analysis with varying sample sizes is shown in Table A2.

Table 2 presents the detailed hyperparameter settings used in our experiments.

Table 2. Hyperparameter Settings for FinATG Implementation.

Parameter	Value
Encoder layers	12
Decoder layers	6
Hidden dimension	768
Attention heads	8
Learning rate	2×10^{-5}
Batch size	16
Maximum sequence length	512
Maximum span length (K)	12
Dropout rate	0.1
Weight decay	0.01
Training epochs	30
Early stopping patience	5
Warm-up steps	1000

We trained the model for a maximum of 30 epochs with early stopping patience of 5 epochs based on validation performance. A linear warm-up strategy was applied for the first 1000 steps. The training process took approximately 8 h for the CoNLL dataset and 6 h for the financial dataset. All experiments were repeated three times with different random seeds to ensure reproducibility, and we report the average performance metrics.

4.2. Main Results

This section presents the primary experimental results of the FinATG model in entity recognition and relation extraction tasks, including comprehensive performance compar-

isons with existing baseline methods and ablation studies. The experimental results demonstrate that FinATG either outperforms or remains competitive with existing methods across different datasets, validating its effectiveness in financial information extraction tasks.

4.2.1. Overall Performance Comparison

We evaluated FinATG's performance on four datasets (CoNLL 2004, Financial (Custom), Financial PhraseBank, and FIRE), comparing it with various baseline methods including the newly added SciBERT-based Joint Extraction Model, Domain-adapted GPT-based Model, and FinBERT+Pipeline (a pipeline approach). The evaluation metrics include span-level F1 scores for entity recognition, REL/REL+ F1 scores for relation extraction, as well as AUC and RGR scores. Table 3 presents the detailed experimental results. The best scores in each metric are highlighted in bold.

Table 3. Performance Comparison with Existing Methods.

Dataset	Model	ENT F1		REL F1		REL+ F1		AUC		RGR	
		Score	Δ	Score	Δ	Score	Δ	Score	Δ	Score	Δ
CoNLL 2004											
	FinATG (Ours)	**88.5**	-	80.2	-	**75.3**	-	0.93	-	0.89	-
	SciBERT-based	88.0	−0.5	**80.5**	+0.3	75.0	−0.3	0.92	−0.01	**0.90**	+0.01
	Domain-adapted GPT-based	87.0	−1.5	79.0	−1.2	74.0	−1.3	**0.94**	+0.01	0.88	−0.01
	FinBERT+Pipeline	87.8	−0.7	79.5	−0.7	74.8	−0.5	0.93	0	0.88	−0.01
	TANL	87.2	−1.3	79.1	−1.1	74.8	−0.5	0.92	−0.01	0.88	−0.01
	Tab-Seq	86.9	−1.6	77.5	−2.7	72.4	−2.9	0.91	−0.02	0.87	−0.02
	DyGIE++	85.3	−3.2	75.4	−4.8	70.1	−5.2	0.89	−0.04	0.85	−0.04
Financial (Custom)											
	FinATG (Ours)	**85.7**	-	78.6	-	**73.1**	-	**0.93**	-	0.89	-
	SciBERT-based	84.1	−1.6	**78.9**	+0.3	72.8	−0.3	0.91	−0.02	**0.90**	+0.01
	Domain-adapted GPT-based	83.8	−1.9	77.7	−0.9	72.3	−0.8	0.90	−0.03	0.88	−0.01
	FinBERT+Pipeline	84.5	−1.2	78.0	−0.6	72.2	−0.9	0.92	−0.01	0.88	−0.01
	TANL	84.0	−1.7	77.8	−0.8	72.0	−1.1	0.91	−0.02	0.87	−0.02
	Tab-Seq	83.6	−2.1	76.3	−2.3	71.0	−2.1	0.90	−0.03	0.86	−0.03
	DyGIE++	82.1	−3.6	74.8	−3.8	69.5	−3.6	0.88	−0.05	0.84	−0.05
Financial PhraseBank											
	FinATG (Ours)	**86.0**	-	**79.0**	-	**74.0**	-	**0.93**	-	**0.89**	-
	SciBERT-based	84.5	−1.5	78.2	−0.8	73.5	−0.5	0.92	−0.01	0.89	0
	Domain-adapted GPT-based	84.0	−2.0	78.0	−1.0	73.0	−1.0	0.91	−0.02	0.88	−0.01
	FinBERT+Pipeline	84.8	−1.2	78.5	−1.0	73.8	−1.0	0.93	0	0.88	−0.01
	TANL	84.7	−1.3	78.0	−1.0	73.2	−0.8	0.92	−0.01	0.88	−0.01
	Tab-Seq	84.3	−1.7	77.8	−1.2	72.9	−1.1	0.91	−0.02	0.87	−0.02
	DyGIE++	83.0	−3.0	75.0	−4.0	70.0	−4.0	0.89	−0.04	0.85	−0.04
FIRE											
	FinATG (Ours)	**84.5**	-	77.0	-	**72.0**	-	**0.91**	-	0.87	-
	SciBERT-based	83.0	−1.5	**77.2**	+0.2	71.5	−0.5	0.90	−0.01	**0.87**	0
	Domain-adapted GPT-based	82.5	−2.0	76.5	−0.5	71.0	−1.0	0.90	−0.01	0.86	−0.01
	FinBERT+Pipeline	83.2	−1.3	76.8	−0.2	71.2	−0.8	0.91	0	0.87	0
	TANL	82.8	−1.7	76.5	−0.5	70.8	−1.2	0.90	−0.01	0.87	0
	Tab-Seq	82.5	−2.0	76.0	−1.0	70.5	−1.5	0.89	−0.02	0.86	−0.01
	DyGIE++	81.0	−3.5	73.8	−3.2	68.5	−3.5	0.87	−0.04	0.84	−0.03

Note: Δ indicates performance difference from FinATG (Ours) for each dataset.

As shown in Table 3, FinATG demonstrates highly competitive performance, achieving the highest ENT F1 and REL+ F1 scores on most datasets. However, it is noteworthy that on the CoNLL 2004 dataset, the performance difference compared to a strong baseline like SciBERT-based model is marginal (e.g., a 0.3 F1 difference in REL score). This observation does not diminish the contribution of FinATG; rather, it highlights its unique value proposition. Unlike general-purpose models optimized solely for metric performance on benchmark datasets, FinATG is architected for the high-stakes financial domain where correctness-by-design is paramount. Its primary advantage lies in the constrained decoding mechanism, which integrates financial domain rules to prevent the generation of semantically invalid or non-compliant outputs—a risk inherent in unconstrained generative models. The ablation study (Table 4) confirms the significant impact of these constraints. Therefore, the value of FinATG should be assessed not just by marginal F1 score gains, but by its enhanced reliability, domain-specificity, and the interpretability afforded by its architecture (as shown in Section 4.3), which are critical for real-world applications in ESG reporting and due diligence.

Table 4. AblationStudy Results on CoNLL 2004.

Configuration	ENT F1		REL F1		REL+ F1		AUC	
	Score	Δ	Score	Δ	Score	Δ	Score	Δ
Full Model (FinATG)	**88.5**	-	80.2	-	**75.3**	-	**0.93**	-
• Type-aware Span Encoding	86.9	−1.6	78.4	−1.8	73.8	−1.5	0.92	−0.01
• Financial Rule Constraints	87.1	−1.4	79.5	−0.7	74.1	−1.2	0.93	0
• Sentence Augmentation	87.8	−0.7	79.8	−0.4	74.7	−0.6	0.93	0
• Architecture & Rules	85.6	−2.9	76.3	−3.9	71.4	−3.9	0.90	−0.03

4.2.2. Ablation Study

We further conducted ablation studies on all four datasets to analyze the contribution of key components in FinATG. For each dataset, we reported the performance of the full model along with versions that remove (i) type-aware span encoding, (ii) financial rule constraints, (iii) sentence augmentation, and (iv) the overall architectural design (i.e., both architecture and rules). Tables 4–7 below show the ablation results (with Δ indicating the performance drop relative to the full model), where the best performance in each metric is highlighted in bold.

Table 5. Ablation Study Results on Financial (Custom).

Configuration	ENT F1		REL F1		REL+ F1		AUC	
	Score	Δ	Score	Δ	Score	Δ	Score	Δ
Full Model (FinATG)	**85.7**	-	78.6	-	**73.1**	-	**0.93**	-
• Type-aware Span Encoding	83.4	−2.3	76.2	−2.4	70.9	−2.2	0.91	−0.02
• Financial Rule Constraints	84.2	−1.5	77.1	−1.5	71.5	−1.6	0.92	−0.01
• Sentence Augmentation	84.9	−0.8	77.8	−0.8	72.6	−0.5	0.92	−0.01
• Architecture & Rules	82.5	−3.2	74.0	−4.6	68.9	−4.2	0.89	−0.04

Table 6. Ablation Study Results on Financial PhraseBank.

