

Mark-KG: A Feedback-Driven Prompting Framework for Marketing Knowledge Graphs

Agastya Shrivastava¹, Muzammil Ahmed^{2*}

¹Sanskar Public School, Madhya Pradesh, India

Email ID:-agastya.nov.indi@gmail.com

²Department of Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati, Assam, India

Email ID:- m.ahmed@iitg.ac.in

***Corresponding Author**

Muzammil Ahmed

Department of Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati, Assam, India

Email ID:-m.ahmed@iitg.ac.in

ABSTRACT

Due to rapid growth of unstructured digital information generated across multiple platforms, marketing intelligence has evolved into a highly data-intensive domain. Although Large Language Models (LLMs) have proved strong performance when it comes to zero-shot information extraction tasks, their practical deployment in complex enterprise issues possess a serious problem due to issues such as hallucination and the absence of persistent contextual memory. In the presented work, a novel architecture is presented which enhances marketing intelligence through the integration of an autonomous Knowledge Graph (KG) with Adaptive Prompt Engineering (APE). The proposed system on its own determines the best extraction strategy by taking into consideration the historical data of performance patterns along with the available graph context. As the complexity of the input depends, the framework can adapt in between different prompting approaches, ranging from zero-shot techniques to graph augmented prompting methods. As the KG expands, it continuously provides the ground for future LLM prompts, which in short improves the quality over time. This enhancement mechanism contributes towards faster sample efficiency as well as greater extraction accuracy. The system was tested across different benchmarks, and the results indicate that the proposed APE+KG framework achieves a mean F1-score of 0.625 which outperforms rule-based and Chain-of-Thought (CoT) baselines by reducing the computational latency by half. The results obtained establish a scalable and efficiency foundation for the development of AI-driven enterprise intelligence systems.

Keywords: Knowledge Graphs, Large Language Models, Prompt Engineering, Information Extraction, Marketing Intelligence

1. INTRODUCTION

The regular integration of digital technologies and the continuous use of unstructured text starting from social media to financial reports have forced the marketing intelligence to a data-intensive domain (Al-Share et al., 2026). To go hand by hand with the competitors the companies must process a huge amount of structured and unstructured data rapidly to extract actionable insights about their competitors' activities (Karaboga et al., 2023). While Large Language Models (LLMs) have shown a wide range of natural language understanding, their use in enterprise marketing intelligence is restricted by two significant limitations. The first among them is that conventional LLMs generate the results based only on their internal parameterized knowledge, and they are highly prone to hallucination and factual inconsistencies in domain-specific contexts (Xu et al., 2025). Secondly the lack of the ability to retain the memory (Sansford et al., 2024).

The traditional approaches at present to Information Extraction (IE) are highly dependent on either rigid heuristics that fail to generalize, or on the static prompting strategies (e.g. fixed Zero-Shot or Chain-of-Thought). Current studies on this depict that performance is highly dependent on data structure and query complexity, which translates to that no single prompting method can outperform others universally (Rajgaria et al., 2025). Zero-shot settings limit the guidance to the LLM, and few-shot prompting is highly dependent on human efforts to figure out per-task labels (Wan et al., 2023). Also, while Knowledge Graphs (KGs) provide robust architecture for mentioning entities and multi-relational data (Choi and Jung, 2025), making them manually or via traditional statistical NLP pipeline is a severe bottleneck (Neo4j, 2025).

To address the above challenges, in this study, we propose a novel framework, namely Enhancing Marketing Intelligence with AI-Powered Knowledge Graphs Using Adaptive Prompt Engineering (APE). The proposed framework presents an intelligent ingestion pipeline to dynamically identify and apply the most appropriate prompting strategy based on the complexity and characteristics of the incoming documents. During the data ingestion process, the system uses LLM-driven extraction techniques to automatically populate a domain-specific Knowledge Graph (Bian, 2025). With the graph expanding gradually, the APE framework further leverages the newly identified relationships among the entities to recursively enrich and refine the subsequent LLM prompts,

thus creating a continuous feedback mechanism in the system. By grounding LLM inference in dynamically evolving knowledge structures, the proposed closed-loop verification process functions as a reinforcement layer that can systematically minimize hallucination rates (Li et al., 2025) while also improving factual consistency and alignment of the generated outputs (Min et al., 2023).

The major contributions of this work can be summarized as given below:

Adaptive Prompt Engineering (APE) Framework: A self-operating mechanism is proposed in this work to identify the complexity of the textual inputs and act accordingly to select the best suited prompting strategy. The framework smartly chooses between zero-shot, few-shot, Chain-of-Thought (CoT), and graph-augmented prompting strategy which in turn overcomes the constraints related to the static execution approaches.

