Value to User's Voice: A Generative AI Framework for Actionable Insights from Customer Reviews in Consumer Electronics

Radhika Mundra
Samsung R&D India
m.radhika@samsung.com

Bhavesh Kukreja
Samsung R&D India
b.kukreja@samsung.com

Aritra Ghosh Dastidar
Samsung R&D India
aritra.d@samsung.com

Kartikey Singh
Samsung R&D India
kartikey.s@samsung.com

Javaid Nabi
Samsung R&D India
javaid.nabi@samsung.com

Abstract

Customer reviews are a valuable asset for businesses, especially in the competitive consumer electronics sector, where understanding user preferences and product performance is critical. However, extracting meaningful insights from these unstructured and often noisy reviews is a challenging task that typically requires significant manual effort. We present "Value to User's Voice" (V2U), a domain-adaptable generative AI framework designed for automating large-scale review processing and providing real-time actionable insights. At the core of V2U is the Data Enrichment Module (DEM), which automatically processes unstructured reviews by extracting key product features, filtering noise, and structuring the data for analysis. DEM enables V2U to adapt to the unique terminology and characteristics of different consumer electronics products, such as smartphones, TVs, and digital appliances. Paired with the Analytic Dashboard AI Module (ADAM) for real-time querying and visualizations, V2U allows businesses to quickly identify trends and respond to evolving customer needs. Evaluations show that V2U achieves a structured data extraction accuracy of 88%, comparable to GPT-4's 90%, while requiring 40% fewer computational resources. Additionally, V2U reduces manual analysis efforts by 70%, resulting in significantly faster time-to-insight.

1 Introduction

Customer reviews are a rich source of valuable insights, providing businesses with a window into user experiences, product preferences, and emerging trends (Oelke et al., 2009; Pankaj et al., 2019). By understanding the customer perspective product managers can identify product strengths and weaknesses, prioritize feature development, and respond to user concerns effectively. This feedback is crucial for optimizing products, fostering innovation, and ensuring customer satisfaction (Li et al., 2024).

However, the sheer volume of reviews generated daily, coupled with their diversity in languages, platforms, and the ways customers express opinions poses a significant challenge. Manual analysis is impractical due to the overwhelming amount of data, as well as the variability in language, slang, and sentiment nuances that make accurate interpretation difficult. Consequently, automating the analysis of customer reviews is essential to unlock their full potential and generate actionable insights efficiently.

1.1 Existing Approaches and Their Limitations

Two primary methods have been explored in automating review analysis: rule-based and model-based approaches. **Rule-based systems** rely on predefined rules and patterns to identify keywords, phrases, and sentiment indicators. While useful for highly structured and predictable data, they are rigid, require constant updates, and struggle with the contextual variability found in customer reviews. See Appendix A for examples of simple rules, including one example of failure.

Model-based approaches, leveraging machine learning and deep learning, offer more flexibility. They include supervised classification models like BERT (Devlin et al., 2019) and DistilBERT (Sanh et al., 2020) for sentiment analysis, as well as topic models like Latent Dirichlet Allocation (LDA) (Jelodar et al., 2018) for identifying underlying themes in reviews. While more powerful than rule-based systems, these models still face significant challenges. Supervised models require large amounts of labeled data, which makes them costly and time-consuming to train. Additionally, models trained on one domain often fail to generalize to new product categories, leading to inconsistent performance across different datasets.

Moreover, existing models struggle to handle more complex language patterns like sarcasm, irony, and subtle sentiment nuances. These limitations highlight the need for a more sophisticated, adaptive approach to fully leverage customer feedback for business insights.

1.2 Proposed Solution

To address these limitations, we propose **Value to User's Voice (V2U)**, a novel generative AI-based framework specifically designed to automate large-scale customer review analysis and data visualization. Existing models struggle with complex language patterns such as sarcasm and irony. V2U leverages the power of Large Language Models (LLMs) (Zhao et al., 2023) to better understand the nuanced context of reviews, including sarcasm and irony, and to transform unstructured, noisy reviews into structured, meaningful data ready for actionable insights.

The V2U framework consists of two core modules as illustrated in **Figure 1**:

1. **Data Enrichment Module (DEM)**: DEM employs LLMs to perform end-to-end data processing. It extracts key product features, refines noisy reviews, splits and rephrases them based on topic context and sentiment, and performs sentiment and review classification. DEM also calculates review severity scores and identifies entities, providing a rich, structured dataset for further analysis.

2. **Analytic Dashboard AI Module (ADAM)**: ADAM offers an intuitive interface for users to interact with structured data. Through custom filters and queries, it enables users to explore insights, generate dynamic reports, and visualize trends through summary charts and comparative analysis, all powered by LLM-driven automation.