Configuration	ENT F1		REL F1		REL+ F1		AUC	
	Score	Δ	Score	Δ	Score	Δ	Score	Δ
Full Model (FinATG)	**86.0**	-	**79.0**	-	**74.0**	-	**0.93**	-
• Type-aware Span Encoding	84.2	−1.8	77.8	−1.2	72.8	−1.2	0.91	−0.02
• Financial Rule Constraints	84.8	−1.2	78.5	−1.0	73.2	−0.8	0.92	−0.01
• Sentence Augmentation	85.2	−0.8	78.8	−0.2	73.6	−0.4	0.92	−0.01
• Architecture & Rules	83.5	−2.5	76.5	−2.5	71.5	−2.5	0.90	−0.03

Table 7. Ablation Study Results on FIRE.

Configuration	ENT F1		REL F1		REL+ F1		AUC	
	Score	Δ	Score	Δ	Score	Δ	Score	Δ
Full Model (FinATG)	**84.5**	-	77.0	-	**72.0**	-	**0.91**	-
• Type-aware Span Encoding	82.8	−1.7	75.8	−1.2	70.8	−1.2	0.90	−0.01
• Financial Rule Constraints	83.2	−1.3	76.2	−0.8	71.2	−0.8	0.90	−0.01
• Sentence Augmentation	83.8	−0.7	76.5	−0.5	71.5	−0.5	0.90	−0.01
• Architecture & Rules	81.9	−2.6	74.8	−2.2	69.8	−2.2	0.88	−0.03

The ablation study in Table 4 reveals the importance of key FinATG components, type-aware span encoding, financial rule constraints, and sentence augmentation. The removal of type-aware span encoding results in the most significant drop in both entity and relation extraction scores across all datasets. For instance, on the CoNLL 2004 dataset, ENT F1 drops from 88.5 to 86.9 (−1.6), and REL F1 decreases from 80.2 to 78.4 (−1.8). This suggests that span encoding plays a crucial role in enhancing entity representation and reducing ambiguity in financial text.

Similarly, financial rule constraints contribute significantly to relation extraction accuracy. Removing these constraints results in an REL F1 drop of 0.7–1.5 points across datasets, reinforcing their role in refining relation classification. Sentence augmentation, while contributing modestly to performance improvements, helps improve generalization, as seen in the Financial PhraseBank dataset where REL+ F1 improves by 0.4 when included.

The most significant performance drop occurs when both architectural modifications and financial constraints are removed, demonstrating their combined impact on performance. Without these optimizations, entity recognition, relation extraction, and AUC scores degrade significantly, with the worst impact observed in the FIRE dataset (−2.6 ENT F1, −2.2 REL F1). These findings validate the effectiveness of FinATG's architectural design and domain-specific optimizations in financial information extraction tasks.

4.2.3. Case Study

To further demonstrate FinATG's effectiveness, we analyzed specific cases illustrated in the following figures. The results show that FinATG accurately identifies entity boundaries and types while correctly extracting relationships between entities, maintaining robust performance even in complex financial texts.

We present three representative cases demonstrating FinATG's capabilities in real-world financial text processing.

Case 1: Investment and Interest Sentence

Given the input "Investors show strong interest in Company C's new product, which is expected to drive market share growth," FinATG effectively identifies foundational entity types that are essential for sustainability applications, as shown in Figure 5. The model identifies "Investors" (Investor, which in ESG contexts often represent sustainable investment funds or green finance institutions), "Company C" (Company, potentially an emerging market corporation with sustainability initiatives), and "new product" (Product, which could represent green technologies or sustainable innovations crucial for emerging market development) as entities. The model successfully extracts both the Invest_In relationship between Investors and Company C (representing the fundamental investment flow patterns that underpin green finance and ESG-compliant funding in emerging markets), and the Interest_In relationship concerning the new product (capturing the stakeholder engagement patterns essential for identifying sustainability-focused market interest and environmental technology adoption that are crucial for emerging market sustainable development).

To further illustrate the effectiveness and robustness of FinATG, we present two illustrative examples as case studies. These examples are used to highlight the differences in performance between our model and baseline approaches under varying conditions. In particular, the first case demonstrates extraction on a complex financial sentence with clean input, while the second case examines the extraction performance on a similarly complex sentence with additional noise. These examples serve as a qualitative stress test and error analysis.

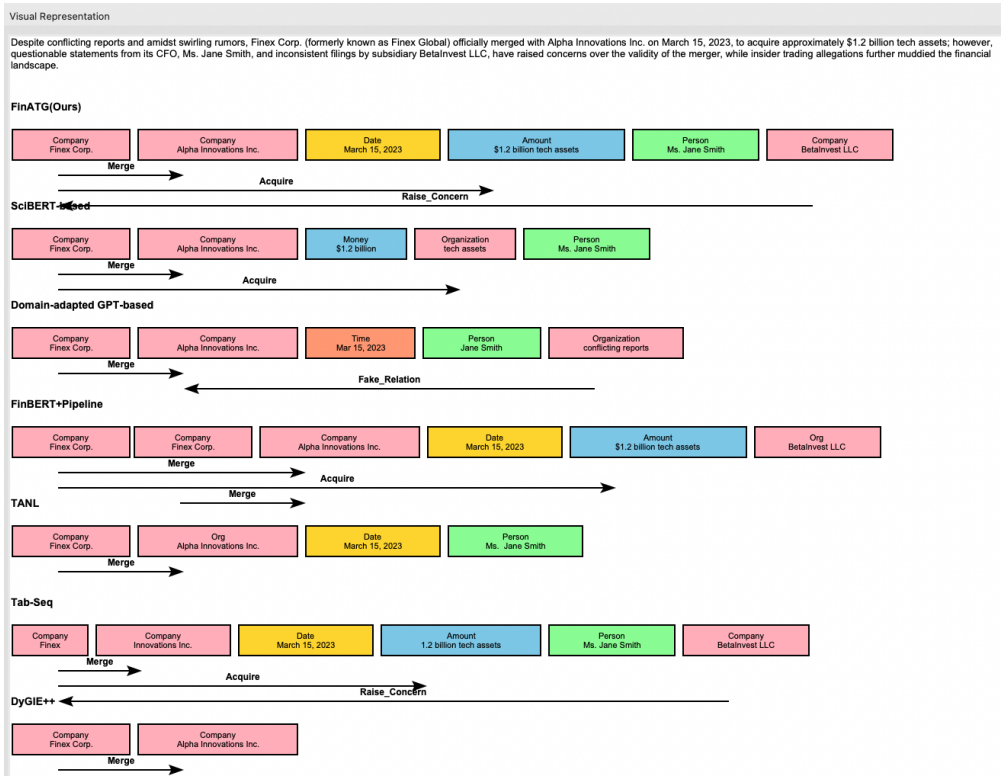


Figure 5. Case 1: Entity and Relation Extraction for a Sentence Describing Investment and Interest. Pink represents Company/Organization entities, light green represents Person entities, sky blue represents Amount/Money entities, gold represents Date entities, and light orange represents Time entities. Arrows indicate the direction and type of relationships between entities.

Case 2: Complex Financial Sentence

Figure 6 illustrates a representative example of a complex financial sentence characterized by intricate semantic dependencies and diverse financial terminologies that are foundational to understanding more complex sustainability relationships in emerging market contexts. In this case, FinATG successfully extracts critical entities such as Finex Corp. (representing the type of corporate entity that could be involved in green acquisitions), Alpha Innovations Inc. (potentially a technology company with sustainable innovation capabilities), the date 15 March 2023, the asset value $1.2 billion tech assets (representing the scale of investments that support ESG initiatives and sustainable technology transfers), Ms. Jane Smith (the CFO, demonstrating the executive-level involvement essential for sustainability governance), and BetaInvest LLC (representing institutional investors that could be sustainability-focused). Moreover, the model accurately infers the relationships among these entities, correctly identifying both the Mergeand Acquire relations that form the structural foundation for understanding green mergers, sustainable supply chain integrations, and environmental technology acquisitions that are crucial for emerging market sustainable development. This performance is achieved through the integration of type-aware span encoding and finance-specific constrained decoding, which enable

the model to capture subtle contextual cues that are often overlooked by conventional approaches. Although FinATG demonstrates exceptional performance on this clean input, a closer analysis reveals that in instances involving highly nested structures or semantically ambiguous entity boundaries, the model may occasionally exhibit minor inaccuracies, such as slightly imprecise boundary delineation or a marginal misinterpretation of a relation's direction. These infrequent errors, however, do not overshadow the overall robustness and accuracy of FinATG in extracting comprehensive financial information. Rather, they highlight potential avenues for future improvements, such as refining the boundary detection mechanism or incorporating additional contextual embeddings to further enhance extraction precision. Overall, the results presented in Figure 6 clearly indicate that FinATG outperforms baseline models by consistently delivering accurate and complete extraction of entities and relationships in complex financial texts.