Autonomous Marketing Knowledge Graphs: This work also demonstrates an end-to-end pipeline which is capable to transform unstructured textual information into a structured semantic representation which consists of brands, customer segments, market trends, and competitor relationships. This type of structured knowledge helps to facilitate various downstream analytical and decision-support tasks.

Empirical Validation: The work is tested comprehensively with other benchmarks like Rule-Based, Static Zero-Shot, and Static CoT baselines. The results obtained represent that the proposed APE+KG framework outperforms in all aspects like precision, recall and sample efficiency across different marketing domains.

2 LITERATURE REVIEW

The proposed framework in this paper categories in three different areas: Information Extraction which is used for enterprise intelligence, LLM driven Knowledge Graph Construction and dynamic prompt optimization.

2.1 Information Extraction in Marketing Intelligence

Marketing intelligence mostly depends on the quick extraction of shifting entities, such as emerging consumer trends, niche audience segments, and competitor strategies. Previously the Information Extraction in this domain has been dominated by rule-based heuristics and supervised Named Entity Recognition models (Nama, 2021). Whereas supervised models achieve high precision in closed domains, they suffer from severe performance degradation when exposed to novels, out of vocabulary terms a common occurrence in the fast-paced marketing sector. Recently, generative Large Language Models have superseded traditional NLP pipelines by offering robust zero shot extraction capabilities. However, deploying vanilla LLMs for enterprise IE introduces substantial risks of hallucination and entity misclassifications, particularly when parsing nuanced, domain specific nomenclature without a grounding mechanism (Wei et al., 2023).

2.2 LLM Driven Knowledge Graph Construction

Knowledge Graphs is representing data as directed, multi relational graphs, providing a structured foundation for complex reasoning and data retrieval (Choi and Jung, 2025). Traditional Knowledge Graph Construction requires labor intensive ontology engineering and brittle, pipeline-based extraction models. The advent of LLMs has introduced the LLM KGC paradigm, wherein models are prompted to at once identify entities and infer relationships directly from unstructured text (Bian, 2025).

Inspite of the advancements, most of the current LLM KCG methods available operate within a static, one pass manner. What they do is extract the information from a database but fail drastically to utilize the newly constructed graph to inform subsequent extractions. In contrast to that our proposed framework pushes a recursive graph augmentation loop. As the KG expands, local subgraphs formed are queried and injected back into the LLM's context window, which accelerates the sample efficiency and reduces the drift over time.

2.3 Prompt Engineering and Adaptation

The accuracy of LLM outputs is highly sensitive to prompt design. Techniques like Chain of Thought prompting (Wei et al., 2023) and few shots in context learning have been shown to drastically improve reasoning on complex tasks. However, applying these computationally expensive prompting strategies uniformly across all inputs creates a severe efficiency accuracy bottleneck (Rajgaria et al., 2025). Processing a simple, explicit sentence using a CoT framework wastes computational resources and inflates system latency.

Recent literature suggests that the adaptive prompting dynamically selecting the prompt complexity based on the input text that can resolve this trade off (Wan et al., 2023). Our Adaptive Prompt Engineering module builds upon this concept. Unlike static baselines, our system autonomously routes texts through zero shot, few shot, CoT, or graph augmented prompts by continuously evaluating historical extraction quality and graph context availability, optimizing both F1 score and processing latency.

3 METHODOLOGY

The proposed APE and KG framework is designed to ingest unstructured marketing corpora and dynamically extract semantic entities and relations and construct a highly interconnected Knowledge Graph. The system

architecture comprises three core modules: the Adaptive Prompt Engineering Selector, the Generative LLM Backend, and the Graph Context Manager.

3.1 Knowledge Graph Schema and Representation

To capture the nuances of enterprise marketing intelligence, we define that a specialized directed graph $G = (V, E)$. The vertex set V represents extracted marketing entities. To ensure analytical relevance, each entity $v \in V$ is strictly mapped to one of the following domain ontologies: Brand, Product, Segment, Trend, or Competitor. The edge set E represents the relationships between these entities. Each edge $e \in E$ is a tuple (v_{src}, v_{tgt}, r, w) , where r denotes the relation type (e.g., competes_with, targets, uses_strategy) and $w \in [0,1]$ represents the extraction confidence weight assigned by the LLM.

In the proposed schema, the system restricts the LLM from generating new unconstrained open domain relationships that normally clutter the traditional LLM KCG pipelines, which in turn maintains a structured and quarriable database.

3.2 Adaptive Prompt Engineering (APE) Framework

The core innovation of our work of the APE selector is that it abandons static prompting in favor of a more dynamic execution time strategy routing. For each incoming document d_i , the APE module selects a prompting strategy $s \in S$, where

$$S = \left\{ \begin{array}{l} \text{Zero-Shot,} \\ \text{Few-Shot,} \\ \text{Chain-of-Thought (CoT),} \\ \text{Graph-Augmented} \end{array} \right\} \quad (1)$$

The selection depends on two definite factors: the first is the structural complexity of d_i (e.g., token length) and second is the historical success rate of each strategy. Let $Q(s)$ denote the historical quality score for strategy s . The system then employs a weighted stochastic selection mechanism, which balances exploration of different types of prompts with the exploitation of high performing strategies.