In this paper, we provide a comprehensive exploration of the V2U framework and its two core modules DEM and ADAM. We will demonstrate how these modules leverage LLMs to overcome the challenges of existing rule-based and model-based methods, providing a flexible, scalable solution for customer review analysis. Finally, we present empirical evidence of V2U's performance, showing its effectiveness in extracting meaningful insights from large volumes of customer reviews in the consumer electronics domain.

2 Related Work

2.1 Literature Review

Extracting actionable insights from customer reviews has been a significant area of research within Natural Language Processing (NLP). Early efforts focused primarily on sentiment classification, determining whether a review expressed a positive, negative, or neutral sentiment (Chandra and Jana, 2020). While this approach offered a general overview of customer feedback, it lacked the granularity to analyze specific aspects that influence customer opinions.

To overcome this limitation, **Aspect-Based Sentiment Analysis (ABSA)** (Yan et al., 2021; Zhu et al., 2022; Negi et al., 2024) emerged as a method for identifying and analyzing sentiment directed towards specific product attributes, such as battery life or screen quality in electronic devices. ABSA provides a more nuanced understanding of customer feedback by associating sentiments with distinct product features.

The introduction of language models has further advanced ABSA, with models like **BERT** and **RoBERTa** (Liu et al., 2019) excelling in capturing complex linguistic relationships and extracting aspects and sentiments with high accuracy. Additionally, recent research, such as the **SOLAR framework** (Gao et al., 2023; Hanni et al., 2016), has built on ABSA to provide concise summaries of helpful reviews, integrating sentiment and topic information for a more comprehensive understanding of customer feedback.

2.2 Limitations of Existing Methods

Despite significant progress in sentiment extraction and review summarization, several challenges remain in directly deriving actionable insights from customer reviews. Existing methods often focus solely on sentiment extraction or topic-based summarization, but fall short of integrating these capabilities into a holistic framework for generating insights that businesses can act on.

Microsoft's 2024 work, AllHands: Ask Me Anything on Large-scale Verbatim Feedback via Large Language Models (Gao et al., 2023), proposed an analytical framework using LLMs for customer review interaction. While this framework offers advancements in review classification and abstractive

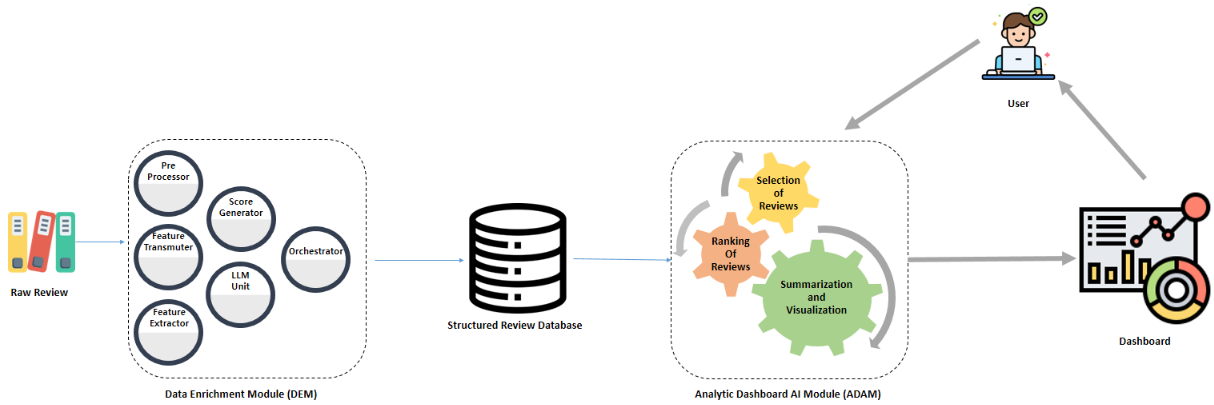

Figure 1: V2U Architecture

topic modeling, several limitations persist:

- **Review Segmentation:** Current approaches lack the ability to split reviews into granular segments based on sentiment and topic context, which is critical for detailed insights.
- **Abstractive Summarization:** Existing methods often fail to generate concise summaries that preserve the original review context after segmenting.
- **Reliance on Closed-Source LLMs:** Many frameworks rely on expensive, closed-source models like GPT-4 (Gao et al., 2023), which limit flexibility and affordability.
- **Review Ranking:** Few approaches address the ranking of reviews based on issue severity or frequency, leaving businesses without clear prioritization of critical issues.

3 Methodology

The V2U framework automates the transformation of unstructured customer reviews into actionable insights. This is achieved through two core components: the **Data Enrichment Module (DEM)** and the **Analytic Dashboard AI Module (ADAM)**. Each module plays a vital role in processing, filtering, and extracting insights from large-scale customer datasets.