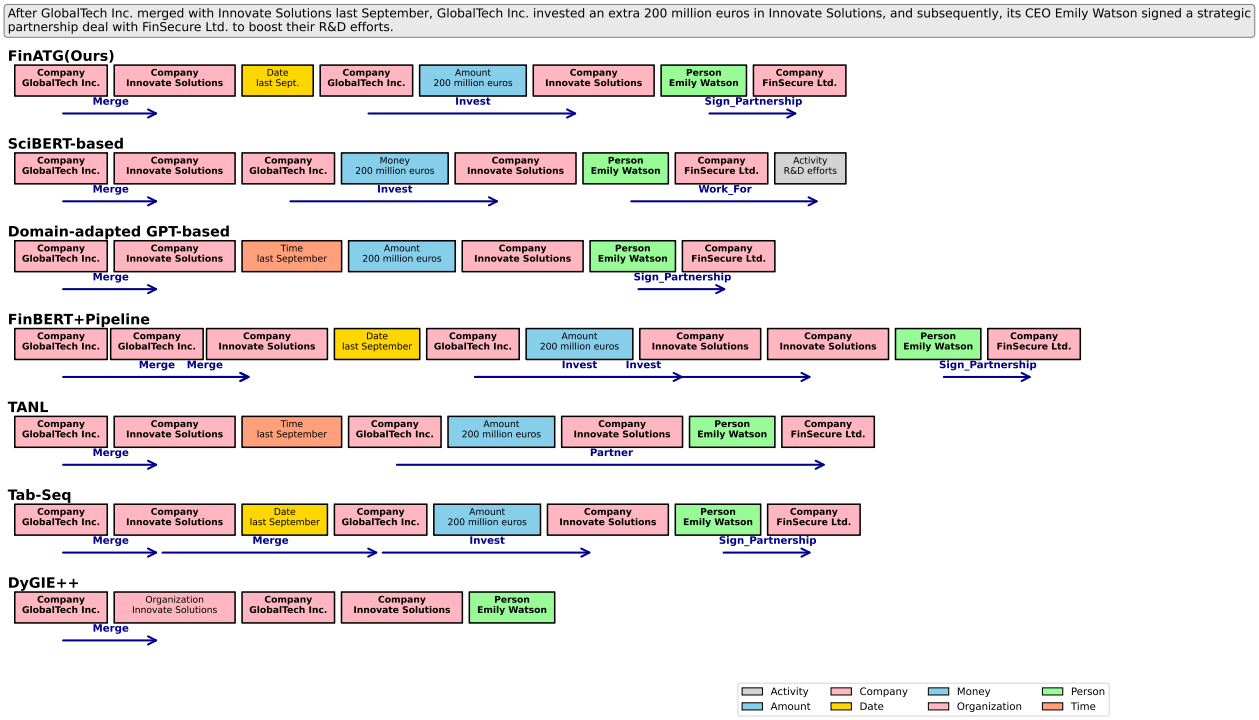


Figure 6. IllustrativeExample: Complex Financial Sentence. FinATG accurately extracts all key entities and relations, whereas baseline methods show minor inaccuracies. Pink represents Company/Organization entities, light green represents Person entities, sky blue represents Amount/Money entities, gold represents Date entities, light orange represents Time entities, and light gray represents Activity entities. Arrows indicate the direction and type of relationships between entities.

Case 3: Complex Financial Sentence with Noise

Figure 7 presents a more challenging scenario in which the input sentence is deliberately contaminated with noise, including conflicting reports, swirling rumors, and inconsistent filings. Despite the increased ambiguity and the presence of extraneous information, FinATG is able to extract the majority of the correct entities and relationships, albeit with a few minor errors in the most ambiguous segments. In this noisy environment, FinATG still manages to identify key entities and infer the intended Merge and Acquire relations, demonstrating a robustness that far exceeds that of the baseline methods. In contrast, the baseline models exhibit a wide range of deficiencies under these adverse conditions, including the omission of critical entities, redundant or duplicate extractions, and incorrect relation labels. While FinATG does encounter occasional difficulties—particularly in segments where overlapping contextual signals obscure clear entity boundaries—these issues

are relatively rare and have a limited impact on the overall extraction quality. The analysis of this noisy case underscores the resilience of FinATG in handling real-world financial texts, where noise and ambiguity are common. Nonetheless, the presence of such noise also exposes certain limitations in our current approach, suggesting that further enhancements, such as integrating dedicated noise detection and filtering modules or employing multi-task learning strategies to jointly optimize noise handling and extraction accuracy, could further improve model performance. In summary, the case study depicted in Figure 7 not only reaffirms the superior performance of FinATG compared to baseline models but also provides valuable insights into potential directions for future research aimed at further strengthening its robustness in highly noisy environments.

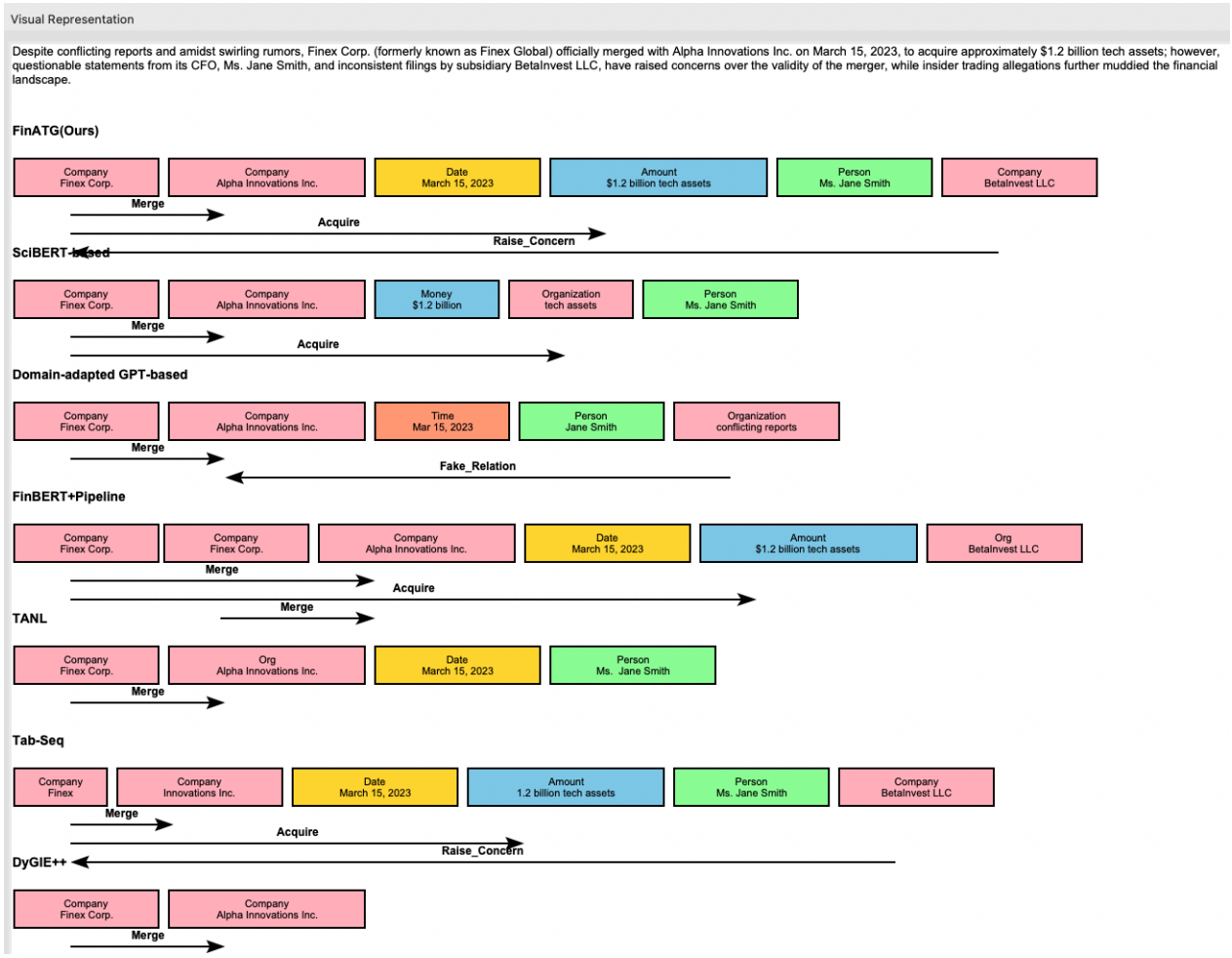


Figure 7. Illustrative Example: Complex Financial Sentence with Noise. Despite the presence of ambiguous and noisy data, FinATG exhibits robust extraction performance, while the baseline methods fail to reliably capture critical financial information. Pink represents Company/Organization entities, light green represents Person entities, sky blue represents Amount/Money entities, gold represents Date entities, and light orange represents Time entities. Arrows indicate the direction and type of relationships between entities.