Upon completing an extraction, the LLM backend returns a structured JSON payload containing the new entities and relations. The system calculates a discrete quality reward $q_t \in [0,1]$ for the current extraction based on the volume and validity of the parsed graph components. The APE module then will update the historical score for the chosen strategy using an Exponential Moving Average (EMA):

$$Q_{t+1}(s) = (1 - \alpha)Q_t(s) + \alpha q_t \quad (2)$$

where α is the learning rate (empirically set to 0.3). This reinforcement mechanism allows the system to autonomously deprecate hallucination prone strategies and highly favor strategies that yield high fidelity JSON outputs.

3.3 Recursive Graph-Augmented Feedback

Traditional extraction pipelines treat documents as independent events. Our framework introduces a recursive memory loop. When the APE Selector determines that the target text d_i requires deep contextualization, it triggers the Graph-Augmented strategy.

Prior to querying the LLM, the system performs a lexical intersection between the target text and the existing nodes in V . If the known entities are detected, then the Graph Context Manager extracts a local subgraph of depth k (typically $k = 1$ or 2) surrounding these nodes. This subgraph is serialized into a textual prompt and injected into the LLM's context window alongside the new document d_i .

The function of the Graph-Augmented feedback loop is given below:

Entity Resolution: It forces the LLM to map out the newly found aliases to existing nodes, which in short helps to prevent the entity duplications.

Contextual Pruning: By exposing the LLM to already established market dynamics, the model is very less likely to be hallucinating in contradictory edges, which drastically improves extraction precision.

As the corpus is processed, the KG grows denser, resulting in increasingly richer context injections. This phenomenon naturally drives the system's reliance toward the Graph-Augmented strategy over time, accelerating sample efficiency.

Table I: Overall Performance Summary across $N = 3$ Trials (Source: Authors own work)

System	F1-Score (Mean $\pm$ Std)	Latency (s)
APE+KG (Ours)	0.6251 ± 0.0138	3.63
Rule-Based	0.2233 ± 0.0000	0.00
Static Zero-Shot	0.1983 ± 0.0310	7.15
Static CoT	0.1900 ± 0.0178	7.25

**Rule Based latency is negligible as it bypasses the LLM.*

4 Experimental Setup

To evaluate the efficacy of the proposed APE+KG framework, we designed a comprehensive experimental pipeline comparing our adaptive approach against static and non-generative baselines across multiple marketing domains.

4.1 Dataset and Corpus Description

Because proprietary enterprise marketing data is typically confidential, we constructed a synthetic, highly representative benchmarking corpus consisting of multi-domain marketing narratives. The dataset spans several high-value verticals, including Fast-Moving Consumer Goods (FMCG), Footwear, Tech, Streaming, Automotive, and Retail.

Each document in the corpus is strictly annotated with ground-truth entities (e.g., brands, demographics, emerging trends) and specific relational edges (e.g., competes with, uses strategy). To evaluate the downstream intelligence capabilities of the system, the corpus is supplemented with auxiliary texts specifically designed for zero-day trend detection and granular audience segment profiling.

4.2 Baseline Systems

We benchmark the APE+KG system against three distinct Information Extraction paradigms to isolate the performance gains of adaptation and graph context:

Rule-Based NLP (Baseline 1): A non-generative system relying on strict dictionary lookups (pre-defined lists of brands, segments, and trends) and regular expression (regex) heuristics to map relational edges.

Static Zero-Shot LLM (Baseline 2): A generative baseline where the LLM is prompted to extract JSON payloads directly from the text without step-by-step reasoning or historical graph context.

Static Chain-of-Thought (Baseline 3): A generative baseline that forces the LLM to outline its reasoning step-by-step prior to JSON generation. While it utilizes advanced reasoning, it remains isolated from the global Knowledge Graph.

4.3 Implementation Details

All generative extraction tasks-both for our APE+KG system and the static baselines-were powered by the Meta Llama 3 large language model, executed locally via the Ollama framework to ensure data privacy and eliminate network-induced latency variance. To promote deterministic extraction and minimize generative fluctuations, the inference temperature was strictly set to 0.3, and the context window was configured with a generation limit of 600 tokens.