3.1 Data Enrichment Module (DEM)

The Data Enrichment Module (DEM) is the cornerstone of the V2U framework, designed to automate the process of transforming unstructured customer reviews into structured data, ready for analysis and visualization. The DEM is composed of six critical components: Pre-Processor, Review Transmuter, Feature Extractor, Score Generator, LLM Unit, and Orchestrator, as depicted in the **Figure 2**. For a detailed flow of the DEM architecture, along with a sample input-output walkthrough, refer to **Appendix B**.

1. **Pre-Processor**:

   The pre-processor is responsible for cleaning and filtering the raw input data. It contains two blocks:

   - **Review Cleanser**: This block handles the removal of irrelevant elements such as URLs, hashtags, special characters, and other noise from the review data. The aim is to produce clean text that is free from extraneous symbols that might hinder analysis.
   - **Review Filter**: The review filter applies three different filters to eliminate unhelpful or irrelevant reviews:
     - **Gibberish Filter**: Detects and removes reviews that contain nonsensical or meaningless content, ensuring that these reviews do not negatively affect downstream analysis.
     - **Out-of-Context Filter**: This filter removes reviews that are unrelated to the domain of interest. For example, if the focus is on digital appliances reviews about unrelated categories like fashion or food are discarded.
     - **Non-Informative Review Filter**: Reviews that merely list product names, features, or vendor names without providing any meaningful

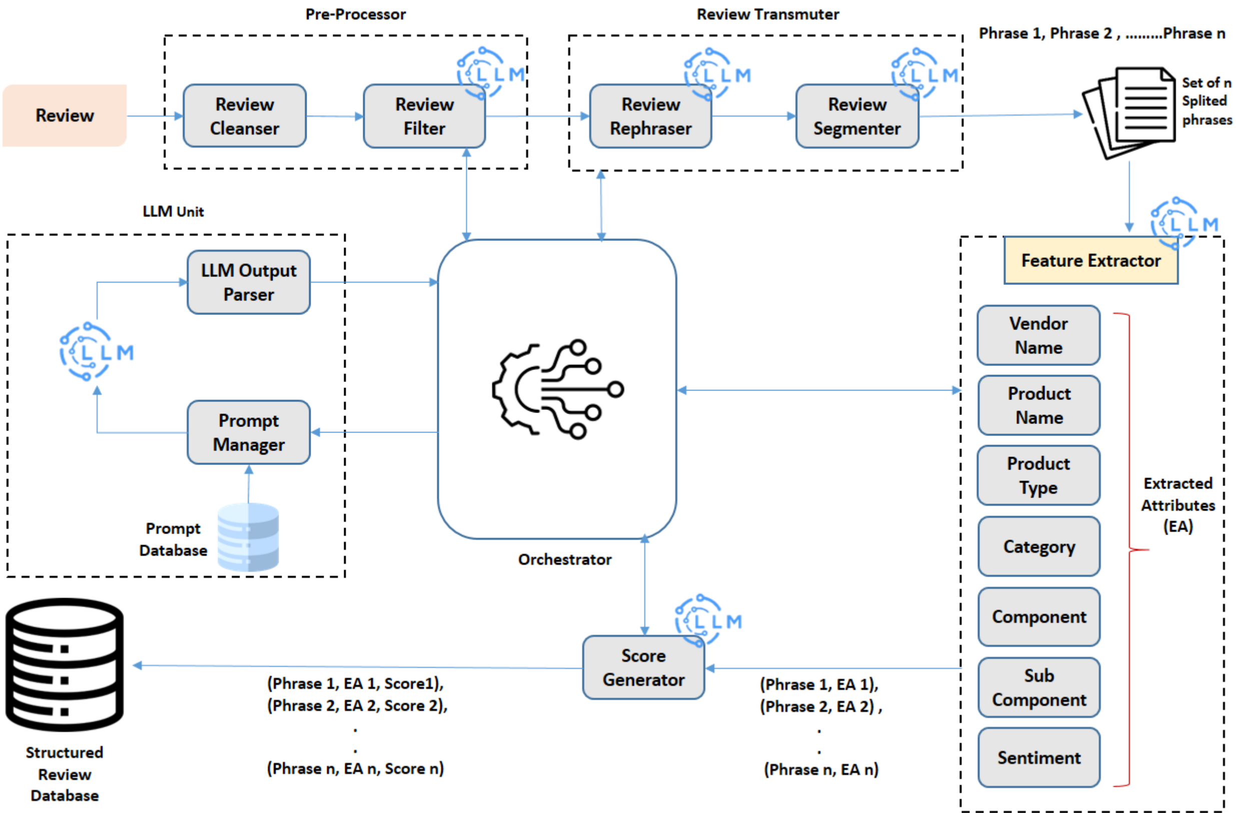

Figure 2: Data Enrichment Module (DEM) Architecture

feedback are removed by this filter. Such reviews may contain keywords but lack the actionable insights needed for the analysis.