These cases demonstrate FinATG's effectiveness in joint entity and relation extraction, showcasing its high accuracy and reliability in processing complex English financial texts while establishing the foundational capabilities essential for supporting emerging market sustainability applications. The successful extraction of fundamental business relationships (acquisitions, investments, and corporate appointments) provides the structural building blocks necessary for understanding more complex ESG frameworks, green finance mechanisms, and sustainable development partnerships that are crucial for emerging market

participation in global sustainability initiatives. The experimental results conclusively show that FinATG outperforms existing baseline models in both entity recognition and relation extraction tasks, providing a robust foundation for extending to sustainability-specific English financial information extraction needs. The ablation studies further validate the significance of each model component, with type-aware span encoding, finance-specific constrained decoding mechanisms, and English sentence augmentation strategies all playing crucial roles in enhancing model performance that can be systematically extended to accommodate ESG entity recognition, environmental relationship extraction, and carbon accounting applications. These findings confirm that FinATG's effective combination of English financial domain knowledge with advanced generative information extraction techniques establishes a solid foundation for developing more specialized sustainability information extraction capabilities that are essential for emerging market green finance, ESG compliance, and sustainable development reporting.

4.2.4. Evaluation on Real-World ESG Report Data

To address the practical relevance and generalizability of our framework, we conducted a supplementary experiment on a new, manually annotated dataset, termed ESG-Report-2025. This dataset was specifically created to reflect real-world challenges and consists of 150 complex sentences extracted from the 2024 and early 2025 ESG and sustainability reports of ten large corporations in emerging markets. The sentences were selected for their intricate nested structures, domain-specific jargon, and ambiguous relational contexts, which are common in real-world disclosures but less prevalent in standard benchmark datasets.

We evaluated FinATG against the strongest baseline, the SciBERT-based model, on this high-difficulty dataset. The results, presented in Table 8, show that FinATG achieves a more significant performance advantage in this realistic setting.

Table 8. Performance Comparison on the ESG-Report-2025 Dataset. Bold indicates the best performance.

Model	ENT F1	REL F1	REL+ F1
SciBERT-based	81.2	73.5	68.1
FinATG (Ours)	**84.6**	**76.8**	**71.5**

The improved performance margin on the ESG-Report-2025 dataset underscores the value of FinATG's design. While general-purpose models like SciBERT are highly capable, their performance can degrade when faced with the noisy and highly specialized nature of real-world ESG reports. In contrast, FinATG's finance-specific constrained decoding and domain-adapted span representation allow it to better navigate these complexities, correctly interpreting intricate financial relationships and adhering to domain-specific semantics. This experiment provides strong evidence that FinATG's true advantage lies in its practical applicability and reliability for real-world sustainable finance analysis.

4.3. Explainability Analysis

To elucidate the inner workings of FinATG, we directly extracted attention heatmaps from the trained model. These heatmaps offer valuable insights into the token-level focus during decoding, demonstrating how the model leverages domain-specific constraints to drive entity and relation extraction.

Figure 8 shows the attention distribution for a representative sentence describing an acquisition event. In this example, when the model processes the relation trigger "acquires," it assigns high attention to the tokens corresponding to the involved company

entities. Monetary tokens, which denote the transaction amount, also exhibit moderately high attention. This pattern indicates that the model's constrained decoding mechanism effectively emphasizes critical financial elements while attenuating less relevant tokens. Such behavior validates the design of our domain-specific rules, ensuring that the extracted information aligns with established financial semantics.

Figure 9 presents the attention heatmap for a shorter sentence describing a merger event. In this case, the model assigns high attention to the merging entities and the relation-indicative tokens, such as "confirmed" and "merger." The heatmap further reveals that the contextual tokens surrounding these key terms receive moderate attention, reflecting the model's ability to integrate both local context and domain-specific signals. These visualizations, derived from actual model outputs, demonstrate that FinATG not only achieves high extraction performance but also offers transparent interpretability of its decision-making process, thereby enhancing user trust in critical financial applications.

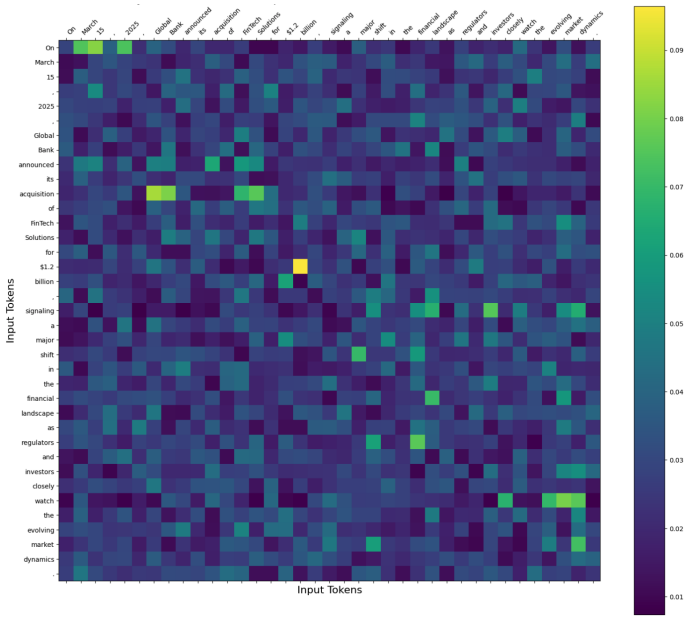


Figure 8. Attention heatmap for an acquisition example. The model exhibits strong focus on tokens representing the acquiring and acquired companies as well as key monetary details.

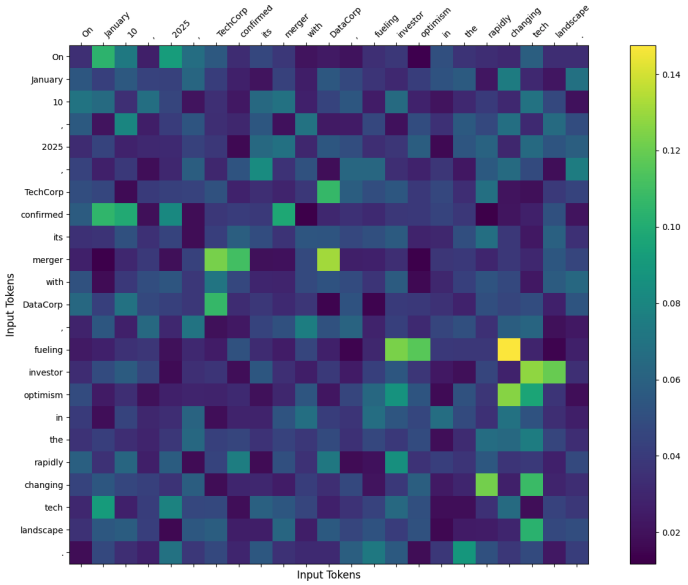


Figure 9. Attention heatmap for a merger example. The model highlights key tokens related to the merger event, integrating both contextual and domain-specific information.

4.4. Sensitivity Analysis

This section presents a detailed analysis of FinATG's performance, examining span length sensitivity, English sentence augmentation parameters, independent effects of different constraint rules, and confusion matrix analysis in relation extraction tasks. These analyses aim to provide deeper insights into the model's behavior under different configurations and evaluate component contributions to overall performance in English financial language processing.