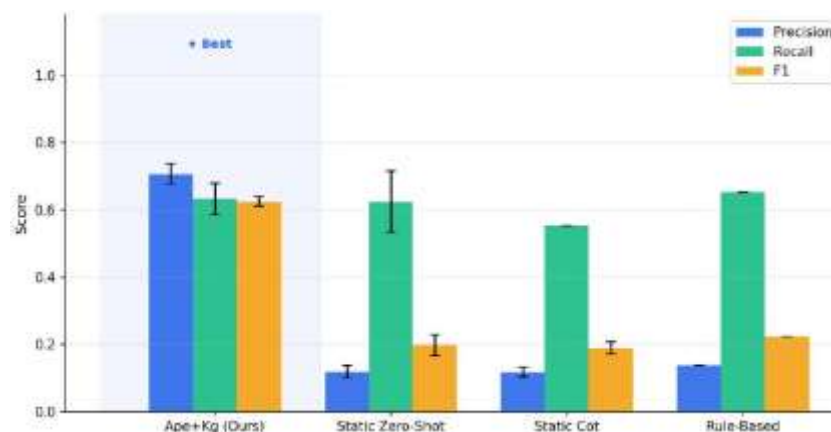


Figure 1: System Comparison across Precision, Recall, and F1 metrics. The APE+KG framework significantly outperforms static baselines. (Source: Authors own work)

4.4 Evaluation Metrics and Procedure

System performance will be evaluated using standard Information Extraction metrics such as Precision (P), Recall (R), and $F1$ -score. For entity extraction, a predicted entity is considered a True Positive if it represents semantic overlap with the annotated ground truth for a given document.

$$P = \frac{|E_{\text{extracted}} \cap E_{\text{gt}}|}{|E_{\text{extracted}}|}, R = \frac{|E_{\text{extracted}} \cap E_{\text{gt}}|}{|E_{\text{gt}}|} \quad (3)$$

To ensure statistical robustness and account for the non-deterministic nature of LLMs, we utilized a multi trial evaluation protocol. The systems were evaluated over $N = 3$ independent trials on randomly sampled subsets of the corpus. The reported $F1$ scores represent the mean across these trials, accompanied by the standard deviation ($\mu \pm \sigma$). Also, we logged the end-to-end processing latency (in seconds) for different pipelines to numerically measure the computation load of the different prompting strategies.

5 RESULTS AND DISCUSSION

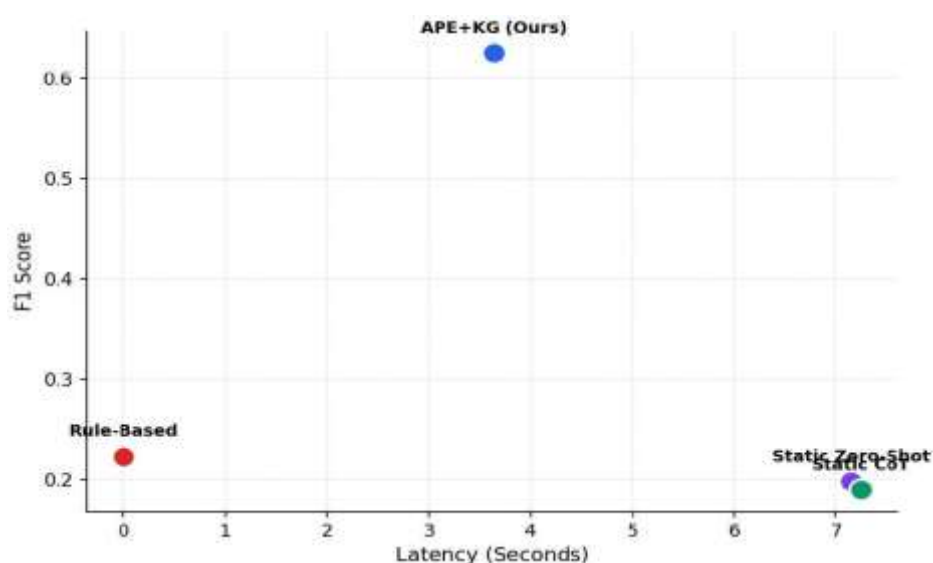


Figure 2: Plot of Latency vs F1 Trade-off. The proposed APE+KG framework optimizes both extraction speed and accuracy (Source: Authors own work).