The pre-processor stage ensures that only relevant and clean review data moves forward to the next stage, significantly reducing noise and improving the quality of the data being processed.

2. **Review Transmuter**:

   The review transmuter further processes the cleansed reviews by focusing on improving readability and splitting reviews into manageable segments. This stage contains two blocks:

   - **Review Rephraser**: The review rephraser corrects grammatical and spelling errors, enhances readability, and removes unnecessary content while retaining the core feedback. The rephrased review is clearer and more focused, allowing for more accurate feature extraction downstream.
   - **Review Segmenter**: This block splits reviews into smaller segments, each focusing on a specific product component, sub-component, or sentiment. Many reviews cover multiple topics. For example, a review might mention both battery life and screen quality. Splitting these topics into individual segments allows for finer-grained analysis. The segmentation ensures that each topic or sentiment can be independently processed by subsequent modules.

3. **Feature Extractor**: Once reviews are segmented, the feature extractor processes each segment to identify key attributes related to the product and review context. This block extracts a variety of structured data from each segment, including:

   - **Vendor Name**: Identifies the brand or manufacturer mentioned in the review (e.g., Samsung, Apple)
   - **Product Name**: Extracts the specific product being discussed (e.g., iPhone 14, Galaxy S23)

- **Product Type**: Categorizes the product (e.g., smartphone, laptop, refrigerator)
- **Review Category**: Classifies the type of feedback provided (e.g., issue, query, comparison, suggestion, or appreciation)
- **Component and Sub-component**: Captures the specific product features discussed, both at a major (component) and more granular (sub-component) level. For example, "camera" might be the component, with "zoom" being the sub-component.
- **Sentiment**: Classifies the sentiment expressed in each segment as positive, negative, or neutral.

By breaking down reviews into these components, the system can generate detailed insights that are highly specific to individual product features. Examples of unstructured reviews and their corresponding structured outputs can be found in **Appendix C**.

4. **Score Generator**:

The score generator assigns a severity score to each review segment based on both qualitative and quantitative factors. This score is used to prioritize reviews that need immediate attention or highlight critical product issues. The score generator has two key components:

- **Qualitative Score**: This score evaluates the severity or intensity of the sentiment expressed in the review segment. This module employs LLM to generate the sentiment score for each review segment.
- **Quantitative Score**: This score is based on the importance of the product, component and sub-component mentioned in the review. For instance, reviews about high-priority features or products is given greater weight.

The combined severity score is used to rank the review segments, ensuring that reviews with critical feedback are given priority for analysis.

5. **LLM Unit**:

The LLM Unit is the core of the DEM, responsible for carrying out advanced language processing tasks that require deep contextual understanding. The LLM Unit contains the following sub-components:

- **Prompt Manager**: The prompt manager retrieves and manages prompt templates for various tasks in the pipeline. The few-shot prompting mechanism is used to fine tune the LLM for task-specific needs.
- **LLM Output Parser**: After the LLM generates an output, the output parser cleanes up any formatting errors, hallucinations, or irrelevant information provided by the LLM.

6. **Orchestrator**:

The orchestrator manages the overall workflow of the DEM, ensuring smooth communication between different components. It handles the following tasks:

- **Data Routing**: The orchestrator routes data from one block to another in sequence.
- **Task Coordination**: The orchestrator coordinates tasks, invoking the LLM Unit for tasks execution.
- **Output Management**: Once all stages are completed, the orchestator ensures that the processed and enriched data is stored in a structured format, ready for retrieval by the ADAM.

3.2 Analytic Dashboard AI Module (ADAM)

The Analytic Dashboard AI Module (ADAM), depicted in the **Figure 3**, is an intelligent system designed to transform raw structured data into actionable insights. At its core, ADAM leverages a LLMs orchestrated by a dedicated LLM Orchestrator, ensuring efficient and context-aware processing of information. This Orchestrator works in conjunction with a Prompt Manager, which dynamically crafts prompts to guide the LLM in generating targeted summaries, analyses, and visualizations. ADAM seamlessly integrates with DEM database to process vast quantities of structured data. Through a natural language interface, users can query the system conversationally. Users can request information and analysis on various products, generating actionable insights.The agentic flow of ADAM includes the following stages:

- **Natural Language Query Interpretation**: User interaction with ADAM begins with the submission of a natural language query (e.g., "What are the top complaints about battery life in Samsung smartphones from the last six months?"). ADAM's NLU engine parses and interprets this query, extracting key attributes such as the target product (Samsung smartphones), the relevant feature (battery life), the desired sentiment (negative complaints), and the specified timeframe (last six months). The engine then constructs a structured query suitable for execution against the underlying database. This stage translates user intent from a natural language into a machine-readable format.