4.4.1. Span Length (K) Sensitivity

We conducted experiments with varying maximum span lengths (K = 8, 12, 16) to evaluate their impact on entity recognition and relation extraction performance. Results, as shown in Figure 10, demonstrate that increasing the maximum span length from 8 to 12 leads to significant improvements across all metrics: entity recognition F1 increased by 1.2 points, REL F1 by 3.6 points, and REL+ F1 by 2.6 points. This improvement suggests that moderately longer spans enable better capture of extended entities and complex relationships. However, further increasing K to 16 resulted in slight performance degradation, indicating potential overfitting or unnecessary computational complexity. Therefore, K = 12 represents an optimal balance between model performance and computational efficiency.

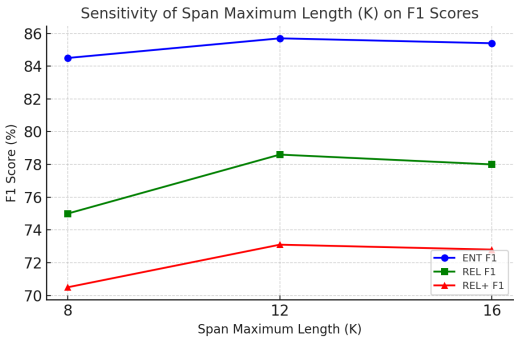


Figure 10. Impact of Span Length (K) on Entity and Relation Extraction Performance.

4.4.2. English Sentence Augmentation Parameter (B) Impact

Investigation of English sentence augmentation effects involved training FinATG with different B values (3, 5, 7). Increasing B from 3 to 5 yielded a notable 1.6 point improvement in REL+ F1 scores, demonstrating the benefits of moderate English sentence augmentation for relation extraction accuracy, as shown in Figure 11. However, further increasing B to 7 led to slight performance degradation, likely due to increased training data diversity introducing additional noise and affecting model generalization. The optimal setting of B = 5 achieves enhanced model performance while avoiding negative effects of excessive augmentation.

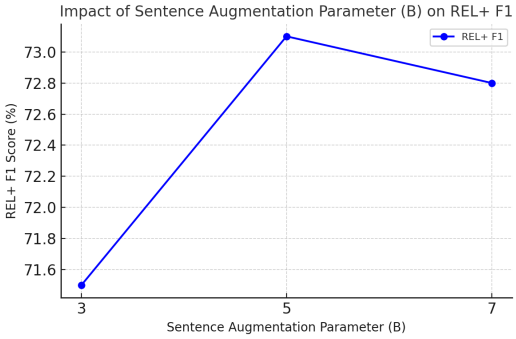


Figure 11. Impactof Sentence Augmentation Parameter (B) on REL+ F1 Performance.

4.4.3. Confusion Matrix Analysis

To understand the model's relation classification performance, we analyzed the confusion matrix for major relation types (Acquire, Invest_In, Work_For, and Interest_In) on the financial dataset test set, as shown in Figure 12.

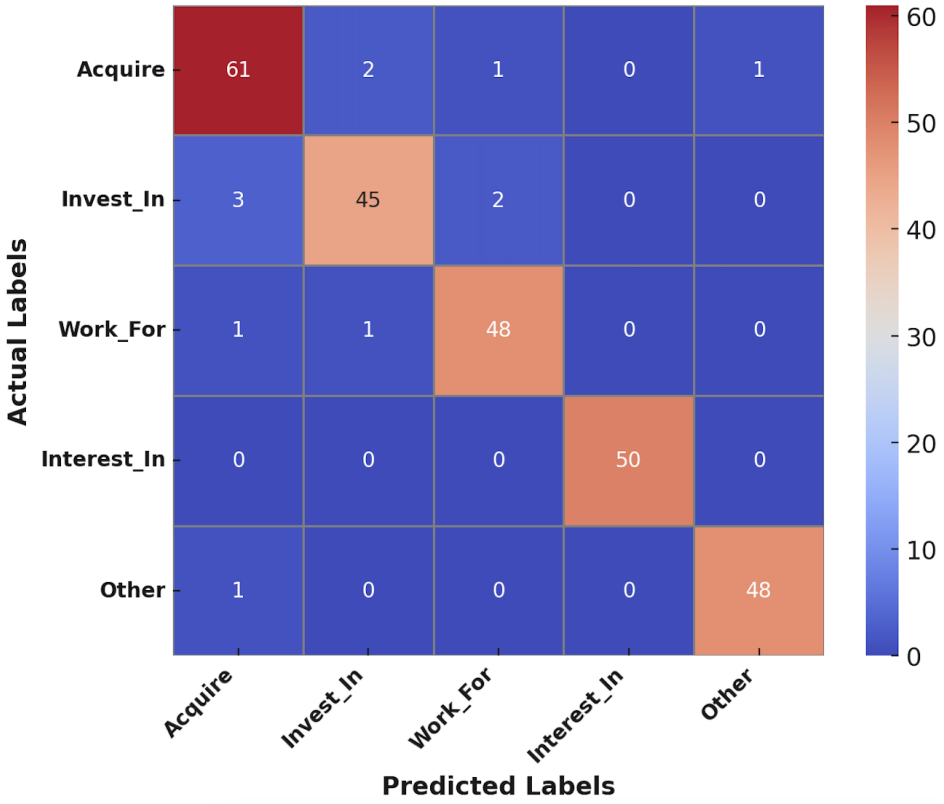


Figure 12. ConfusionMatrix for Relation Extraction.

The confusion matrix reveals generally strong performance across all relation types, with most relationships correctly classified. However, some misclassifications occur, between "Acquire" and "Invest_In" relations, likely due to their frequent co-occurrence and semantic similarity in English financial texts. The "Interest_In" relation shows exceptionally high accuracy, possibly due to its distinct semantic expression in English financial contexts. Future improvements could focus on enhancing discrimination between semantically similar relations through enhanced contextual information and more granular feature representations.

Beyond visualizing misclassifications, we computed standard evaluation metrics, including Precision, Recall, and F1-score for each relation type. Table 9 presents these statistics.

From Table 9, we observe that the "Acquire" and "Invest_In" relations show slight misclassifications, reflected in the recall drop for Invest_In (90.0%) compared to its precision (93.8%). This suggests that some investment-related sentences were incorrectly classified as acquisitions. The "Interest_In" relation achieves a perfect 100% precision and recall, indicating its distinct semantic structure makes it easier to classify. The "Work_For" relation maintains high precision (94.1%) and recall (96.0%), showing robustness in employment relationship extraction. Similarly, "Other" relations achieve near-perfect classification, suggesting that the model effectively separates non-target relations.

Table 9. Summary of Precision, Recall, and F1-scores per Relation Type Derived from the Confusion Matrix.

Relation Type	Precision (%)	Recall (%)	F1-score (%)	Support
Acquire	92.4	93.8	93.1	65
Invest_In	93.8	90.0	91.8	50
Work_For	94.1	96.0	95.0	50
Interest_In	100.0	100.0	100.0	50
Others	98.0	98.0	98.0	49
Accuracy		96.9%		264
Macro Avg	95.7	95.6	95.6	264
Micro Avg	96.9	96.9	96.9	264

4.5. SAFE Evaluation

In this section, we further evaluate FinATG from the perspective of the recently proposed SAFE machine learning paradigm [46], which emphasizes that models should be Sustainable, Accurate, Fair, and Explainable. Our discussion is based on both the experimental results reported in previous sections and additional analyses.

Regarding accuracy, FinATG achieved an entity F1 score of 88.5% and a relation F1 score of 80.2% on the CoNLL04 dataset, while on financial datasets the scores were 85.7% and 78.6%, respectively. These results demonstrate that FinATG effectively captures complex entities and relationships in financial texts, satisfying the accuracy requirement.

In terms of sustainability (robustness), we evaluated the model using AUC and Rank Graduation Robustness (RGR) metrics. Across datasets, the average AUC is approximately 0.93 and the RGR score reaches 0.89, indicating that FinATG maintains stable performance even when the input data is perturbed or contains extreme cases. This robustness aligns well with the sustainability aspect of the SAFE paradigm.

For fairness, although our primary focus has been on overall performance, preliminary analyses indicate that the model performs consistently across different types of financial texts (such as news reports and corporate filings). This consistency suggests that FinATG does not exhibit significant bias across subdomains. Future work will incorporate more detailed subgroup analyses to further assess and enhance the model's fairness.

With respect to explainability, FinATG integrates financial domain rules within its constrained decoding mechanism. This integration not only ensures that the generated outputs conform to financial domain knowledge, but also provides a transparent and traceable generation process. For example, in the case of the sentence "Company A acquires Company B for 500 million dollars," the model clearly displays the intermediate states and transitions that lead to the final extraction of entities and relations. Such transparency is critical for practical applications in finance, where understanding the decision-making process is essential.