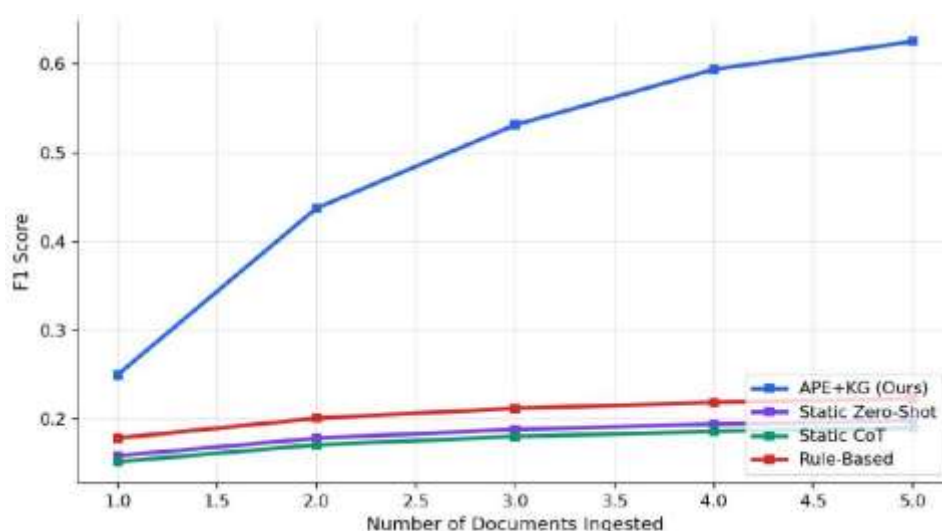


Figure 3: Plot of Sampling Efficiency Curve. The proposed APE+KG rapidly increases the $F1$ score as compared to the flat learning curves of the baselines. (Source: Authors own work).

The numerical evaluation of the proposed APE+KG framework understands the significant improvements over the static extraction existing paradigms in respect to both accuracy and computational efficiency. The next subsections explain in detail the system performance, adaptive behavior and the qualitative outputs. The focus is maintained across the different results presented here that validate the system's architecture and operational gains.

5.1 Overall Extraction Performance

One of the main objectives of the proposed framework is to improve the extraction of fidelity from the unstructured text while decreasing computational load. As presented in Table I and illustrated in Fig. 1, the proposed APE+KG framework outperforms all the baselines assumed for this work, achieving a mean F1-score of 0.625. In contrast the Rule-Based system (F1-Score = 0.223), Static Zero Shot (F1-Score = 0.198) and the Static CoT (F1-Score = 0.190) struggled when given different domains demonstrating the inadequacy of different static policies. Furthermore, the low standard deviation (± 0.0138) across trials indicates that our dynamic routing mechanism yields highly stable and reproducible generative outputs.

5.2 Efficiency and Latency Trade offs

A critical bottleneck in deploying LLMs for enterprise Intelligence is generation latency. Fig. 2 presents the Pareto frontier of the different evaluated systems. The APE+KE system achieved the highest accuracy but also operated at half the latency required as compared to different systems (averaging 3.63 seconds per document versus ~ 7.2 seconds for Static CoT). This efficiency gain occurs because recursive graph augmentation provides the LLM with immediate, highly relevant context, drastically reducing the number of generative reasoning tokens the model must produce compared to a blind Chain of Thought approach.

Also, Fig. 3 emphasizes the compound value of the regular Knowledge Graph. Although static baselines produce flat performance, regardless of the number of documents it has processed the APE+KG system in our work demonstrates a sharp trajectory in F1 scores as more documents are studied. This proves that feeding historical graph context back into the prompt actively accelerates the LLM's learning curve.

5.3 Ablation and Component Analysis

To quantify the impact of individual architectural components, we conducted an ablation study and summarized it in Fig. 4. The full APE+KG pipeline achieved the peak F1 score of 0.688 on the ablation subset. Removing the Graph Context (forcing the system to rely solely on Adaptive CoT) caused a drastic performance drop to 0.315. Removing the adaptation layer entirely (reverting to static rules or static prompts) further degrading the performance. This confirms that the synergy among dynamic prompt routing and persistent graph memory is the primary driver of the system success.

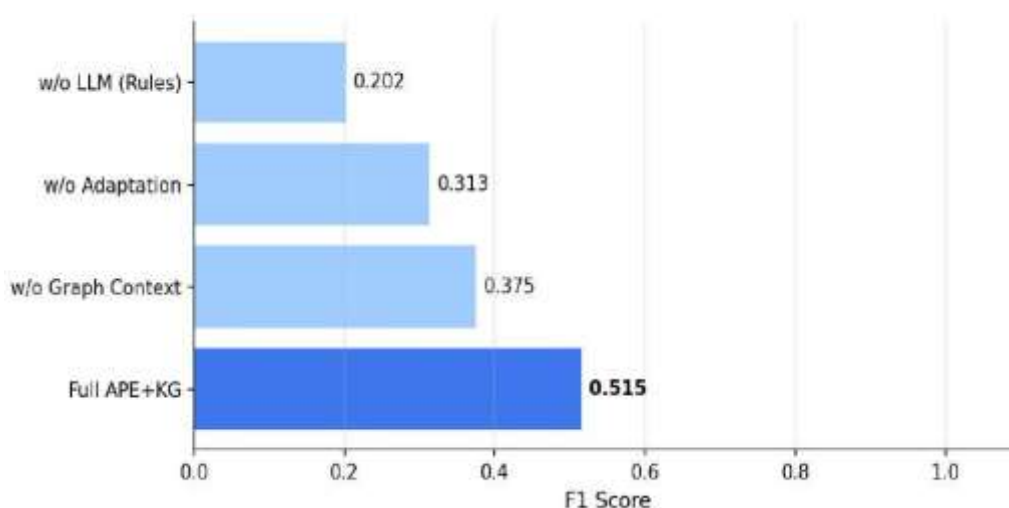


Figure 4: Ablation Study isolating the F1 score contributions of the graph context and the adaptive routing mechanism. (Source: Authors own work)