- **Data Retrieval and Ranking**: Once filters are applied (manually or via natural language queries), ADAM retrieves the relevant customer reviews from the structured database. The reviews are then ranked using the combined contextual and quantitative scores generated by the DEM. This ranking prioritizes the most critical reviews, such as those highlighting major product issues.

- **Summarization and Comparative Analysis**: ADAM uses an LLM to generate concise, insightful summaries of the retrieved data. Before summary generation, a quantitative analysis, termed Attribute Frequency Aggregation (AFA), is performed. The AFA results and the raw data are then provided as input to the LLM's summarization module. Carefully crafted prompts guide the LLM. The summaries incorporate both quantitative insights from quantitative insights from AFA and qualitative details extracted from the data, including relevant excerpts from user reviews. This ensures comprehensive overview of user sentiment and product performance. For comparative analyses (e.g., comparing two products), ADAM performs this process for each product. The LLM-generated summaries side-by-side, highlighting key differences and similarities.

- **Visualization and Reporting**: ADAM integrates with graphical libraries to present data visually. This enhances user comprehension and facilitates the identification of key trends and patterns. Visualizations include bar charts, line graphs, trend charts, comparative tables, and word clouds. These graphics help users visualize trends in sentiment distribution and product performance. For example, a trend chart could display changes in user sentiment regarding battery life over the past year, while a comparative table could juxtapose the performance of two competing products.

4 Experiments, Results & Observations

4.1 Experimental Setup

The V2U framework was deployed and tested on Google Cloud Platform (GCP). Two versions of the Mistral-7B-Instruct model (Jiang et al., 2023) were employed: a base model for review cleaning, rephrasing, attribute extraction, sentiment analysis (including scoring), and the Analytic Dashboard Module (ADAM) module; and a fine-tuned model for component and sub-component classification. The VLLM inference engine (Kwon et al., 2023) was used for efficient processing and query response. The framework's performance was evaluated on four tasks—attribute extraction, Aspect Based Sentiment Analysis (ABSA), review summarization, and review ranking. For experiment we have used a in-house dataset of 1 million English-language customer reviews for smartphones, TVs, and digital appliances. Baseline models for comparison included GPT-4 and RoBERTa (for attribute extraction) and GPT-4 for ABSA, summarization, and review ranking.

4.2 Attribute Extraction (NER and Classification Tasks)

We evaluated the attribute extraction performance of V2U's DEM pipeline for tasks such as vendor name, product type, product name, review category, and sentiment classification. Precision, recall, and F1-score were used to assess performance.

Table 1 shows that V2U achieves competitive performance compared to GPT-4 across all classification and NER tasks, while maintaining cost-effectiveness by utilizing an open source model. These results demonstrate that V2U is well-suited for transforming unstructured reviews into structured data for further analysis.

4.3 Aspect-Based Sentiment Analysis(ABSA)

The V2U framework was evaluated for its ability to perform ABSA by extracting both major components (e.g., camera, battery, display) and their respective sub-components (e.g., battery life, screen