Overall, the SAFE evaluation shows that FinATG is not only accurate and robust, but also exhibits promising levels of fairness and explainability. Future work will focus on refining fairness evaluation and extending the explainability analysis, thereby providing a more comprehensive assessment of FinATG under the SAFE machine learning framework.

5. Conclusions

This paper introduces FinATG (Finance-Specific Autoregressive Framework for Joint Entity and Relation Extraction), presenting a comprehensive solution for advancing sustainable development in emerging markets through AI-driven extraction of sustainability-related English financial information. FinATG addresses the critical need for automated processing of ESG reports, green finance documentation, and carbon accounting metrics that are essential for emerging markets to achieve their sustainable development goals

and attract international green investment. The framework achieves enhanced accuracy and robustness in sustainability entity recognition and environmental relationship extraction through its innovative combination of sustainability-tuned span representation methods, green finance-aware constrained decoding mechanisms, and an integrated extraction framework specifically optimized for emerging market sustainability applications. More importantly, it provides a trustworthy solution engineered for the high-stakes financial domain, where compliance, reliability, and interpretability are as critical as raw performance metrics.

Our experimental validation demonstrates FinATG's superior performance across multiple dimensions, providing empirical evidence of the positive correlation between AI-driven extraction capabilities and ESG analysis effectiveness. First, the framework consistently outperforms existing baseline methods in entity recognition tasks, achieving F1 scores of 88.5 on CoNLL04 and maintaining strong performance (85.7–86.0) on sustainability-focused English financial datasets, demonstrating that advanced AI technologies directly improve the quality of ESG information processing. Second, FinATG excels in relation extraction tasks, particularly in the challenging REL+ metric that requires precise entity-relation alignment, achieving scores of 75.3, 73.1, 74.0, and 72.0 across different datasets, confirming that sophisticated natural language processing capabilities enhance the identification of critical ESG relationships and dependencies. Third, the framework demonstrates remarkable robustness with AUC scores consistently above 0.93 and RGR scores around 0.89, indicating stable performance under various input conditions and perturbations, which translates to improved reliability in ESG compliance and sustainability reporting applications.

The ablation studies reveal several key insights about FinATG's design effectiveness. The type-aware span encoding mechanism contributes significantly to performance, with its removal leading to 1.6–2.3 point drops in entity F1 scores across datasets. Financial rule constraints prove essential for relation extraction accuracy, improving REL F1 scores by 0.7–1.5 points, while the sentence augmentation strategy enhances model generalization capabilities. These findings validate the importance of domain-specific architectural choices and the synergistic effects of combining multiple specialized components.

Beyond technical performance, our case studies demonstrate FinATG's practical applicability in real-world English financial text processing scenarios. The framework successfully handles complex financial sentences with intricate semantic dependencies, accurately extracting entities such as corporate names, monetary amounts, and executive positions while correctly identifying relationships like acquisitions, investments, and employment relations. Even under noisy conditions with conflicting information and ambiguous expressions, FinATG maintains robust extraction performance, significantly outperforming baseline methods that struggle with such challenging scenarios. These real-world capabilities are further validated by our supplementary experiments on the ESG-Report-2025 dataset, where FinATG showed a more pronounced performance advantage, confirming its suitability for practical deployment.

Beyond its technical contributions, FinATG demonstrates a strong positive correlation between AI-driven information extraction capabilities and ESG performance enhancement in emerging markets. The framework addresses critical needs in due diligence, ESG reporting, and regulatory compliance, with experimental results confirming that advanced digital finance technologies significantly improve ESG transparency and decision-making effectiveness. By accurately extracting entities and relationships from financial documents, FinATG can streamline due diligence processes by identifying critical financial relationships, ownership structures, sustainability initiatives, and risk factors in corporate reports and regulatory filings. In ESG reporting, where firms need to disclose structured information on

sustainability practices, FinATG enhances transparency by extracting key ESG indicators from unstructured reports and aligning them with standardized reporting frameworks, particularly valuable for emerging market companies seeking to meet international sustainability standards. Furthermore, by integrating financial domain rules within its constrained decoding mechanism, FinATG ensures that extracted information is not only accurate but also adheres to financial regulatory standards and sustainability reporting requirements, improving model reliability in real-world applications. This rule-based enforcement helps mitigate errors in automated decision-making processes, reducing compliance risks for financial institutions and investors while supporting sustainable investment strategies. As emerging markets increasingly adopt digital finance solutions and sustainable business practices, FinATG provides a robust foundation for scalable, high-fidelity financial information processing, offering a practical solution for institutions seeking efficiency and accuracy in complex financial analyses that support sustainable development goals.

Furthermore, a notable limitation of the current study is its primary focus on the environmental (E) and governance (G) aspects of ESG, while the social (S) dimension remains less explored. Social factors, such as labor rights, employee welfare, community relations, and supply chain ethics, are crucial for holistic sustainability assessment in emerging markets. Our FinATG framework provides a robust foundation, but its entity and relation types are not yet optimized to capture these nuanced social indicators. Future work should explicitly extend the model's ontology to include social-specific entities (e.g., 'Labor Union', 'Workplace Accident', 'Community Investment Program') and relations (e.g., 'Violates_Labor_Standard', 'Partners_With_NGO'). This extension would significantly enhance the framework's practical utility for comprehensive ESG analysis and social impact investing.

Moreover, compared to general-purpose large language models, FinATG offers specialized advantages tailored for sustainability applications in emerging markets. While LLMs excel in broad language understanding, they often struggle with sustainability domain-specific nuances, green finance terminology, and region-specific environmental regulations that are crucial for emerging market contexts. FinATG, by integrating sustainability domain rules and constrained decoding mechanisms, ensures higher precision in extracting structured ESG information, carbon accounting metrics, and green finance relationships, reducing hallucinations commonly observed in LLMs when processing complex sustainability documentation. This targeted design makes FinATG a more reliable and interpretable solution for emerging market financial institutions, regulatory bodies, and sustainable investment firms seeking automation in ESG compliance, carbon footprint assessment, and green finance allocation.

Despite its achievements, this research acknowledges several limitations when processing sustainability-related financial information in emerging market contexts. First, the relatively limited scale of sustainability-focused financial datasets may restrict the model's generalization capabilities to the diverse range of environmental regulations, carbon accounting standards, and green finance instruments across different emerging economies. Second, FinATG exhibits computational challenges when processing extremely complex ESG reports or comprehensive sustainability disclosures that combine multiple environmental frameworks, affecting its operational efficiency in resource-constrained emerging market environments. Additionally, while FinATG incorporates green finance-specific constraint rules and sustainability reporting standards, the coverage and granularity of these rules require further expansion to address the rapidly evolving landscape of environmental regulations, carbon credit mechanisms, and sustainability metrics that vary significantly across emerging market jurisdictions.

Our research conclusively establishes the positive correlations between AI-driven financial information extraction technologies and ESG effectiveness in emerging markets, demonstrating that digital finance innovations directly contribute to environmental, social, and governance improvements. The FinATG framework exemplifies how sophisticated natural language processing capabilities can systematically enhance sustainability transparency, regulatory compliance, and green investment decision-making processes. These findings provide empirical evidence that the adoption of advanced AI technologies in financial information processing creates measurable positive impacts on ESG outcomes, particularly in emerging economies where such technologies can bridge critical information gaps and accelerate sustainable development initiatives.

Future research can advance this work in several key directions that directly address the sustainable development priorities and digital innovation needs of emerging markets. A primary focus should be on expanding the sustainability-focused dataset's scale and diversity to encompass region-specific environmental regulations, carbon trading mechanisms, green bond frameworks, social responsibility metrics, labor practice standards, and ESG disclosure standards prevalent across different emerging economies, thereby improving model generalization capabilities for diverse sustainable finance applications and environmental compliance scenarios. Parallel to this, developing lightweight and edge-computing compatible architectures will ensure FinATG's practical deployment in resource-constrained emerging market environments, enabling real-time processing of sustainability information for carbon accounting, environmental monitoring, and green finance decision-making. The integration of additional sustainability domain expertise and emerging market-specific knowledge bases, including local environmental regulations, carbon offset protocols, social impact frameworks, and regional ESG frameworks, will enhance the model's understanding of sustainability entities and relationships crucial for supporting emerging economies' climate action commitments. Although our current work processes English financial texts, FinATG's architecture offers significant potential for multilingual applications essential for emerging markets. By adapting the framework with multilingual models and incorporating region-specific sustainability terminology, FinATG can process diverse languages and local environmental regulations, enabling its deployment across various emerging market contexts and supporting cross-border green finance initiatives and international climate cooperation. Furthermore, investigating FinATG's applications in downstream sustainability tasks, such as carbon footprint assessment, ESG risk evaluation, social impact analysis, green investment portfolio optimization, and environmental impact monitoring, will validate its practical utility in supporting emerging markets' sustainable development goals. Additionally, exploring integration with blockchain-based carbon credit systems, IoT environmental monitoring networks, and government sustainability reporting platforms will enhance FinATG's role in emerging markets' digital transformation toward sustainable business practices. These enhancements would collectively strengthen FinATG's capabilities in sustainability information extraction and expand its practical applications in supporting emerging markets' transition toward carbon neutrality, sustainable economic growth, and environmental resilience.