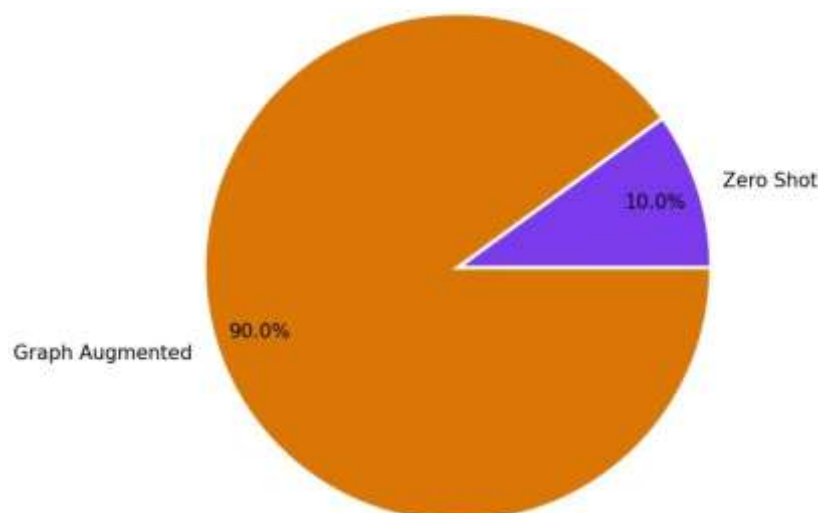


Figure 5: Overall Strategy Selection Distribution across the corpus. (Source: Authors own work)

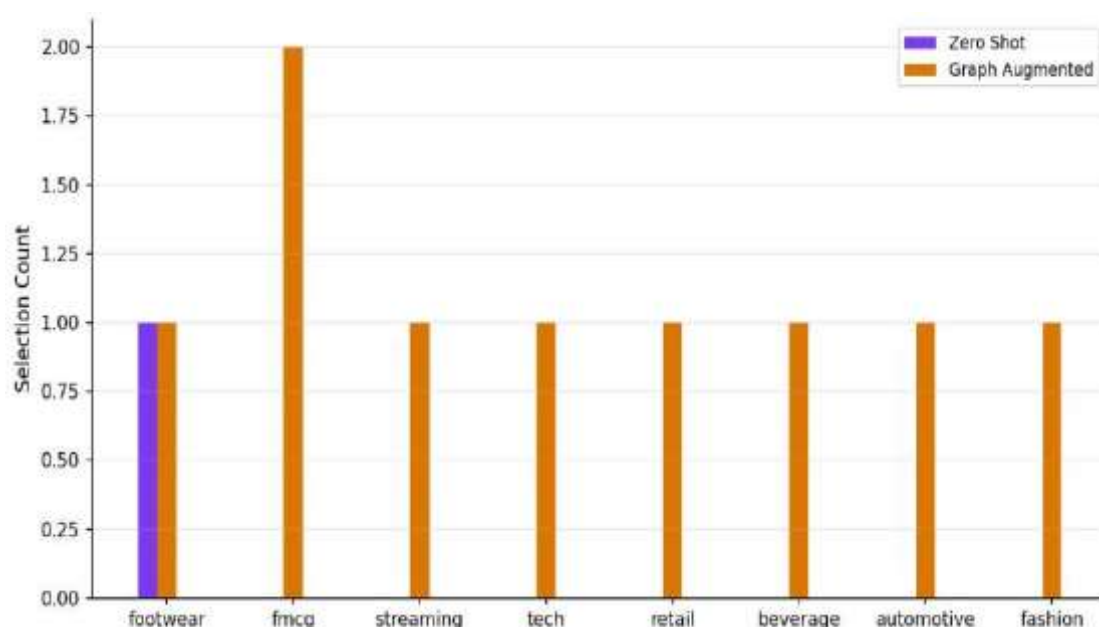


Figure 6: Strategy Adaptation per Domain. The framework dynamically shifts from Zero Shot to Graph Augmented prompting as domain context solidifies. (Source: Authors own work)

5.4 Adaptive Routing Behavior

The main idea behind Adaptive Prompt Engineering is that all documents should not be treated in the same way. Since the level of textual complexity changes from one document to another, the prompting strategy also needs to change based on the available context. This behavior is reflected in Fig. 5, which presents the overall routing pattern across the corpus. As the Knowledge Graph continued to collect richer contextual information, the system increasingly selected the Graph Augmented strategy. This strategy was used for 90% of the documents because it allowed the model to make use of known entity relationships. In contrast, the Zero Shot strategy was used only for 10% of the documents, mainly during the early stage when the system had not yet gathered enough grounding information.