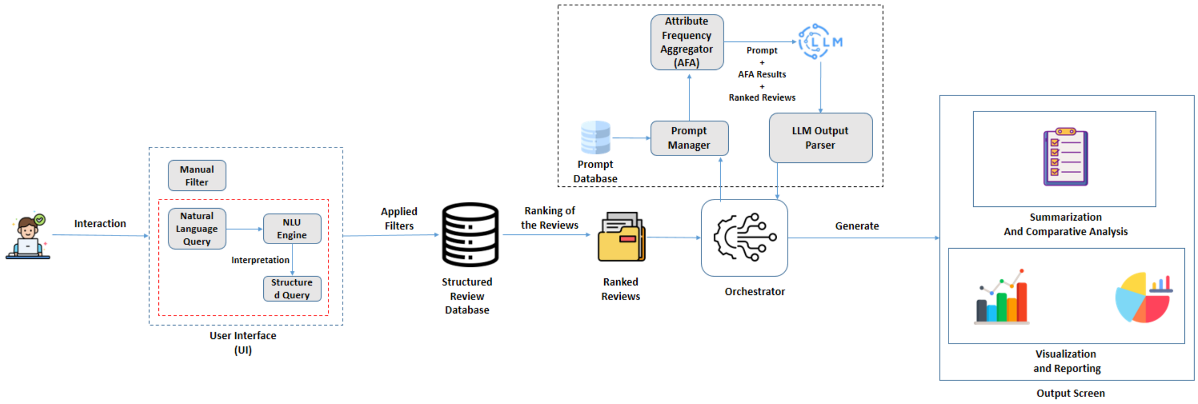

Figure 3: Analytic Dashboard AI Module (ADAM) Architecture

Task	Model	Precision (%)	Recall (%)	F1-Score (%)
Vendor Name	V2U	88.5	88.1	88.3
	GPT-4	90.7	90.3	90.5
	RoBERTa	82.3	81.9	82.1
Product Type	V2U	87.3	86.9	87.1
	GPT-4	89.8	89.4	89.6
	RoBERTa	75.6	75.0	75.3
Product Name	V2U	88.9	88.5	88.7
	GPT-4	90.5	90.1	90.3
	RoBERTa	73.5	72.9	73.2
Review Category	V2U	87.8	87.4	87.6
	GPT-4	89.2	88.8	89.0
	RoBERTa	72.9	72.1	72.5
Sentiment Classification	V2U	87.5	87.0	87.2
	GPT-4	89.7	89.1	89.4
	RoBERTa	83.9	83.5	83.7

Table 1: Attribute Extraction: Precision, Recall, and F1-Score

brightness). The following table compares the precision, recall, and F1-Score of V2U with GPT-4.

Table 2 shows that V2U performs at a high level, achieving similar results comparable to GPT-4, while processing data at lower computational costs. This makes V2U an ideal tool for consumer electronics companies looking to analyze product features based on customer feedback.

4.4 Review Summarization

We evaluated the summarization capabilities of V2U's ADAM using Retrieval Augmented Generation (RAG) metrics such as context recall, answer relevance, conciseness, summary coherence, and readability (Lewis et al., 2021).

Table 3 shows that V2U demonstrates strong summarization capabilities, producing concise, relevant summaries with accuracy comparable to GPT-4. This is essential for providing actionable insights to product managers, especially when analyzing product features such as camera quality or battery life from a large volume of customer reviews.

4.5 Review Ranking Based on Component Severity

A unique feature of V2U is its ability to rank reviews based on major component, sub-component severity, product priority, frequency, and recency. This capability helps business to prioritize the most critical issues that require immediate attention.

Table 4 demonstrates how V2U ranks reviews to help product teams identify the most pressing product issues, enabling faster decision-making and more efficient resource allocation for product development.

Major Component	Sub-Component	Model	Precision (%)	Recall (%)	F1-Score (%)
Camera	Low-Light	V2U	87.6	87.0	87.3
		GPT-4	89.3	89.0	89.2
	Zoom	V2U	86.5	85.8	86.1
		GPT-4	88.9	88.6	88.7
Battery	Battery Life	V2U	87.9	87.5	87.7
		GPT-4	89.8	89.1	89.4
	Charging Speed	V2U	88.5	88.0	88.2
		GPT-4	90.0	89.6	89.8

Table 2: Aspect-Based Sentiment Analysis Results: Major Components and Sub-Components

Metric	V2U	GPT-4
Context Recall (%)	84.9	86.1
Answer Relevancy (%)	85.7	87.0
Conciseness (%)	83.5	85.2
Summary Coherence (%)	84.2	86.0
Readability (%)	82.8	84.5

Table 3: Review Summarization Results: Comparison between V2U and GPT-4

4.6 Latency and Scalability

To evaluate V2U's DEM scalability, we processed review datasets of increasing sizes: 1000, 10,000, 50,000 and 100,000 reviews. As shown in **Table 5**, it maintains a near-linear increase in processing time, handling 100,000 reviews in approximately 140 minutes, with an average latency per review of 85ms. This demonstrates V2U's DEM scalability, ensuring efficient processing as the dataset size grows.

5 Conclusion

The experiments and results demonstrate that the V2U framework performs comparably to GPT-4 in tasks such as attribute extraction, ABSA, summarization, and review ranking. Additionally, V2U's scalability makes it a powerful and practical solution for business seeking to extract actionable insights from customer reviews. The system's ability to process large datasets with low latency and provide real-time insights underscores its value for consumer electronics companies.

Limitations

Currently, V2U supports only English reviews. Extending it to handle multiple languages would require exploration of multilingual models to ensure accurate processing across different languages. This presents an additional challenge, as multilingual sentiment analysis and aspect extraction may involve different linguistic structures, sentiments, and context-specific nuances. Finally, while V2U efficiently processes data in batches, it lacks real-time data handling capabilities, which may be necessary for businesses requiring immediate insights. These limitations highlight areas for further refinement and adaptation of the V2U framework to enhance its utility and effectiveness.