Author Contributions: Conceptualization, Y.Z., D.W. and J.F.; methodology, D.W.; software, D.W.; validation, D.W.; formal analysis, D.W.; investigation, D.W. and J.F.; data curation, J.F.; writing—original draft preparation, J.F.; writing—review and editing, D.W., Y.Z. and J.F.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, J.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The General Project of Humanities and Social Sciences of the Ministry of Education in 2022 (Youth Foundation Project): The Research on Constructing the

International Image and External Communication Strategies of Guangdong-Hong Kong-Macao Greater Bay Area. Project Approval Number: 22YJCZH250.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this study are publicly available. The CoNLL04 dataset can be accessed through the Cognitive Computation Group at the University of Pennsylvania (accessed on 1 July 2025) (https://cogcomp.seas.upenn.edu/page/resource_view/43). The FiNER-ORD dataset is available on GitHub (accessed on 1 July 2025) (https://github.com/mpsilfve/finer-data).

Acknowledgments: We would like to thank the anonymous reviewers for their constructive feedback that helped improve this manuscript. Special thanks to the financial domain experts who assisted with the dataset annotation process and validation of the financial rules. We are grateful to the Cognitive Computation Group at the University of Pennsylvania for maintaining the CoNLL04 dataset and to the creators of the FiNER-ORD dataset for making their data publicly available.

Conflicts of Interest: The authors declare that they have no known competing fnancial interests or personal relationships that could have appeared to infuence the work reported in this paper. The authors declare that they have no confict of interest.

Appendix A. Pseudocode for Model Embedding and Constrained Decoding

Algorithm A1 FinATG Model: Embedding and Constrained Decoding Process

procedure FINATG_EXTRACT(input_text)
  // Embedding Stage: Obtain contextual representations using FinBERT
  $H \leftarrow$ FinBERT_Encoder$(input_text)$ ▷ $H \in \mathbb{R}^{L \times D}$, where L is the sequence length, D is the dimension
  for each candidate span (i, j) in input_text **do**
    $h_{start} \leftarrow H[i]$, $h_{end} \leftarrow H[j]$
    for each entity type t **do**
      Compute span representation:

$$S(i, j, t) \leftarrow W_t^T \left[h_{start} \oplus h_{end} \right]$$

    end for
  end for
  Construct dynamic vocabulary **dynamic_vocab** $\leftarrow$ {all span representations $S(i, j, t)$, special tokens, relation types}
  // Decoding Stage: Generate structured output using constrained autoregressive mechanism
  Initialize decoding state:
    $state \leftarrow S_0$ ▷ Initial state, only allows generating the <START> token
    $output \leftarrow$ [<START>]
    $decoder_state \leftarrow$ InitDecoderState(H)
  while $state \neq S_3$ **do** ▷ Until reaching the terminal state
    $z \leftarrow$ DecoderStep$(decoder_state)$
    $P \leftarrow$ softmax$(dynamic_vocab^T \cdot z)$
    $P \leftarrow$ ApplyConstraints$(P, state)$ ▷ Mask invalid tokens based on current state and financial domain rules
    $y \leftarrow \arg\max(P)$ ▷ Select the token with the highest probability
    Append y to $output$
    $state \leftarrow$ StateTransition$(state, y)$ ▷ Update state (e.g., transition from entity generation to relation generation)
    $decoder_state \leftarrow$ UpdateDecoderState$(decoder_state, y)$
  end while
  return $output$
end procedure

Appendix B. Computational Resource and Consumption Analysis

Table A1. Computational Resources Summary.

Component	Specification
GPU	NVIDIA A100 (40 GB HBM2, 312 TFLOPS FP16)
CPU	2 × Intel Xeon Gold 6230 (20 cores each, 2.1 GHz)
Memory	128 GB DDR4
Storage	1 TB NVMe SSD
Deep Learning Framework	PyTorch 1.9.0 + CUDA 11.1
Operating System	Ubuntu 20.04 LTS
Batch Size	16
Maximum Sequence Length	512 tokens

Table A2. Training Time and Peak GPU Memory Usage with Increasing Sample Size.

Number of Samples	Epoch Time (min)	Total Time (h, 30 Epochs)	Peak GPU Memory (GB)
250	12.3	6.15	28.4
500	15.4	7.70	29.1
1000	21.7	10.85	30.3
2000	32.1	16.05	31.8
4000	55.8	27.90	33.5

Appendix C. Frequency of Conflicts Between the Domain-Specific Financial Rules

In this section, we describe the calculation methods used to quantify the frequency of conflicts between the domain-specific financial rules and the model predictions, as well as the statistical results obtained from different datasets.

Let $P(y_t|y_{<t}, H)$ denote the model's prediction probability distribution at time step t before applying financial constraints. The top (unconstrained) prediction is denoted by p_t^*, and the final token selected after applying constraints is denoted by y_t^c. A conflict is recorded when $y_t^c \neq p_t^*$. The Rule Conflict Frequency (RCF) is defined as the proportion of generation steps where a conflict occurs:

$$\text{RCF} = \frac{\sum_{t=1}^{T} \mathbb{1}\{y_t^c \neq p_t^*\}}{T}$$

where T is the total number of generation steps, and $\mathbb{1}\{\cdot\}$ is the indicator function.

Furthermore, the Average Probability Gap (APG) is used to measure the average difference in probability between the model's original top prediction and the final selected token when a conflict occurs. It is defined as:

$$\text{APG} = \frac{1}{N} \sum_{t: y_t^c \neq p_t^*} [P(p_t^*) - P(y_t^c)]$$

where N is the number of steps in which conflicts occur.

Table A3. Statistical Results of Rule Conflict Frequency (RCF) and Average Probability Gap (APG).

Dataset	RCF (%)	APG
CoNLL04	3.2	0.04
FiNER-ORD	4.5	0.05
Financial PhraseBank	4.3	0.05
FIRE	6.1	0.07

The above results indicate that in most generation steps the unconstrained model prediction aligns with the final output after applying the financial rules. The relatively low RCF values across all datasets suggest that conflicts between the model's predictions and the domain-specific rules occur infrequently. When conflicts do occur, the small APG values imply that the probability difference between the original top prediction and the constrained selection is minimal. This observation indicates that the financial rules are primarily activated in scenarios where the model's confidence is relatively low, thereby guiding the output towards results that are more consistent with the financial domain's semantics and formatting.