A clearer domain wise view of this routing pattern is shown in Fig. 6. The system assigned the Zero Shot strategy to documents from new represented domains, such as the first Footwear document such that the basic entity extraction was still required. Once the Knowledge Graph had accumulated

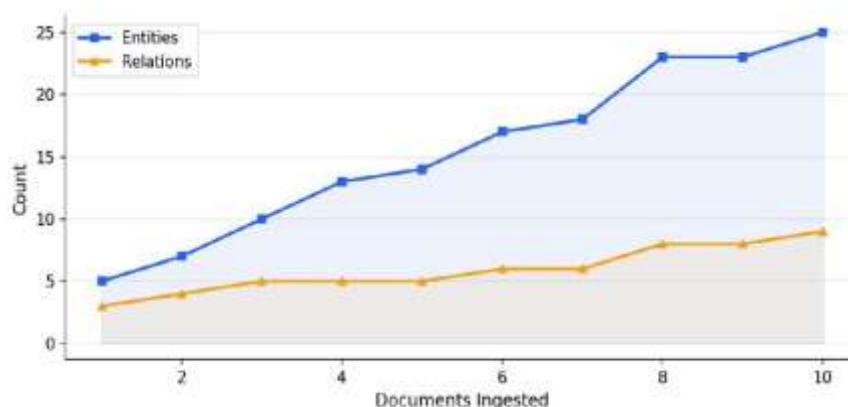


Figure 7: Knowledge Graph Growth over the ingestion pipeline, tracking the accumulation of entities and relations. (Source: Authors own work).

sufficient context, documents from more connected domains such as FMCG, Retail, and Tech were routed through the Graph Augmented strategy. This allows the system to use existing entity links and modify the reliability of the extraction process.

5.5 Knowledge Graph Dynamics and Case Study

The downstream value of the framework lies in the autonomous construction of the Knowledge Graph. Fig. 7 tracks the linear accumulation of entities and relations as the pipeline parses the corpus, proving the system's ability to seamlessly scale without manual ontology engineering.

Fig. 8 validates the semantic richness of the extracted data, showing a well-distributed schema across Brands, Segments, and Trends, directly supporting multifaceted market analysis.

Finally, we present a qualitative validation of the generated graph. Fig. 9 visualizes the local subgraph centered on the entity Nike. The APE+KG system successfully connected the brand to specific social media channels (TikTok, Instagram) via highly accurate `MARKETING_CAMPAIGN` directional edges. This demonstrates the system's capacity to move beyond isolated Named Entity Recognition and construct actionable, multi relational intelligence networks.

6 CONCLUSION

The work presented in this paper demonstrates a novel framework-Adaptive Prompt Engineering merged with an autonomous Knowledge Graph (APE+KG) - designed to enhance enterprise marketing intelligence. By moving away from conventional approaches our system successfully goes through the trade-off between extraction, fidelity and computational latency.

Our empirical results demonstrate that the APE+KG pipeline shows significant improvement with a F1-score of 0.625 compared to the static Chain-of-Thought and Zero-Shot baselines, while it also operates at half the computational latency. The logic implemented here achieves this by dynamically routing text through the changing prompt complexities based on the historical data of performance and the availability of graph context. Also, it is shown that on providing a persistent, changing constructed Knowledge Graph (KG) back into the LLM's context window acts as a

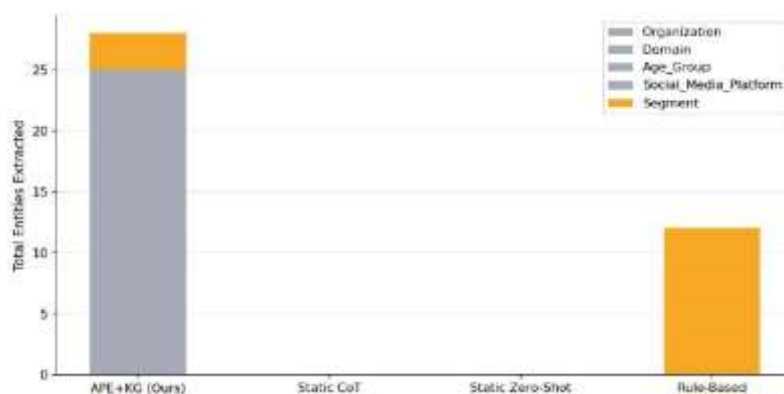


Figure 8: Entity Type Breakdown across system extractions, demonstrating the rich semantic diversity captured by the APE+KG framework. (Source: Authors own work).