Ethics Statement

This research utilizes company internal data for experimentation purposes. All personally identifiable information (PII) has been rigorously anonymized and de-identified before analysis, ensuring the privacy of employees and other stakeholders. The data was handled securely in accordance with [Samsung's Data Privacy Policy/Relevant Regulations]. Access to the data was restricted to authorized personnel only, and all analyses adhered to the company's strict data security protocols to safeguard against unauthorized disclosure or use.

Acknowledgments

We thank Javaid Nabi and Kartikey Singh for reviewing the draft. The following individuals contributed to the study: Radhika Mundra (conceptualization and leadership); Aritra Ghosh Dastidar, Kartikey Singh, and Bhavesh Kukreja (methodology and investigation); and Radhika Mundra and Bhavesh Kukreja (original draft preparation).

References

Yogesh Chandra and Antoreep Jana. 2020. Sentiment analysis using machine learning and deep learning. In *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 1–4.

Issue	V2U's Score	V2U's Rank	GPT-4 Score	GPT-4 Rank
Camera Quality	4.8	1	4.7	1
Battery Life	4.3	2	4.5	2
Application Freezing	3.9	3	3.7	5
Processor Speed	3.8	4	3.9	4
Display Quality	3.7	5	4.0	3
Speaker Quality	3.3	6	3.5	6
Button responsiveness	3.0	7	3.1	7

Table 4: Review Ranking Based on Severity Score

No. of Reviews	Process Time	Latency
1,000	1.5 mins	80ms
10,000	15 mins	82ms
50,000	70 mins	85ms
100,000	140 mins	87ms

Table 5: Latency and Scalability

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Cuiyun Gao, Yaoxian Li, Shuhan Qi, Yang Liu, Xuan Wang, Zibin Zheng, and Qing Liao. 2023. Listening to users' voice: Automatic summarization of helpful app reviews. *IEEE Transactions on Reliability*, 72(4):1619–1631.

Akkamahadevi R Hanni, Mayur M Patil, and Priyadarshini M Patil. 2016. Summarization of customer reviews for a product on a website using natural language processing. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2280–2285.

Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2018. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-augmented generation for knowledge-intensive nlp tasks.

Ze Shi Li, Nowshin Nawar Arony, Kezia Devathasan, Manish Sihag, Neil Ernst, and Daniela Damian. 2024. Unveiling the life cycle of user feedback: Best practices from software practitioners. In *2024 IEEE/ACM 46th International Conference on Software Engineering (ICSE)*, pages 641–653.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Gaurav Negi, Rajdeep Sarkar, Omnia Zayed, and Paul Buitelaar. 2024. A hybrid approach to aspect based sentiment analysis using transfer learning.

Daniela Oelke, Ming Hao, Christian Rohrdantz, Daniel A. Keim, Umeshwar Dayal, Lars-Erik Haug, and Halldor Janetzko. 2009. Visual opinion analysis of customer feedback data. In *2009 IEEE Symposium on Visual Analytics Science and Technology*, pages 187–194.

Pankaj, Prashant Pandey, Muskan, and Nitasha Soni. 2019. Sentiment analysis on customer feedback data: Amazon product reviews. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pages 320–322.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter.

Hang Yan, Junqi Dai, Tuo ji, Xipeng Qiu, and Zheng Zhang. 2021. A unified generative framework for aspect-based sentiment analysis.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A survey of large language models.

Linan Zhu, Minhao Xu, Yinwei Bao, Yifei Xu, and Xiangjie Kong. 2022. Deep learning for aspect-based sentiment analysis: a review. *PeerJ Computer Science*, 8:e1044.

A Appendix

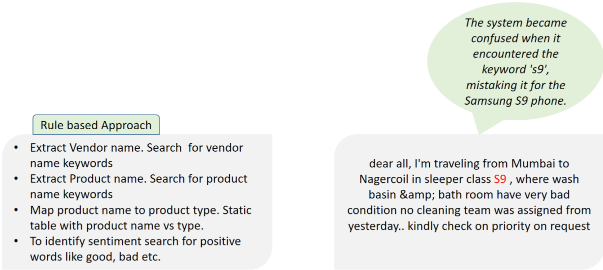

Figure 4: Rule Based Approach And Limitation

Figure 4 illustrates the limitations of a rule-based approach. Contextual ambiguity caused the rule-based system in the example to fail. For instance, a simple rule incorrectly interpreted "S9" as referring to a Samsung S9 phone, despite its use in a different context.

B Appendix

The **Figure 5** illustrates the flow of data through the DEM pipeline, showing inputs and outputs at each processing stage.

C Appendix

The examples of unstructured reviews and their corresponding structured attributes are presented in **Table 6**.