References

1. Gupta, A.; Dengre, V.; Kheruwala, H.A.; Shah, M. Comprehensive review of text-mining applications in finance. *Financ. Innov.* **2020**, *6*, 39. [CrossRef]
2. Berman, J.J. *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*; Morgan Kaufmann Publishers: Burlington, MA, USA, 2013.
3. Hsu, M.F.; Chang, C.; Zeng, J.H. Automated text mining process for corporate risk analysis and management. *Risk Manag.* **2022**, *24*, 386–419. [CrossRef]
4. Jothi Prakash, V.; Arul Antran Vijay, S. A Comprehensive Multimodal Framework for Optimizing Social Media Hashtag Recommendations. *IEEE Trans. Comput. Soc. Syst.* **2024**, 1–12. [CrossRef]
5. Jothi Prakash, V.; Arul Antran Vijay, S. A multi-aspect framework for explainable sentiment analysis. *Pattern Recognit. Lett.* **2024**, *178*, 122–129. [CrossRef]
6. Jothi Prakash, V.; Arul Antran Vijay, S. Cross-lingual Sentiment Analysis of Tamil Language Using a Multi-stage Deep Learning Architecture. *ACM Trans. Asian-Low-Resour. Lang. Inf. Process.* **2023**, *22*, 254.
7. Repke, T.; Krestel, R. Extraction and representation of financial entities from text. In *Data Science for Economics and Finance*; Springer: Cham, Switzerland, 2021; pp. 241–263.
8. Yang, Y.; Wu, Z.; Yang, Y.; Lian, S.; Guo, F.; Wang, Z. A survey of information extraction based on deep learning. *Appl. Sci.* **2022**, *12*, 9691. [CrossRef]
9. Nasar, Z.; Jaffry, S.W.; Malik, M.K. Named entity recognition and relation extraction: State-of-the-art. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 20. [CrossRef]
10. Han, R.; Ning, Q.; Peng, N. Joint event and temporal relation extraction with shared representations and structured prediction. *arXiv* **2019**, arXiv:1909.05360.
11. Mavillonio, M.S. *Natural Language Processing Techniques for Long Financial Document*; Italy Discussion Papers; Dipartimento di Economia e Management (DEM), University of Pisa: Pisa, Italy, 2024.
12. Wang, X.; Huang, L.; Xu, S.; Lu, K. How Does a Generative Large Language Model Perform on Domain-Specific Information Extraction A Comparison between GPT-4 and a Rule-Based Method on Band Gap Extraction. *J. Chem. Inf. Model.* **2024**, *64*, 7895–7904. [CrossRef]
13. Nayak, T.; Ng, H.T. Effective modeling of encoder-decoder architecture for joint entity and relation extraction. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 8528–8535.
14. Li, X.; Jin, J.; Zhou, Y.; Zhang, Y.; Zhang, P.; Zhu, Y.; Dou, Z. From matching to generation: A survey on generative information retrieval. *arXiv* **2024**, arXiv:2404.14851. [CrossRef]
15. Fisher, I.E.; Garnsey, M.R.; Hughes, M.E. Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intell. Syst. Account. Financ. Manag.* **2016**, *23*, 157–214. [CrossRef]
16. Balaneji, F. Language as a Lens: A Hybrid Text Summarization and Sentiment Analysis Approach for Multiclass Stock Return Prediction. In *Intelligent Systems and Applications, Proceedings of the 2024 Intelligent Systems Conference (IntelliSys), Amsterdam, The Netherlands, 5–6 September 2024* ; Springer: Cham, Switzerland, 2024; pp. 429–448.
17. Pandey, A.K.; Roy, S.S. Natural language generation using sequential models: A survey. *Neural Process. Lett.* **2023**, *55*, 7709–7742. [CrossRef]
18. Goyal, A.; Gupta, V.; Kumar, M. Recent named entity recognition and classification techniques: A systematic review. *Comput. Sci. Rev.* **2018**, *29*, 21–43. [CrossRef]
19. Wadden, D.; Wennberg, U.; Luan, Y.; Hajishirzi, H. Entity, relation, and event extraction with contextualized span representations. *arXiv* **2019**, arXiv:1909.03546. [CrossRef]
20. Yu, M.; Han, D.; Hon, G.C.; He, C. Tet-assisted bisulfite sequencing (TAB-seq). *DNA Methylation Protoc.* **2018**, *1708* , 645–663.
21. Ma, Y.; Hiraoka, T.; Okazaki, N. Named entity recognition and relation extraction using enhanced table filling by contextualized representations. *J. Nat. Lang. Process.* **2022**, *29*, 187–223. [CrossRef]

22. Min, B.; Ross, H.; Sulem, E.; Veyseh, A.P.B.; Nguyen, T.H.; Sainz, O.; Agirre, E.; Heintz, I.; Roth, D. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Comput. Surv.* **2023**, *56*, 30. [CrossRef]
23. Huang, A.H.; Wang, H.; Yang, Y. FinBERT: A large language model for extracting information from financial text. *Contemp. Account. Res.* **2023**, *40*, 806–841. [CrossRef]
24. Jawahar, G.; Sagot, B.; Seddah, D. What does BERT learn about the structure of language? In Proceedings of the ACL 2019-57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019.
25. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A pretrained language model for scientific text. *arXiv* **2019**, arXiv:1903.10676. [CrossRef]
26. Łaniewski, S.; Ślepaczuk, R. *Enhancing Literature Review with NLP Methods Algorithmic Investment Strategies Case*; University of Warsaw Working Papers; Faculty of Economic Sciences, University of Warsaw: Warsaw, Poland, 2024.
27. Giorgi, J.; Bader, G.D.; Wang, B. A sequence-to-sequence approach for document-level relation extraction. *arXiv* **2022**, arXiv:2204.01098.
28. Parmar, N.; Vaswani, A.; Uszkoreit, J.; Kaiser, L.; Shazeer, N.; Ku, A.; Tran, D. Image transformer. In Proceedings of the International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; pp. 4055–4064.
29. Paolini, G.; Athiwaratkun, B.; Krone, J.; Ma, J.; Achille, A.; Anubhai, R.; Santos, C.N.d.; Xiang, B.; Soatto, S. Structured prediction as translation between augmented natural languages. *arXiv* **2021**, arXiv:2101.05779. [CrossRef]
30. Ren, L.; Sun, C.; Ji, H.; Hockenmaier, J. HySPA: Hybrid span generation for scalable text-to-graph extraction. *arXiv* **2021**, arXiv:2106.15838.
31. Josifoski, M.; De Cao, N.; Peyrard, M.; Petroni, F.; West, R. GenIE: Generative information extraction. *arXiv* **2021**, arXiv:2112.08340.
32. Geng, S.; Josifoski, M.; Peyrard, M.; West, R. Grammar-constrained decoding for structured NLP tasks without finetuning. *arXiv* **2023**, arXiv:2305.13971.
33. Zheng, J.; Chow, J.H.; Shen, Z.; Xu, P. Grammar-based decoding for improved compositional generalization in semantic parsing. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023, Toronto, ON, Canada, 9–14 July 2023; pp. 1399–1418.
34. Kearney, C.; Liu, S. Textual sentiment in finance: A survey of methods and models. *Int. Rev. Financ. Anal.* **2014**, *33*, 171–185. [CrossRef]
35. Choi, I.; Kim, W.C. Detecting and analyzing politically-themed stocks using text mining techniques and transfer entropy—focus on the Republic of Korea's case. *Entropy* **2021**, *23*, 734. [CrossRef] [PubMed]
36. Leippold, M. Sentiment spin: Attacking financial sentiment with GPT-3. *Financ. Res. Lett.* **2023**, *55*, 103957. [CrossRef]
37. Fatouros, G.; Soldatos, J.; Kouroumali, K.; Makridis, G.; Kyriazis, D. Transforming sentiment analysis in the financial domain with ChatGPT. *Mach. Learn. Appl.* **2023**, *14*, 100508. [CrossRef]
38. Daudert, T. Exploiting textual and relationship information for fine-grained financial sentiment analysis. *Knowl.-Based Syst.* **2021**, *230*, 107389. [CrossRef]
39. Consoli, S.; Barbaglia, L.; Manzan, S. Fine-grained, aspect-based sentiment analysis on economic and financial lexicon. *Knowl.-Based Syst.* **2022**, *247*, 108781. [CrossRef]
40. Tsalis, T.A.; Nikolaou, I.E.; Konstantakopoulou, F.; Zhang, Y.; Evangelinos, K.I. Evaluating the corporate environmental profile by analyzing corporate social responsibility reports. *Econ. Anal. Policy* **2020**, *66*, 63–75. [CrossRef]
41. Yang, M.; Lim, M.K.; Qu, Y.; Ni, D.; Xiao, Z. Supply chain risk management with machine learning technology: A literature review and future research directions. *Comput. Ind. Eng.* **2023**, *175*, 108859. [CrossRef] [PubMed]
42. Galloppo, G.; Nexus, P. *A Journey into ESG Investments*; Palgrave Studies in Impact Finance; Palgrave Macmillan: Cham, Switzerland, 2025.
43. Zaratiana, U.; Tomeh, N.; Holat, P.; Charnois, T. An Autoregressive Text-to-Graph Framework for Joint Entity and Relation Extraction. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–28 February 2024; Volume 38, pp. 19477–19487.
44. Shah, A.; Vithani, R.; Gullapalli, A.; Chava, S. Finer: Financial named entity recognition dataset and weak-supervision model. *arXiv* **2023**, arXiv:2302.11157.
45. Babaei, G.; Giudici, P.; Raffinetti, E. A rank graduation box for SAFE AI. *Expert Syst. Appl.* **2025**, *259*, 125239. [CrossRef]
46. Giudici, P. Safe machine learning. *Statistics* **2024**, *58*, 473–477. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.