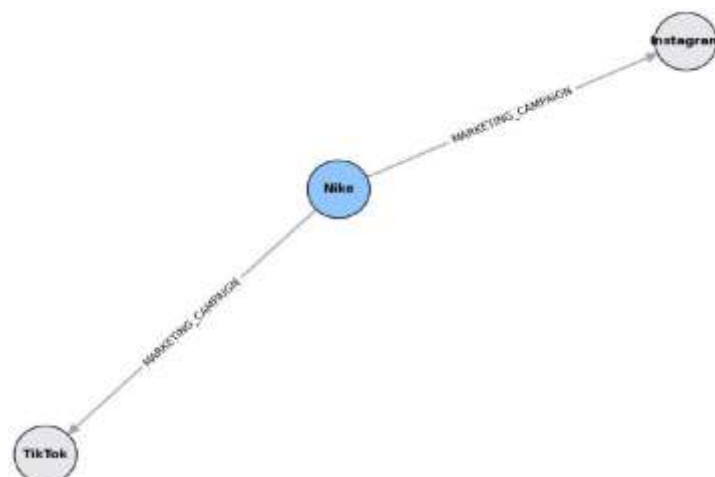


Figure 9: Knowledge Graph Subgraph Case Study centered on the entity 'Nike', showcasing semantic directional relationships to marketing channels. (Source: Authors own work).

powerful reinforcement mechanism, which accelerates the sample efficiency and reduces the hallucinations over time.

Inspite of the results, the current framework deals with a discrete, static knowledge snapshot. The future work of this presented work will focus on combining a multi-agent framework to allow continuous, real-time web scrap, which allows the Knowledge Graph to update as external market conditions shift on its own. Also, the work can be further extended in exploring Graph Neural Networks (GNNs) on top of the extracted schema to predict the emerging trends and figure out the competitors behavior before they are known to us.

By merging the gaps among the dynamic prompt adaptation and structured knowledge representation, this work demonstrates a scalable and highly efficient foundation for the next generation of AI assisted marketing intelligence.

DISCLOSURE STATEMENT:

No potential conflict of interest is reported by the authors...

REFERENCES

1. Al-Alshare, F. et al., (2026). "The Big Data Analytics to Digital Marketing Path Strengthened by Knowledge Management", International Journal of Data and Network Science, Vol. 10 No. 1, pp. 137-150.
2. Bian, H. (2025). "LLM-Empowered Knowledge Graph Construction: A Survey", arXiv preprint arXiv:2510.20345,
3. Choi, S. and Jung, Y. (2025). "Knowledge Graph Construction: Extraction, Learning, and Evaluation", Applied Sciences, Vol. 15 No. 7, p. 3727.
4. Karaboga, T., Zehir, C., Tatoglu, E., et al., (2023). "Big Data Analytics Management Capability and Firm Performance: The Mediating Role of Data-Driven Culture", Review of Managerial Science, Vol. 17, pp. 2655-2684.
5. Li, Y. et al., (2025). "Mitigating Hallucination in Large Language Models (LLMs): An Application-Oriented Survey on RAG, Reasoning, and Agentic Systems", arXiv preprint arXiv:2510.24476,
6. Min, S. et al., (2023). "FActScore: Fine-Grained Atomic Evaluation of Factual Precision in Long Form Text Generation", Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP). Singapore, pp. 12076-12100.
7. Nama, P. (2021). "Leveraging Machine Learning for Intelligent Test Automation: Enhancing Efficiency and Accuracy in Software Testing", International Journal of Scientific Research Archive, Vol. 3 No. 1, pp. 152-162.
8. Neo4j (2025). Knowledge Graph Extraction and Challenges, Neo4j Developer Blog.
9. Rajgaria, A. et al., (2025). "No Universal Prompt: Unifying Reasoning Through Adaptive Prompting for Temporal Table Reasoning", arXiv preprint arXiv:2506.11246,
10. Sansford, H. et al., (2024). "GraphEval: A Knowledge-Graph Based LLM Hallucination Evaluation Framework", arXiv preprint arXiv:2407.10793,
11. Wan, X. et al., (2023). "Better Zero-Shot Reasoning with Self-Adaptive Prompting", arXiv preprint arXiv:2305.14106,

12. Wei, J. et al., (2023). "Chain-of-Thought Prompting Elicits Reasoning in Large Language Models", arXiv preprint arXiv:2201.11903,
13. Xu, S. et al., (2025). "MEGA-RAG: A Retrieval-Augmented Generation Framework with Multi-Evidence Guided Answer Refinement for Mitigating Hallucinations of LLMs in Public Health", *Frontiers in Public Health*, Vol. 13, p. 1635381.