Unstructured	Structured				
Customer Review	**Vendor Name**	**Product Name**	**Product Type**	**Sentiment**	**Category**
Amazing camera! The IPhone 15's low-light performance and zoom are incredible. Photos and videos are crisp and vibrant. Highly recommend!	Apple	iPhone 15	smartphone	positive	appreciation
The sound on my LG C1 is ruined by a constant, high-pitched crackling. Troubleshooting hasn't helped. Beautiful picture, but unusable sound makes this a very expensive disappointment.	LG	LG C1	television	negative	issue
I recently purchased the Samsung WF45R6300AW washing machine and I'm blown away by the spin cycle. My clothes come out significantly less wet than with my old machine, meaning faster drying time and fewer wrinkles. It's incredibly efficient and powerful, yet still quiet. If you're looking for a washing machine with a superior spin cycle, this is it!	Samsung	WF45R6300AW	washing machine	positive	appreciation

Table 6: Unstructured Reviews and Structured Attributes

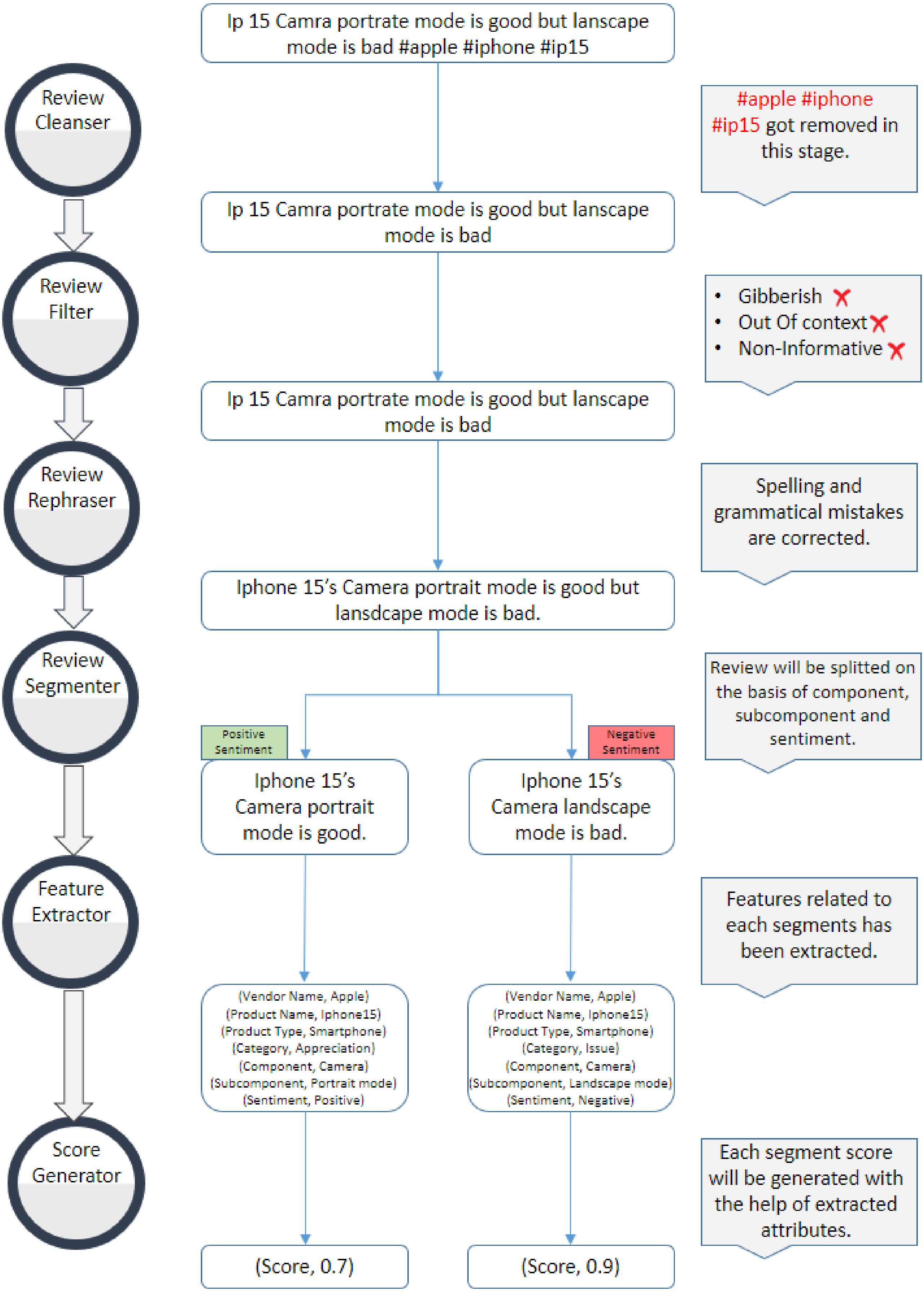

Figure 5: A Detailed Flow of DEM Architecture